Ground Cover Change Model Framework: User’s Guide I

How to source temporally-consistent regional metrics to filter a Ground Cover Change training dataset?

This guide provides detailed steps to supervise candidate training samples automatically and collect large datasets of high-quality training units for land cover classification models. The metrics are temporally-consistent for a pre-defined region of choice in an eco-zone of interest. A guide on “How to collect a Global Ground Cover Change training dataset?” 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. Extract a Spaceborne ICESat-2 ground-track acquisition over a wide and heterogeneous area of interest.
   1. Extraction: North (latitude\_20\_m) from upper terrain/land end to lower terrain/land end. East (longitude\_20\_m) from fid (North) to fid (South).
   2. Copy/paste columns North one beneath the other in a single column.
   3. Copy/paste columns East one beneath the other in a single column.
   4. No sorting, no erasing.
   5. Extract LiDAR Data: Canopy\_H\_20m from upper to lower end fids.
   6. Extract LiDAR Data: Canopy\_Mean from upper to lower end fids.
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. Upload point data to QGIS and upload a pre-existing land cover product.
5. Clean LiDAR point data to avoid autocorrelation problems; erase points landing ≥45 m away from inter-class perimeters (perimeter= class border).
6. Sample the land cover product labels.
7. Upload google earth satellite.cn to manually supervise pixel labels where points have landed. Consider agreement for labels with >50% of the pixel covered by that class.
8. Upload a forest height product to be used in supervision procedures below.
9. Sample the forest height product.
10. Then, use the following filters to supervise your ‘agreed/disagreed’ interpretations with a binary output column for each filter:
    1. Class-Height thresholds (see step 11 below) to filter the vertical structure of each class using ICESat-2 canopy\_mean values, or h\_canopy\_20 values where no canopy\_mean values exist.
    2. Use forest height values to validate ICESat-2 canopy heights in range of 6 m.
11. 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.
12. Sum the three columns (thematic agreement column, class height agreement column, and the lidar vetting column) in a new ‘reliability’ column.
13. Once these are summed, extract level 3 samples from the dataset.
14. Erase level 0 and level 1 to eliminate spurious data units.
15. Retrieve level 3 samples having 500 m buffer distances to validate mapping outcomes.
16. Create a new EXCEL file and paste the remaining level 3 samples.
17. Add level 2 samples with positive agreements and positive class height supervisions to the new EXCEL file. Discard the rest of level 2 samples.
18. Name this file ‘metrics dataset.csv’.
19. Upload .csv file to QGIS and save as .shp file.
20. Upload the .shp file to Google Earth Engine in assets folder. GEE Access: (<https://code.earthengine.google.com/?accept_repo=users/martingarciafry/gf>).
21. Run the code: verify target year in a classification region of interest and then run the code to sample atmospherically-corrected and cloud-masked Sentinel-2 mosaics.
22. 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) for the target year.
23. Export and then download multitemporal datasets from Google Drive.
24. Stack these datasets using each point’s corresponding id value. Note that repeated Lon/Lat values will be displayed.
25. Upload the updated metrics dataset as .csv file to QGIS.
26. Now, append ancillary data with nearest neighbor sampling. Here, we used ‘sample raster values’ to append ancillary data to 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>).
27. Calculate the mean temporal surface reflectance value corresponding to each point in the metrics dataset using ‘QGIS>raster calculator>mean’, or using formulas in EXCEL. The appended data has static temporal values, so they will be used later.
28. 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.
29. In EXCEL, regressions will reveal which features (static or temporal) are sensitive to the independent variables of choice. We followed literature guidelines to select representative features for treed classes (NDVI), wetlands (SAVI), Built-up areas (BN), water (NDWI\_I), and cropland (SAVI). Upper 95th percentile sensitive features should be selected as metric contenders (5-8 metrics per class).
30. After selecting features, proceed to extract data bounds using ‘Data Analysis>Descriptive Statistics’ and obtain the max, min, and standard deviation of each feature per class.
31. Use the standard deviation to calculate the min/max range values (min. + std. & max – std.).
32. These become the official MEAN metrics for supervising training data with a filtering tool, available in the Mendeley Repository: <https://doi.org/10.17632/mzp3k6fmtz.4>.
33. We recommend using multitemporal metrics (avoiding Step 27) because labeled agreements have been potentially well supervised using ICESat-2 canopy heights and a reliable metric sample set has been rendered.

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 compact classified dataset for user-based discriminations. All output files in .csv format are saved locally in the same location where the tool was saved.

Additional information in Mendeley Repository v.4 (<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!