Ground Cover Change Model Framework: User’s Guide II

How to collect a Global Ground Cover Change training dataset?

This guide provides detailed steps to collect large datasets of high-quality training samples for land cover classification models. The resulting dataset is temporally-consistent for a pre-defined region in an eco-zone of interest. A guide to source ‘temporally-consistent regional metrics for filtering a Ground Cover Change training set” is also available. Additional information with regional metrics, mapped products, and land cover models can be accessed through our Mendeley Repository v.4 <https://data.mendeley.com/datasets/mzp3k6fmtz>.

1. For year ‘x’ and region ‘y’: download ICESat-2 ATLAS08 Spaceborne LiDAR data:
   1. Extraction: North (latitude\_20\_m) from upper terrain/land end to lower terrain/land end. Copy/paste columns one beneath the other in a single column.
   2. Extraction: East (longitude\_20\_m) from fid (North) to fid (South).
   3. No sorting, no erasing.
   4. Extract LiDAR Data: Canopy\_H\_20m from upper to lower end ids.
   5. Extract LiDAR Data: Canopy\_Mean from upper to lower end ids.
   6. (Optional) Extract LiDAR Data: Slope from upper to lower end ids.
2. Save LiDAR data to CSV file.
3. Erase LiDAR Data with 3.4E\*\* values for: Canopy\_Mean, Canopy\_H\_20m, Longitude, and Latitude.
4. Repeat steps 1-3 for each strong beam Ground Track (GT) (upon desired training data size). Here, we used one GT as metric samples and pre-processed eighteen GTs to calibrate training data samples.
5. Merge Data: all training samples in a single .csv file.
6. Upload .csv file to QGIS: Save as point .shp file.
7. Upload land cover product to QGIS 3.
8. Clean LiDAR point data to avoid autocorrelation problems; erase points that are ≥45 meter away from class perimeter (perimeter= inter-class border).
9. Append and sample the land cover label classifications in band 1. (<https://glad.umd.edu/dataset/global-land-cover-land-use-v1>)
10. Save as SHP file and upload the file to Google Earth Engine in assets folder. GEE Access: (<https://code.earthengine.google.com/?accept_repo=users/martingarciafry/gf>).
11. Run the code for a target year to sample atmospherically-corrected and cloud-masked Sentinel-2 image collections.
12. Sample raster bands during: (i) ‘Cultivation Season’ (March/April), (ii) ‘growing season\_1’ (May/June), (iii) ‘growing season\_2’ (July/August), and (iv) ‘harvest season’ (October) of a target year.
13. Download multitemporal datasets from Google Drive and then stack them using each point’s corresponding id value. Note that repeated Lon/Lat values will be displayed.
14. Once a compact dataset is compiled, add a ‘class code’ column to identify land cover labels with own values (1 – 7). We used: Dense Short Vegetation (1), Open Tree Cover (2), Dense Tree Cover (3), Wetland (4), Built up (5), Water (6), and Cropland (7) based on a desired study of land surface phenology and in accordance with FAO’s land classifications.
15. Now, upload ancillary data to QGIS: MERIT Digital Elevation Model, Geomorpho90m -Topographic Wetness Index, and the Global Forest Cover Height product.
16. Append ancillary data with nearest neighbor sampling: Here we used ‘sample raster values’ to append data for each sample unit:
    1. First, sample DEM elevation data (Ground height) for choice of DEM product. Here, we used the MERIT DEM (http://hydro.iis.u-tokyo.ac.jp/~yamadai/MERIT\_DEM/).
    2. Second, sample Topographic Wetness Index (TWI). Here we used: Geomorpho90m, a MERIT DEM derived data product to measure topographic-related metrics. (http://www.spatial-ecology.net/dokuwiki/doku.php?id=topovar90m).
    3. Third, append the Forest Height data product: we used the Global Forest Cover Change product with height metrics in Band 1 (https://glad.umd.edu/dataset/gedi/).
    4. ‘Raster>Analysis>Aspect’ is estimated using the DEM model in TIFF format (~90m). (Tutorial: https://www.youtube.com/watch?v=B-5RQ9o9EyU).
    5. ‘Raster>Terrain Analysis>Slope’. Or use GDAL>Slope (Tutorial: https://www.youtube.com/watch?v=7eIFvZ4fU6k). Slope is also estimated using the DEM model in QGIS with TIFF format (~90m). We used a WGS 8 UTM CRS projected from the default CRS [4326].
    6. Fourth, sample the topographic solar radiation layer estimated with ‘GRASS>r.sun.insoltime’ for the Day 200 of a target year close to July (due to the apparent lack of clouds while seeking Sentinel-2 images), using the Elevation layer, Aspect layer, and Slope layer in QGIS 3.21. (Tutorial: <https://www.youtube.com/watch?v=0z2trThOYaQ>).
17. Save LiDAR point data with appended ancillary data as .csv file with ids.
18. Estimate the following indices for land surface phenology stages (Use B8 with 10 m bands (i.e., B2, B3, and B4) and use the B8A with near- and short-wave infrared bands (i.e., B5, B6, B11, and B12):
    1. NIR\_Green (B8-B3/B8+B3).
    2. Normalized Difference Vegetation Index (B8-B4/B8+B4).
    3. SWIR\_1\_SWIR\_2 (B11-B12/B11+B12).
    4. Tasseled Cap Greenness (-0.3599\*B2-0.3555\*B3-0.4734\*B4+0.6633\*B8-0.0087\*B11-0.2856\*B12).
    5. Tasseled Cap Wetness (0.2578\*B2+0.2305\*B3+0.0883\*B4+0.1071\*B8-0.7611\*B11-0.5308\*B12).
    6. Normalized Difference water index\_I (B3-B8/B3+B8).
    7. Normalized Difference water index\_II (B8A-B11/B8A+B11).
    8. BN (B2-B11/B2+B11)
    9. SAVI (B8-B4/B8+B4+0.5) x1.5
    10. Finally, using a moving-average time-series of the NDVI column, estimate the first, third, and last quartile for NDVI, Band 3 (Blue), Band 8 (NIR), and Band 11 (SWIR 1) to classify cropland periods of change.
19. Three important quality filters were used to classify data (A filtering tool can be found in our Mendeley repository to automate this part:
    1. First, spectral metric upper/lower bounds using band combinations and static ancillary data for manually-agreed labels only.
    2. Second, class height thresholds are determined to filter each class unit using ICESat-2 values corresponding to canopy\_mean, or h\_canopy\_20 where no canopy\_mean values exist.
    3. Third, Forest height values are used to validate ICESat-2 canopy heights if these are within a range of 6 m.
20. A variation of discrete heights sourced from FAO (<https://www.fao.org/3/x0596e/x0596e01f.htm#p665_54535>) were used; (1) Dense short vegetation (shrubs, percentage of bare areas were available in the land cover product, but more than 95% of our samples returned a 0% bare cover. Therefore, we did not factor in this measure) ≤5 m, (2) Open Tree Cover (coniferous, broadleaf, mixed woody plants, wetlands) >5 m and < 12 m, (3) Dense Tree Cover (closed dry and closed wetland) >12 m, (4) built up (this class was vetted on the basis that most structures in rural areas have up to 2 levels and pixels are combined with canopy cover) < 8 m, (5) water (open water bodies) = 0 m, and (6) cropland (cultivated and managed land) < 5 m for agriculture and > 5 meters for managed forests.
21. Once the dataset has been classified, exclude the least reliable labels (level 0 and level 1) to ascertain a minimum degree of quality training labels.
22. Finally, use the provided training data set template.csv to eliminate unnecessary data columns and export the .csv file to a .geojson file for model training.
23. A link to the model is provided from the initial page of the Mendeley Repository.

For reference, the filtering tool uses a simple Python programming language with the imported Numpy library to classify candidate training units. This tool supervises class-segmented datasets and saves a classified compact dataset for user-based discriminations. All output files in .csv format are saved locally in the folder location where the tool is saved.

Additional information in Mendeley Repository (<https://data.mendeley.com/datasets/mzp3k6fmtz>):

1. “How to collect a Global Ground Cover Change dataset?” – User’s Guide.
2. “How to source temporally-consistent regional metrics to filter a Ground Cover Change training dataset?” – User’s Guide.
3. ‘training dataset\_template’.
4. Filtering Tool
5. Product map classifications using the Ground Cover Change model.
6. Accuracy Assessments
7. Shapefiles, metadata, datasets, etc.

Contact Information:

[globalgroundcoverchange@gmail.com](mailto:globalgroundcoverchange@gmail.com)

Thank you, good luck!