GLAD Tools V2.0

Quick Start Guide

GLAD ARD data and tools are available at https://glad.umd.edu/ard/home

The GLAD Tools and Landsat ARD data are available with no charges and no restrictions on subsequent redistribution or use, as long as the proper citation is provided as specified by the Creative Commons Attribution License (CC BY).

Suggested citation:

Potapov, P., Hansen, M.C., Kommareddy, I., Kommareddy, A., Turubanova, S., Pickens, A., Adusei, B., Tyukavina A., and Ying, Q., 2020. Landsat analysis ready data for global land cover and land cover change mapping. *Remote Sensing*, 2020, 12, 426; doi:10.3390/rs12030426 *https://www.mdpi.com/2072-4292/12/3/426*

The lates update: September 2023.

The full GLAD Tools Manual is available with the software installation:

C:\GLAD_Tools\Documentation\ARD_Manual_v2.0_2023.pdf

and online at https://glad.umd.edu/users/Potapov/ARD/ARD_Manual_v2.0_2023.pdf

This document summarizes the basics of GLAD Tools installations and functions.

1. GLAD Tools V2.0 Installation (Windows 10/11)

1.1. System Requirements

- Windows 10 or 11 (64-bit).
- 16 GB RAM (8GB RAM for limited capacity).
- Enough disk volume for data storage and processing. The disk volume requirement depends on the area of analysis and time interval. The following average data volumes may be used to estimate the required disk space:
  - ARD 16-day data for one tile, one year – 5 GB
  - Phenological metrics for one tile, one year – 6.5 GB
  - Change detection metrics for one tile, one year – 12 GB.

  Example: For mapping LCLU and performing the sample analysis of decadal (2012-2022) LCLU changes within 1x1 geographic degree area, the user should reserve at least 100 GB of HDD space.
- Administrative privileges are required for software installation.

1.2. Installation of Required Open Source and Free Software

1.2.1. PERL

- Download the 64-bit version of PERL programming language from http://strawberryperl.com/
- Install using default installation parameters.
- Restart your computer.

1.2.2. Google Earth Desktop

- Download the latest installer from https://www.google.com/earth/versions/#earth-pro
- Install using default installation parameters.

1.2.3. QGIS and OSGeo4W

- Download the latest QGIS/OSGeo4W installer from: https://qgis.org/en/site/forusers/download.html. **The GLAD Tools require QGIS 3.30 or later version.**
- Install using default installation parameters.
- Restart your computer.
- Open **QGIS Desktop**.
- Open the menu "**Plugins / Manage and Install Plugins**".
- Select and install the following plugins:

  - **Send2GE**

    This plugin will open the Google Erath Desktop app and show the location of the cursor in QGIS map.

  - **QuickMapServices (QMS)**

    This plugin allows adding Google and Bing satellite image layers to the QGIS map.
- Restart QGIS to implement changes.

1.2.4. Text Editor

We recommend **Notepad++**, an open-source editor, for working with parameter files and scripts in GLAD Tools. The latest version of the editor can be downloaded here: https://notepad-plus-plus.org/downloads/

1.2.5. R

- Download and install the latest version of R from https://cloud.r-project.org/
- Install using default installation parameters.
- Restart your computer.
- Open **R Console**.
- Execute the following commands:

```
install.packages("ggplot2", repos="http://cran.rstudio.com/")
install.packages("dplyr", repos="http://cran.rstudio.com/")
```

- Close R console.

1.3. GLAD Tools Installation

- Download the latest complete package of GLAD Tools:
  https://glad.umd.edu/Potapov/ARD/GLAD_Tools.zip
- Create folder **C:\GLAD_Tools** and unpack the content of the **GLAD_Tools.zip** file into this folder keeping the original subfolder structure and files and folder names.
- Open the **C:\GLAD_Tools** folder in File Explorer.
- Right-click on the file **Add_PATH_for_GLAD_v1.1.bat** and select the option "**Run as Administrator**".
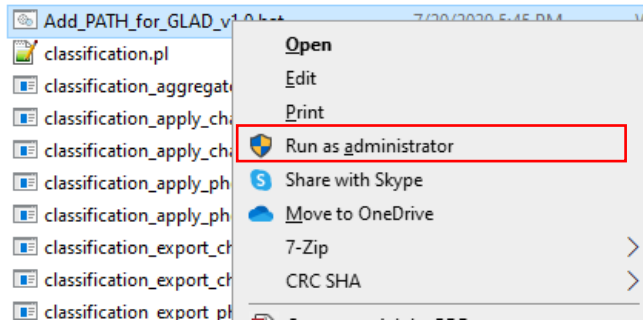
- Restart your computer.
- Rename **C:\GLAD_Tools\dependencies_example.txt** file into **C:\GLAD_Tools\dependencies.txt**.
- Open the file **C:\GLAD_Tools\dependencies.txt** in a text editor.
- Find the paths to files **OSGeo4w.bat** and **Rscript.exe** and fill in the corresponding parameters in the text file.

Example of **dependencies.txt** file
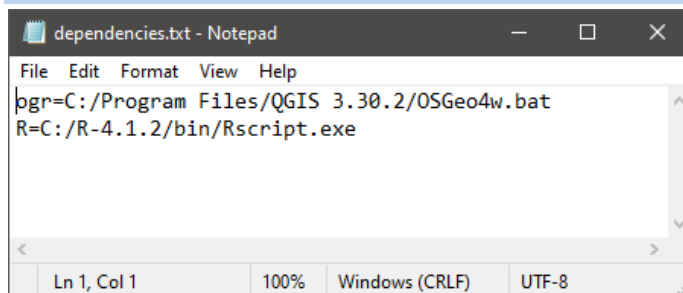
- Save the **C:\GLAD_Tools\dependencies.txt** file.

2. Data Management Guidelines

GLAD Tools requires users strictly follow data organization rules. All classification projects and datasets should be created in **separate folders** to avoid software errors. The following section presents an optimal data organization approach. We suggest storing all GLAD Tools-related files in a single workspace, e.g., **C:\GLAD_Workspace**. The **folder and file names should not have spaces**, otherwise, some of the Tools will fail to run.

2.1 Data Storage

The Landsat ARD data and SRTM DEM data should be stored in separate subfolders. We advise users to keep all ARD data in the same folder to avoid duplication. The download code will only process new data, reducing the data download time. The tile sub-folders will be created automatically by the software.
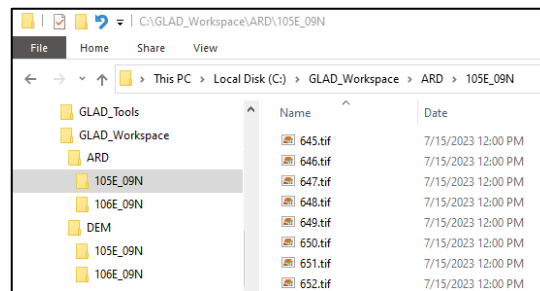
Example of the ARD and DEM storage folders

C:\GLAD_Workspace\ARD

*Landsat ARD data tiles*

C:\GLAD_Workspace\DEM

DEM tiles

2.2 Parameter Files and Metrics

- The **data management parameter files** may be stored in any sub-folder of the workspace. We recommend keeping them **together with the list of commands** and the **list of tiles** to simplify GLAD Tools application and to document your work.
- The **Phenological and Change Detection metrics** should be stored in **separate subfolders** for **each year**.
- **Image mosaics** may be stored in the same folder given that the user specified a **unique output name** for each mosaic.
- Land cover **classifications** must be stored **separately for each thematic class and each year.**
- Multitemporal classifications (**section 8**) are stored separately for each thematic class, but the years are stored together.

Example of the workspace structure
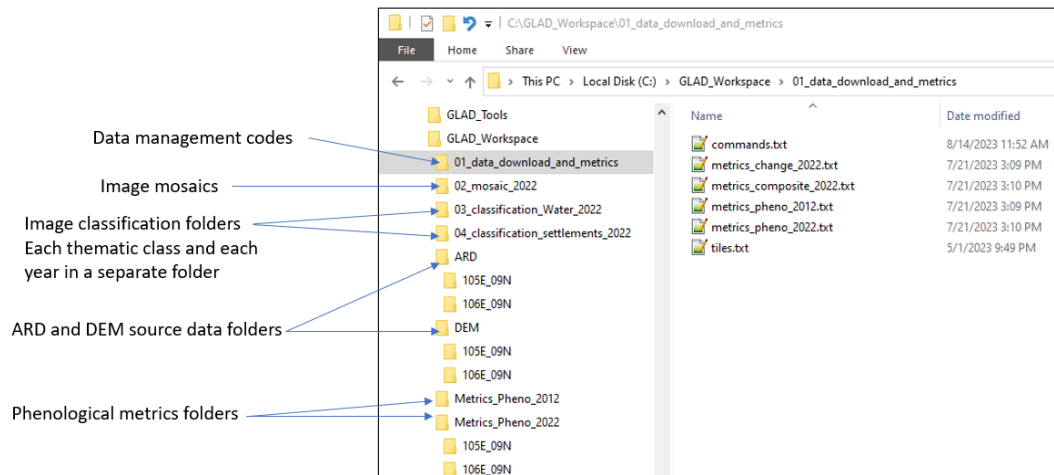
2.3. Using Workspace Templates

To simplify GLAD Tools operations, we provide a set of **workspace templates** that illustrate software applications. The templates are provided in **C:\GLAD_Tools\Templates**.

To use a template, make a new workspace (e.g., **C:\GLAD_Workspace**) and copy there selected sub-folder from **C:\GLAD_Tools\Templates**. Each template includes test tile lists, parameter files, a list of commands, and empty shapefiles for classification training data. Check and correct all parameter files, list of tiles, and commands before application.

To execute a command in the workspace, open **CMD** in the workspace folder. The fastest way to call the **CMD** terminal from a workspace directory is to type "**cmd**" in the address bar of the Windows File Explorer. To do this, first, navigate to the workspace directory, then type "**cmd**" in the File Explorer address bar and press **Enter**. This will open the **CMD** terminal in this workspace directory.
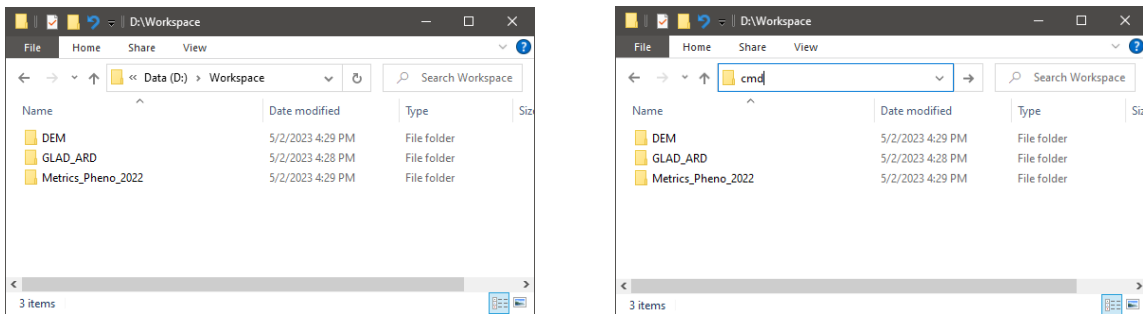
The list of commands in each workspace template is named **commands.txt**. The comments are marked with the "**rem**" prefix. The **command.txt** file may serve as a Windows batch file to run multiple commands in a sequence.

The Quick Start Guide uses workspace templates as examples to illustrate GLAD Tools functions.

3. GLAD Landsat ARD

3.1 GLAD ARD Data Format

The GLAD ARD represents a 16-day time series of globally consistent, tiled Landsat normalized surface reflectance from 1997 to the present, updated every 16 days.

The Landsat ARD data are stored as multi-layer raster tiles. The spatial resolution of the data is **0.00025 degrees per pixel**, which corresponds to 27.83 m per pixel on the Equator.

The ARD product is stored in geographic coordinates using the World Geodetic System (WGS84). The coordinate system is defined by EPSG Geodetic Parameter Dataset:

EPSG:4326 (https://spatialreference.org/ref/epsg/wgs-84/)

Alternatively, it can be defined using the PROJ standard (http://proj.org):

+proj=longlat +ellps=WGS84 +datum=WGS84 +no_defs

The data for each 16-day interval for a tile are stored as 8-band, 16-bit unsigned, LZW-compressed GeoTIFF file.

3.2. Global GLAD ARD Tile System

The global Landsat ARD product is provided as a set of 1x1 geographic degrees tiles. The size of one raster tile is **4004x4004 pixels**, corresponding to an extent of 1.0005 by 1.0005 degrees. The tile system features a 2-pixel overlap. **Tile names** are derived from the tile center and refer to their integer value of the tile center degrees. Tile naming example: The name of a tile with center 17.5E and 52.5N is 017E_52N.

To select ARD data tiles for your area of analysis, use the tile boundary shapefile located in **C:\GLAD_Tools\Data\Global_tiles\glad_ard_tiles.shp**.

To select the list of tiles for your analysis, follow these steps:

1. Open **C:\GLAD_Tools\Data\Global_tiles\glad_ard_tiles.shp** in QGIS together with Google Satellite layer (QMS) or any other map data.

2. Select tile or tiles that overlap your area of interest and check the tile names (The "**Tile**" field of **glad_ard_tiles.shp**).

3. Create a list of tiles as a text file (**tiles.txt**), one line per tile. The text file should not have empty lines or spaces before/after the tile names. Save the file in your workspace (**C:\GLAD_Workspace\tiles.txt**)

4. The list of tiles should be present in each data management and classification folder for convenience.

See the tile list file example here: **C:\GLAD_Tools\Templates\01_Data_Download\tiles.txt.**
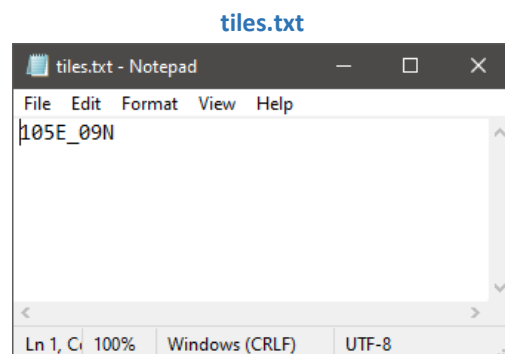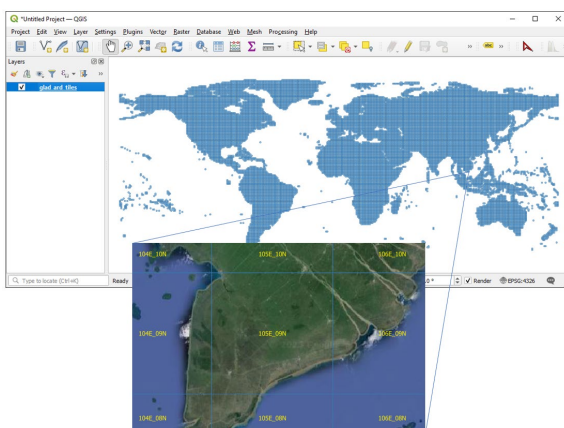
3.3. GLAD ARD 16-day Composites

Landsat image data collected during a 16-day interval are stored as a single ARD GeoTIFF file (ARD composite). There are 23 composites (TIFF files) per year. Each composite (TIFF file) has a unique numeric ID, starting from the first composite of the year 1980. Use the 16-day interval ID table (**C:\GLAD_Tools\Documentation\16d_intervals.xlsx**) to select composites for your analysis.

To create annual **phenological metrics**, **five years of ARD data** is required (present year and 4 preceding years). For example, to create metrics for the year 2022, the range of the required years is 2018 - 2022, which corresponds to the range of 16-day composites from 875 to 989.

Change detection metrics require **four years of ARD data** (present year and 3 preceding years).

The table below provides annual start/end composite dates to simplify annual data selection. The range of composites for phenological and change detection metrics should always correspond to a complete year.

Annual start and end dates of the ARD composites

Year	First composite	Last composite	Year	First composite	Last composite
1997	392	414	2010	691	713
1998	415	437	2011	714	736
1999	438	460	2012	737	759
2000	461	483	2014	783	805
2001	484	506	2015	806	828
2002	507	529	2016	829	851
2003	530	552	2017	852	874
2004	553	575	2018	875	897
2005	576	598	2019	898	920
2006	599	621	2020	921	943
2007	622	644	2021	944	966
2008	645	667	2022	967	989
2009	668	690	2023	990	1012

3.4. SRTM Data

The SRTM data (extracted from NASA product SRTMGL1v003) includes **elevation**, **slope**, and **aspect** at Landsat pixel resolution. The data is used as inputs to most classification and change detection models.

3.5. ARD and SRTM Data Downloading

The data download commands are illustrated by the **C:\GLAD_Tools\Templates\01_Data_Download** template which includes examples of ARD and SRTM download commands. The folder contains **tiles.txt**, the list of tiles to download. The command examples in **commands.txt** provide examples of commands to download the data.

To run the data download command, copy and paste the command from the text file to the CMD prompt:

```
perl C:/GLAD_Tools/download_ARD.pl tiles.txt 875 989 C:/GLAD_Workspace/ARD
```

This command will download ARD data for the years 2018-2022 to the local storage. You need to wait until the **Download complete** message. In case the software outputs the **Download errors. Check log.txt** message, user needs to check the log file. If none of the images were downloaded, check your network settings. If only some of the images failed to download, re-run the same command. The ARD download tool checks images that exist on disk and only gets new, missing, or corrupted data.

To download the SRTM data, run the following command:

```
perl C:/GLAD_Tools/download_SRTM.pl tiles.txt C:/GLAD_Workspace/DEM
```

The SRTM data is the same for all years, so it should only be downloaded once for each tile.

4. Multitemporal Metrics

Metric processing is fully automated and requires minimal user input. The following workflow is used to generate a metric set:

- Download all required 16-day composites.
- Make a list of ARD tiles to process (single column, tile names only): **tiles.txt**.
- Make sure you have enough space on the disk. Each annual phenological metrics set occupies 6.5 GB and change detection metrics set occupies 12 GB.
- Make a parameter file. The parameter file is a text file located in the project folder.
- Run the automated tool to create the metric set.

4.1. Phenological Metrics

The annual phenological metrics serve as source data for land cover, land use, and vegetation structure mapping models. The complete set of phenological metrics is explained in the **C:\GLAD_Tools\Documentation\Metrics_pheno.xlsx** supplementary table.

The workspace template **C:\GLAD_Tools\Templates\02_Metrics** provides examples of commands (**commands.txt**) and parameter files (**metrics_pheno_2022.txt**) for metric processing. To create metrics, follow these steps:

1. Copy the template folder to your workspace (e.g., **C:\GLAD_Workspace\02_Metrics**).

2. Check the parameter file (**metrics_pheno_2022.txt**). The parameter file defines all variables for the metrics generation process. The file should have the following structure:

mettype=pheno	Metric type (**pheno**)
tilelist=tiles.txt	Name of the ARD tile list file
year=2022	Target year
input=C:/GLAD_Workspace/ARD	Input ARD 16-day data folder (should exist and contain ARD files)
output=C:/GLAD_Workspace/Metrics_Pheno_2022	Output folder (will be created if new)
threads=1	Number of parallel processes *
gapfill=4	Number of years to use for gap-filling (values 0 … 4). The default value is 4. Required only for metric types pheno and composite**

* The number of parallel processes should be set to 1 unless a computer has a multi-core processor and available RAM is suitable for several processes simultaneously.

**Gap-filling is a process of filling missing 16-day composite data with clear-sky data from a preceding year. By default, four preceding years are used to fill the gap. The value 0-4 defines how many preceding years will be used. Value 0 means that gap-filling is not performed and only data from the current year is used.

3. Open **CMD** in the **02_Metrics** folder.

4. Use the following command to calculate phenological metrics:

```
perl C:/GLAD_Tools/build_metrics.pl metrics_pheno_2022.txt
```

Metric processing takes time, you need to wait until the **Process complete** message.

Phenological metrics

Annual spectral reflectance/index value distribution metrics

Spectral Bands

Blue (482 nm) [blue]
Green (561 nm) [green]
Red (654 nm) [red]
Near Infrared (864 nm) [nir]
Shortwave Infrared 1 (1609 nm) [swri1]
Shortwave Infrared 2 (2201 nm) [swri2]

Derived Indices

(NIR-Green)/(NIR+Green) [GN]
(NIR-Red)/(NIR+Red) [RN]
(NIR-SWIR1)/(NIR+SWIR1) [S1N]
(NIR-SWIR2)/(NIR+SWIR2) [S2N]
(SWIR1-SWIR2)/(SWIR1+SWIR2) [S1S2]
Spectral variability index** [SVVI]
Tasseled Cap Greenness*** [TCG]

Distribution statistics

Minimum [min]
Maximum [max]
Median [median]
Average between min and Q1 [avmin25]
Average between Q3 and max [av75max]
Average between Q1 and Q3 [av2575]
Average of all values [avminmax]
Standard Deviation [sd]
Total Absolute Difference [absdif]

Amplitudes*

max - min
av75max - avmin25
max - median

Annual seasonality metrics

Spectral Bands

Blue [blue]
Green [green]
Red [red]
Near Infrared [nir]
Shortwave Infrared 1 [swri1]
Shortwave Infrared 2 [swri2]

Distribution methods
(observation dates distributed by the following corresponding inputs)

(NIR-Red)/(NIR+Red) [RN]
(NIR-SWIR2)/(NIR+SWIR2) [S2N]
Brightness temperature [LST]

Distribution statistics

Minimum [min]
Maximum [max]
Average between min and Q1 [avmin25]
Average between Q3 and max [av75max]

Amplitudes*

max - min
av75max - avmin25

Phenology metrics (based on the annual NDVI time series)

Start of season value [RNph_sos]
End of season value [RNph_eos]
Start of season slope [RNph_sos_slope]
End of season slope [RNph_eos_slope]
Start of season amplitude [RNph_sos_amp]
End of season amplitude [RNph_eos_amp]
Growing season average [RNph_ave]
Growing season total [RNph_sum]

Technical metrics

Number of observations used [TEC_count]
Percent water observations [TEC_prcwater]
Data quality flag [TEC_pf]

*Amplitude metrics are not recorded as files.

** See https://www.sciencedirect.com/science/article/pii/S0034425716302693

*** Tasseled Cap coefficients are from
https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0147121#pone.0147121.s002

File naming convention:

2022_GN_av2575.tif

Year   Band or   Statistic
         index

4.2. Change Detection Metrics

The annual change detection metrics are designed to highlight inter-annual changes in spectral reflectance while reducing false detections due to reflectance fluctuations and inconsistent clear-sky observations availability. These metrics are primarily designed to map changes in tree canopy cover, and they may not be suitable for mapping changes in other land cover types. The complete set of phenological metrics is explained in the **C:\GLAD_Tools\Documentation\Metrics_change_detection.xlsx** supplementary table.

The workspace template **C:\GLAD_Tools\Templates\02_Metrics** provides examples of commands (**commands.txt**) and parameter files (**metrics_change_2022.txt**) for metric processing. To process the metrics set, follow these steps:

1. Copy the template folder to your workspace (e.g., **C:\GLAD_Workspace\02_Metrics**).

2. Check the parameter file (**metrics_change_2022.txt**). The parameter file defines all variables for the metrics generation process. The file should have the following structure:

mettype=change	Metric type (**change**)
tilelist=tiles.txt	Name of the ARD tile list file
year=2022	Target year
input= C:/GLAD_Workspace/ARD	Input ARD 16-day data folder (should exist and contain ARD files)
output=C:/GLAD_Workspace/Metrics_Change_2022	Output folder (will be created if new)
threads=1	Number of parallel processes *
gapfill=4	Number of years to use for gap-filling (values 0 … 4). The default value is 4. Required only for metric types pheno and composite**

* The number of parallel processes should be set to 1 unless a computer has a multi-core processor and available RAM is suitable for several processes simultaneously.

**the gap-filling parameter is not used for this type of metrics.

3. Open **CMD** in the **02_Metrics** folder.

4. Use the following command to calculate phenological metrics:

```
perl C:/GLAD_Tools/build_metrics.pl metrics_change_2022.txt
```

Metric processing takes time, you need to wait until the **Process complete** message.

Change Detection Metrics

Spectral data and indices

Statistics

Spectral Bands

Blue [blue]
Green [green]
Red [red]
Near Infrared [nir]
Shortwave Infrared 1 [swri1]
Shortwave Infrared 2 [swri2]

Derived Indices

(NIR-Red)/(NIR+Red) [NDVI]
(NIR-SWIR1)/(NIR+SWIR1) [NDWI]
(SWIR1-SWIR2)/(SWIR1+SWIR2) [SWSW]

Interannual metrics for the corresponding [c] and preceding [p] years

Distribution statistics

Minimum [min]
Maximum [max]
Second lowest value [smin]
Second highest value [smax]
Median [median]
Average of all values [ave]
Average of all values except min and max [avett]
Last annual observation [last]

Annual seasonality metrics

Distribution methods

(NIR-Red)/(NIR+Red) [NDVI]
Brightness temperature [LST]

Distribution statistic

Minimum [min]
Maximum [max]
Second lowest value [smin]
Second highest value [smax]
Median [median]

Amplitudes*

Difference between all metrics extracted from corresponding and preceding years.

Seasonal reflectance change (calculated from the per-16-day difference time series)

Minimum [min]
Maximum [max]
Second lowest value [smin]
Second highest value [smax]
Average of all values [ave]
Average of all values except min and max [avett]
Value after minimum [amin]
Value after maximum [amax]

Interval time series

Slope of linear regression of band/index value vs. observation date [reg]
Standard deviation of the band/index value [sd]

Technical metrics

Number of observations used [count]
Percent water observations [prcwater]
Percent land observations [prcland]
Data usability code [code]
Data quality flag [pf]

*Amplitude metrics are not recorded as files.

4.3. Annual Clear Sky Composite

GLAD Tools provides an instrument to create cloud-free composites for data visualization. The annual composite represents a subset of phenological metrics that are designed for data visualization only, and not suitable for image classification. The composite includes spectral band reflectance average between Q1 and Q3 (av2575) metrics and technical metrics from the phenological metric set (**section 4.1**).

The workspace template **C:\GLAD_Tools\Templates\02_Metrics** provides examples of commands (**commands.txt**) and parameter files (**metrics_composite_2022.txt**) for metric download. To calculate the metrics set, follow steps provided in **section 4.1** with the following changes:

- The metric type in the parameter file is **composite**.
- The command is `perl C:/GLAD_Tools/build_metrics.pl metrics_composite_2022.txt`

5. Multitemporal Metrics Visualization

The multitemporal metrics are stored as 1x1 degree tiles. To visualize data for a large region, tiles must be mosaicked together. OSGeo4W and GDAL Tools provide several solutions to mosaic the data as a Virtual Raster Table (VRT) and GeoTIFF files (TIF).

5.1. Phenological Metrics Mosaicking

The template **C:\GLAD_Tools\Templates\03_Pheno_Mosaic** provides a set of parameter file examples to calculate a variety of mosaics from a phenological metric set. To implement the mosaicking, copy this template to your workspace. All image mosaics created using these codes will be stored in a single subfolder. The VRT files are linked to the original phenological metrics folders and will not work if the metrics folders are removed or renamed.

Before running the mosaicking tool, check the following preconditions:

- The metrics must be created before image mosaicking. Image mosaicking directly from the ARD 16-day data is not supported.
- The **C:\GLAD_Tools\dependencies.txt** file has a correct link to the **OSGeo4w.bat** software (see GLAD Tools installation, **section 1.3**).

To specify the mosaic extent, GLAD Tools uses the list of tiles **tiles.txt** (same format as for the ARD download and metric generation). The parameter file is required for the tool. The file should have the following structure:

source=D:/Metrics_pheno_2022	Source folder
list=tiles.txt	The name of the tile list
year=2022	Year
outname=median654	Output name
bands=swir1_av2575, red_av2575, nir_av2575	List of metrics to aggregate (comma separated)

To create a VRT mosaic, open the **CMD** terminal and use the following command:

```
perl C:/GLAD_Tools/mosaic_vrt.pl mosaic_median654_2022.txt
```

To create a **GeoTIFF** output file, use the following command:

```
perl C:/GLAD_Tools/mosaic_tif.pl mosaic_median654_2022.txt
```

The full description of each of the mosaic types is provided in the User Manual. The following table summarizes the most common mosaic types:

Parameter file name	Mosaic description and application
mosaic_median567_2022.txt	NIR-SWIR1-SWIR2 median annual composite. Typically used to map vegetation types.
mosaic_median654_2022.txt	SWIR1-NIR-Red median annual composite. Typically used to map land cover types (pseudo-natural color composite)
mosaic_lowNDVI567_2022.txt	NIR-SWIR1-SWIR2 composite of mean reflectance for observations between the minimum and 25th percentile of NDVI value within the year. Typically used to map inundated areas and crop types.
mosaic_highNDVI567_2022.txt	NIR-SWIR1-SWIR2 composite of mean reflectance for observations between the maximum and 75th percentile of NDVI value within the year. Typically used to map permanent water and crop types.
mosaic_amplitude_2022.txt	The amplitude of selected band values and indices (NIR-SWIR1-NDVI). Typically used to map crop types and tree cover.
mosaic_index_2022.txt	Composite of selected vegetation indices. Typically used to map tree canopy cover

Parameter files **mosaic_median567_2022.txt** and **mosaic_median654_2022.txt** are suitable for visualization of the annual cloud-free composite as well (the user needs to change the metric input folder).

5.2. Change Detection Metrics Mosaicking

Change detection metrics may be visualized using two approaches. First, the reflectance metrics of current and preceding years can be visualized separately and compared in the QGIS. Second, a user can visualize per-16-day composite (seasonal) changes in spectral reflectance simplifies.

A set of example parameter files in **C:\GLAD_Tools\Templates\04_Change_Mosaic** provides examples for different visualization. The application of the mosaicking tool is the same as for the phenological metrics (**section 5.1**). The full description of each of the mosaic types is provided in the User Manual.

Parameter file name	Mosaic description and application
mosaic_c_avett654_2022.txt	SWIR1-NIR-Red median annual composite of the current year.
mosaic_c_last654_2022.txt	SWIR1-NIR-Red composite of the last cloud-free observation of the current year.
mosaic_p_avett654_2022.txt	SWIR1-NIR-Red median annual composite of the preceding year.
mosaic_p_last654_2022.txt	SWIR1-NIR-Red composite of the last cloud-free observation of the preceding year.
mosaic_change1_2022.txt	A composite that shows the SWIR1 band difference between the current year (highest reflectance) and the preceding year (average reflectance).
mosaic_change2_2022.txt	A composite that displays the highest seasonal change of the SWIR1 band and the average seasonal change of the NIR/SWIR1 band ratio.

6. Land Cover Classification

The GLAD Tools employ a **machine learning decision tree algorithm** (also called Classification and Regression Trees, or CART) for land cover classification. This algorithm estimates the per-pixel probability of a target land cover class presence using decision rules in the multispectral/multitemporal domain of ARD-based metrics. To reduce the model overfitting, we implement **bootstrap aggregation (bagging)**, generating an ensemble of decision trees for each classification.

The decision tree ensemble is generated automatically from a set of training pixels. Each training pixel has a label that indicates if it belongs to the land cover class of interest (**target**) or other land cover types (**background**). During classification, we only map one land cover type. Multiple types can be mapped using hierarchical classification (see **section 6.5**).

The population of training objects (target and background classes) is defined by vector polygons that are manually mapped by an expert. When the classification tree model is implemented for an entire image, it predicts the land cover class probability for every pixel. The classification result represents the probability of the pixel to represent the target land cover class in percent.

The following **steps must be completed before** starting the classification:

1. The ARD data for the target year of classification should be downloaded. To produce phenological metrics, each year user should download ARD data for the five-year interval: target year and four preceding years (**sections 3.3 and 3.5**).

2. DEM data should be downloaded (**section 3.5**).

3. The phenological metrics for the target year should be created (**section 4.1**)

4. A set of image mosaics from phenological metrics should be created (**section 5.1**).

6.1. Classification workspace

The template **C:\GLAD_Tools\Templates\05_Land_cover_classification** shows the example workspace to implement land cover classification. Each classification (each land cover class) should have its own workspace. The main components of the classification workspace (after implementing classification) are the following:
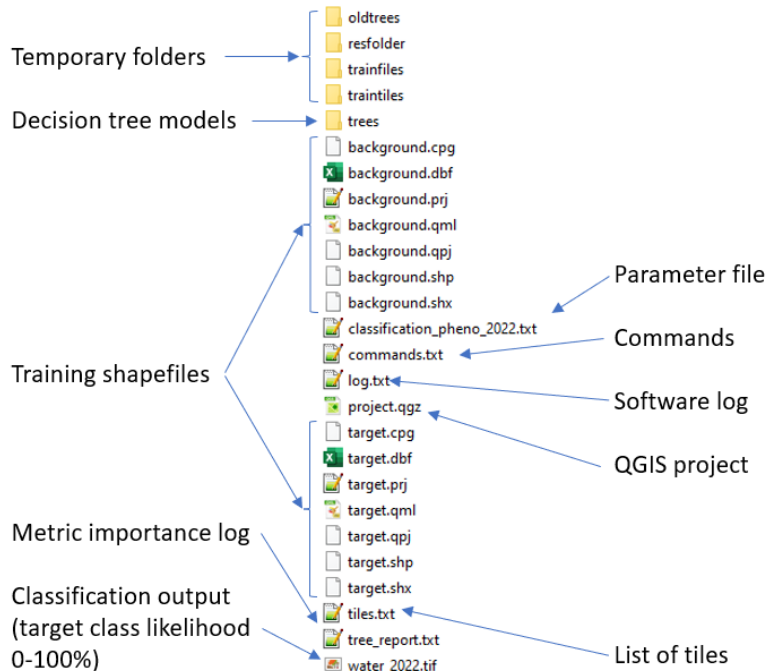
A new classification workspace should have the following elements:

- List of tiles (**tiles.txt**).
- Two training files (**target.shp** and **background.shp**).
- Classification parameter file (**classification_pheno_2022.txt**).
- Optional: the list of commands (**commands.txt**).

To start a new classification, copy template **C:\GLAD_Tools\Templates\05_Land_cover_classification** to your workspace.

6.2. Collecting Training Data
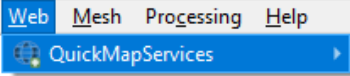
Training data represent two polygon shapefiles, one with areas of land cover class presence (**target**), and the other with areas of land cover class absence (**background**). Both shapefiles should be in the same coordinate system as phenological metrics (*+proj=longlat +ellps=WGS84 +datum=WGS84 +no_defs*). The classification tool uses only the object shape data, with all attributes ignored. The shapefiles may contain overlapping polygons. The correct topology is not required as long as data can be correctly rasterized. The polygons in the **target** and **background** shapefiles may overlap. In case of overlap, the area under the **target** class polygons will be erased from the **background** layer (the **target** class has a higher priority during rasterization).

**To collect training data**, follow the routine described below:

- Open QGIS (new project).
- Add mosaics of selected metrics. Optionally: load the Google Satellite layer using the QMS plugin.
- Load target.shp and background.shp files. Put the target layer onto the top of the background layer in the Layer Panel.
- Start editing (Toggle Editing button) for both shapefiles.

- Use the "Add Polygon" or "Freehand Drawing" tools to add training samples. Avoid creating large training polygons. Distribute samples over the entire area of the image.
- Save layers and project (periodically)

Using QGIS Tools and plugins

|	
	Load the QMS Google Satellite layer: - Open "Search QickMapServices" - Use the "Google" keyword to search for the layer. Add the "Google Satellite" layer.
	Google Satellite mosaic is an important reference data, but it may show images from a different date than the Landsat mosaic. The user should check if the Google images represent the same date as Landsat metrics. To do this, use the "Send2GE" tool to check image dates in Google Desktop.
	Enable the "Raster Toolbar". Use image enhancement tool to improve the Landsat image visualization.
	To start adding training areas, use the "Toggle Editing" button for both the target and background layers.
	To add a new polygon, use the "Add Polygon Feature" button.
	When the "Add Polygon" tool is activated, use the "Digitizing mode" selector to choose "Stream Digitizing Mode".
	Save layers and the project file at least every 10 minutes to avoid losing your work. Use the "Save for All Layers" button for convenience.

For the new QGIS installation, we suggest updating the following program options:
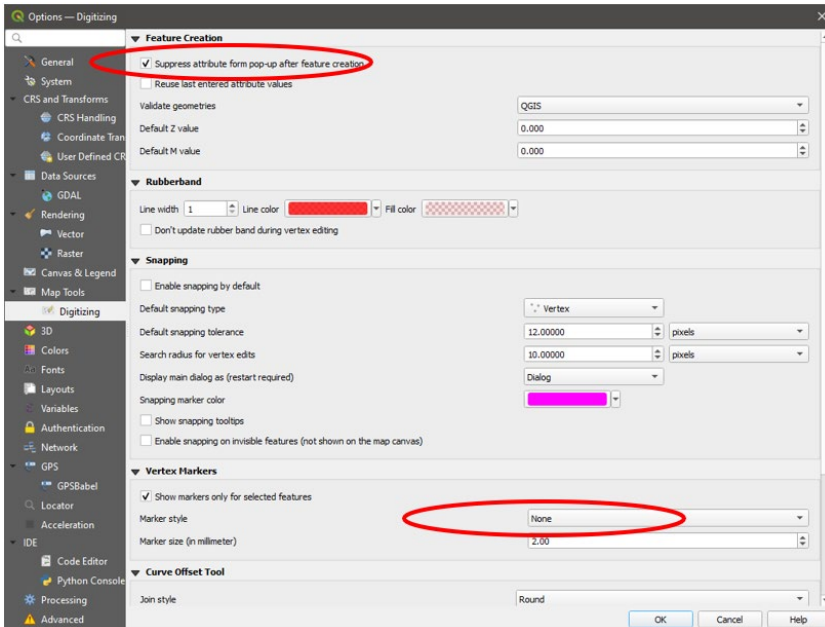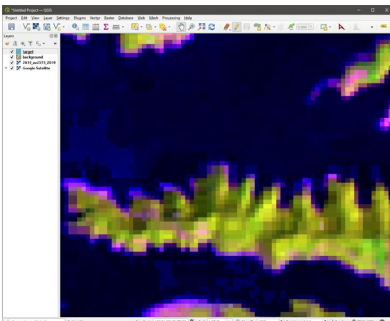

- Open from QGIS Manu: Setting → Options.
- Go to the "Digitizing" tab.
- Check the "Suppress attribute from pop-up after feature creation" box.
- Set Vertex marker style to "None."

Example of the trainign data collection workflow

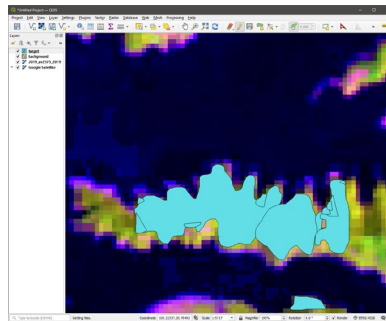QGIS project with training shapefiles, image mosaic, and the Google Maps layer (through QMS).
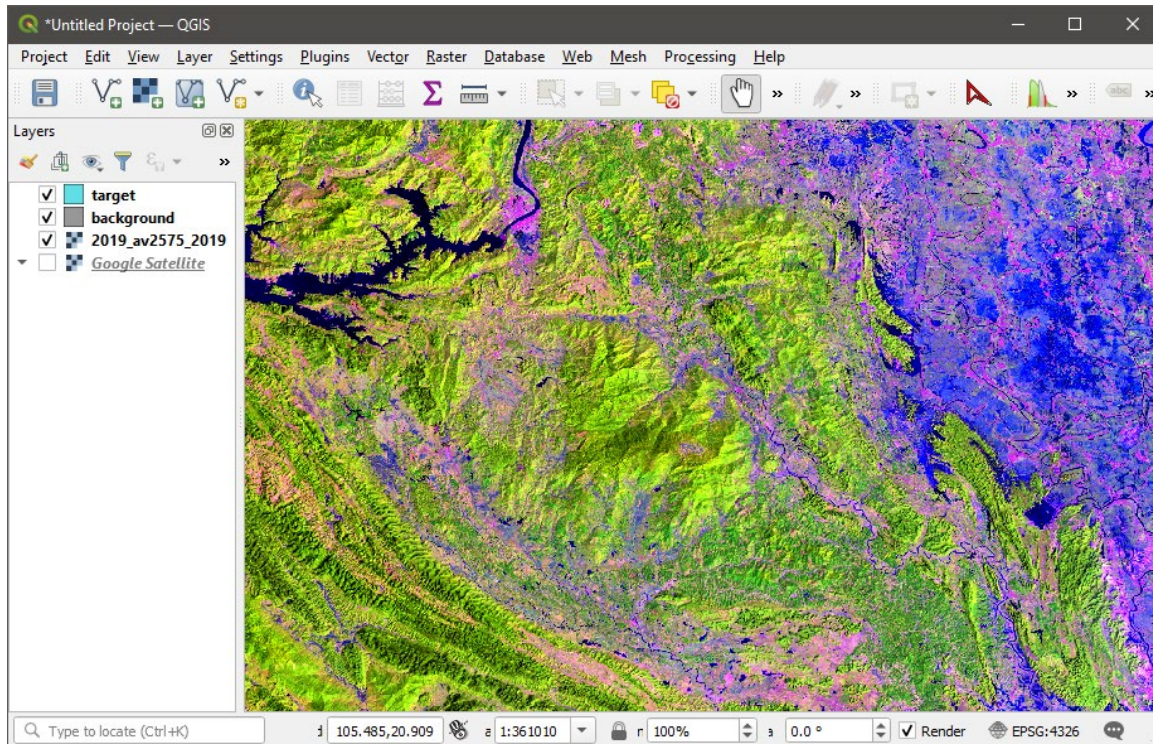
Example of land cover class drawing using image mosaic


1. Zoom into the area (1:5,000 or similar).

Only draw training using the zoom level where you can see individual pixels of Landsat data.

1. Drawing the **target** class presence (forest).

The target class represents the land cover/use theme of interest (e.g., forest).

The class can be mapped by any number of overlapping polygons. All pixels of the target class within the training area must be marked.
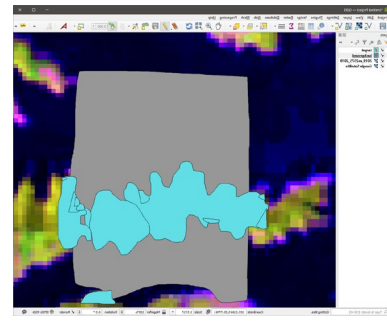
2. Drawing the class absence, or **background** (non-forest).

The background class represents all other land cover types.

The polygons of the background layer can overlap with the polygons of the target class layer.

Good practice guidance for training sample allocation

It is important to distribute training areas throughout the entire mapping region. We recommend using the QGIS **grid tool** to guide training allocation. The grid can be displayed over the metric mosaics using the QGIS menu option **View\Decorations\Grid**. If your mapping area is smaller than 200x200 km, a grid spaced by 0.2 degrees is optimal to guide training data. In the image below, the grid overlaps a single GLAD ARD tile. The training samples should be collected within each grid cell. The placement of a sample inside a grid cell depends on the land cover class configuration.
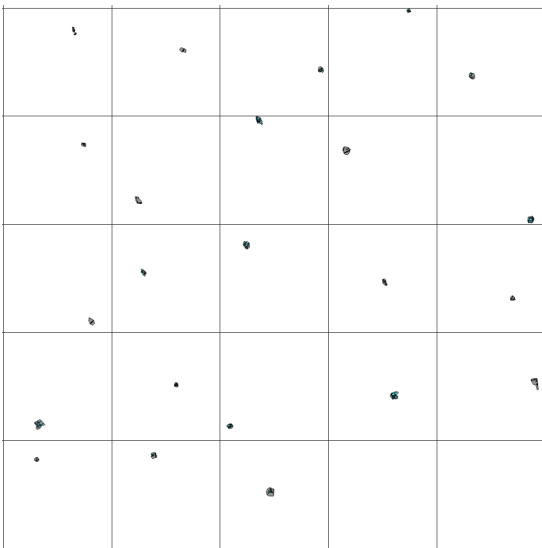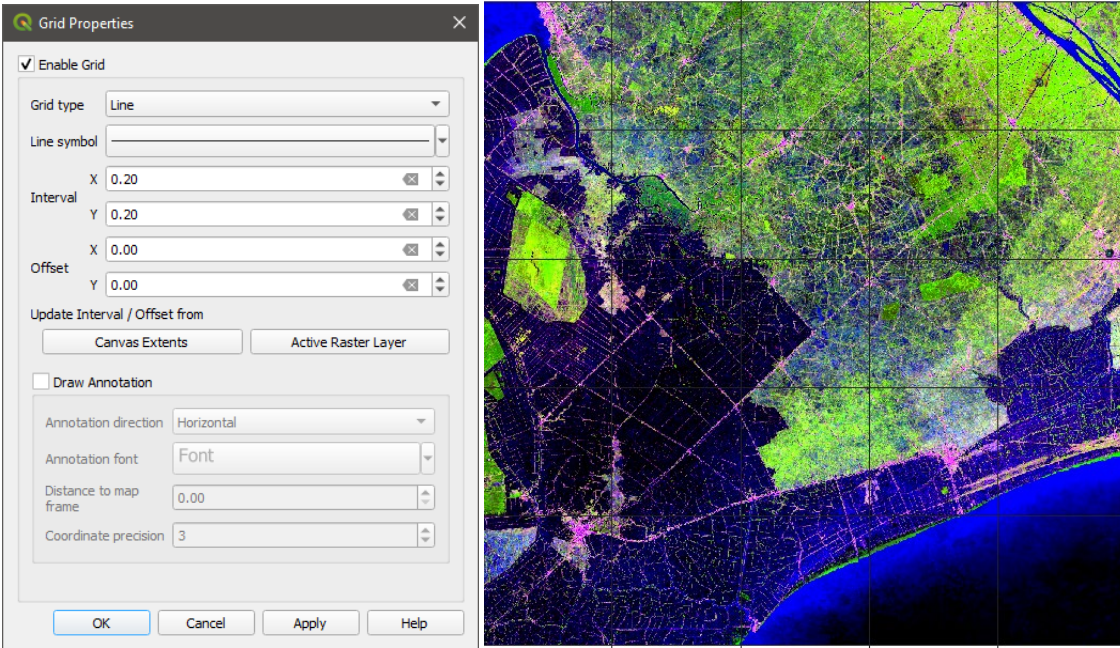
The build-up land cover class training shapefiles. The training distribution is guided by the 0.2x0.2 degrees grid within the area of analysis.

6.3. Applying Classification

To apply classification, follow the routine described below:

- Save all edits and close the QGIS project. **It is important to close the project while running the classification!**
- Check and edit if needed the classification parameter file.

Classification parameter file structure (classification_pheno_2022.txt)

mettype=pheno	Metric type
metrics=C:/GLAD_Workspace/Metrics_Pheno_2022	Multi-temporal metrics source folder
dem= C:/GLAD_Workspace/DEM	Topography metrics source folder
year=2022	Year of multi-temporal metrics
target_shp=target.shp	Target class shapefile name
bkgr_shp=background.shp	Background class shapefile name
tilelist=tiles.txt	Name of the tile list file
outname=water_2022	Output file name
mask=none	Mask file name (none – no mask)
maxtrees=25	Number of trees (odd number in the range 1-25)
sampling=10	Sampling rate (percent training data extracted for each tree)
mindev=0.0001	Tree pruning rule
threads=1	Number of parallel processes
treethreads=25	Number of parallel processes for a tree model
reuse_model=none	Use existing trees (none – create a new model)

You may modify parameters depending on the computer capacity, training size, etc. Specifically:
- Increasing the *maxtrees* parameter will slow classification but improve model generalization.
- Increasing the *mindev* will reduce tree complexity, while reducing it will increase tree complexity.
- Reduce *sampling* if sample areas are too large. Increase it if the "maxtrees" parameter is reduced.
- Reduce *threads* and *treetherads* parameters for a low-capacity computer (minimal value 1)

Parameter file template
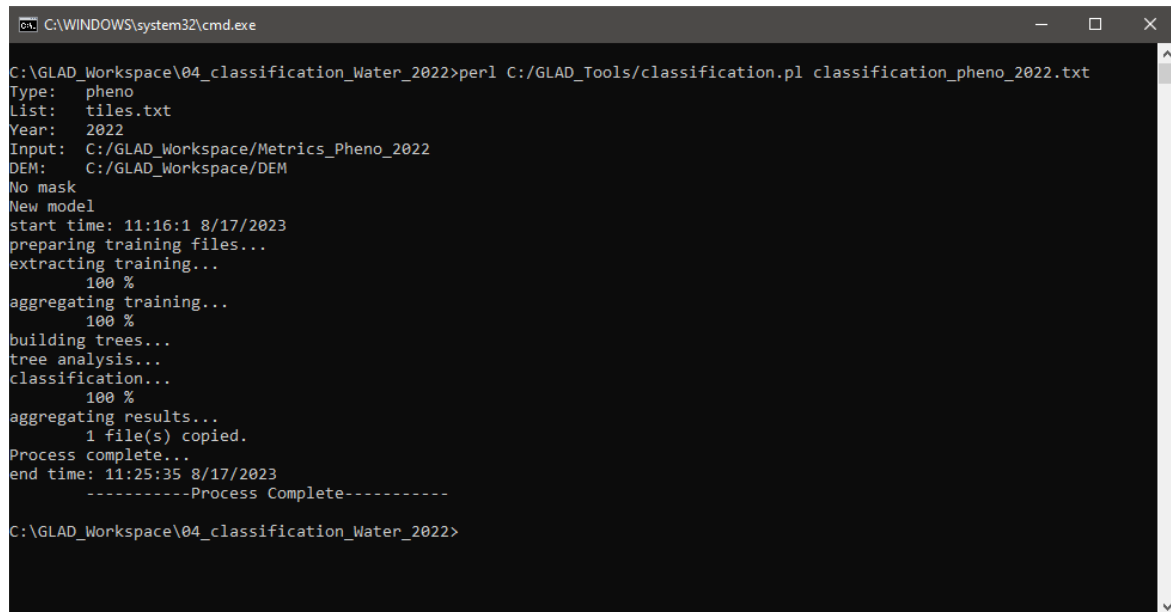**C:\GLAD_Tools\Templates\05_Land_cover_classification\classification_pheno_2022.txt**

- Open CMD in the classification project folder
- Use the following command to implement classification:

```
perl C:/GLAD_Tools/classification.pl classification_pheno_2022.txt
```

- Classification will take time. Wait for the **Process Complete** sign that indicates that the classification was run successfully. In case of an error message, check the software installation and parameter file.

```
C:\WINDOWS\system32\cmd.exe

C:\GLAD_Workspace\04_classification_Water_2022>perl C:/GLAD_Tools/classification.pl classification_pheno_2022.txt
Type:    pheno
List:    tiles.txt
Year:    2022
Input:   C:/GLAD_Workspace/Metrics_Pheno_2022
DEM:     C:/GLAD_Workspace/DEM
No mask
New model
start time: 11:16:1 8/17/2023
preparing training files...
extracting training...
        100 %
aggregating training...
        100 %
building trees...
tree analysis...
classification...
        100 %
aggregating results...
        1 file(s) copied.
Process complete...
end time: 11:25:35 8/17/2023
-----------Process Complete-----------

C:\GLAD_Workspace\04_classification_Water_2022>
```

6.4. Classification Results

The classification output is stored as a raster file (LZW-compressed GeoTIFF). All tiles are mosaicked together. The pixel value is in the range of 0 – 100 and represents the target class likelihood. The commonly used threshold to identify the target class is 50% (values 0-49 represent the background class, and 50-100 - the target class). However, the threshold may be adjusted if needed. The likelihood should not be treated as spectral similarity as it depends on the training population. The class likelihood does not represent the percentage of a target class within a pixel.

To visualize the classification results, open the saved QGIS project and add the classification output layer (e.g., **water_2022.tif**). Classification results represent a single mosaicked 8-bit TIFF file. Set layer symbology to "Singleband pseudocolor" and leave only one value in the legend.
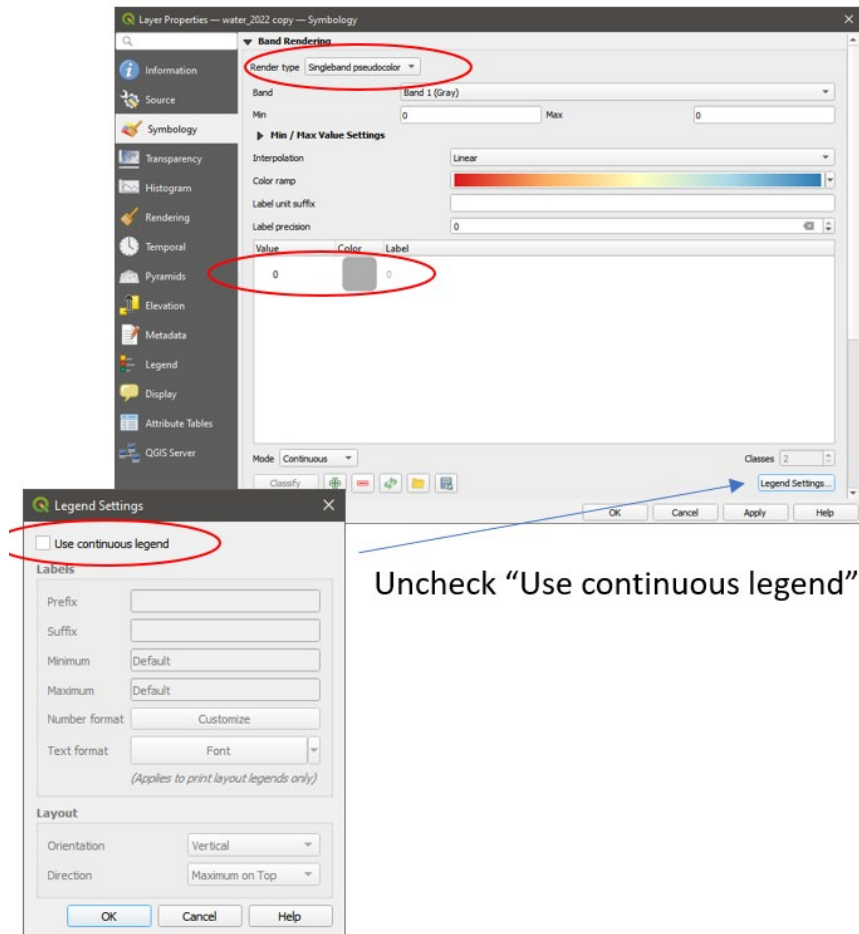
To visualize the target class extent, set layer **transparency** to the **range 0-49**. This visualization displays the class presence and is used to check for omission errors.
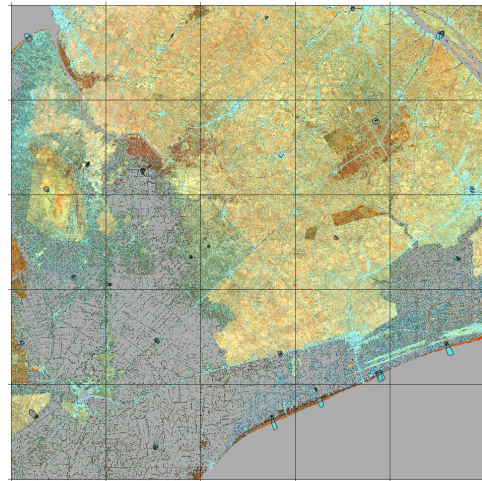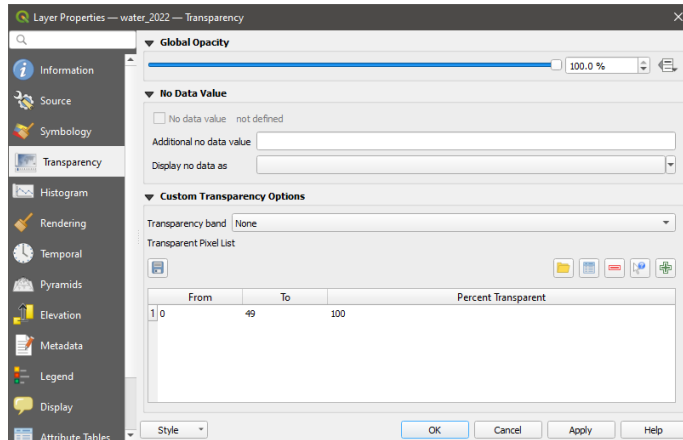
To visualize target class absence, set layer **transparency** to the **range 50-100**. This visualization is used to check for commission errors.
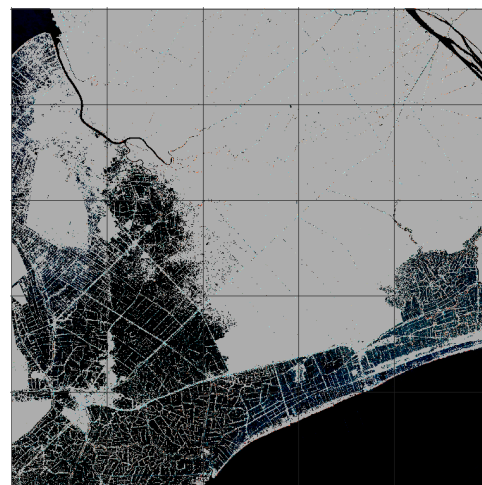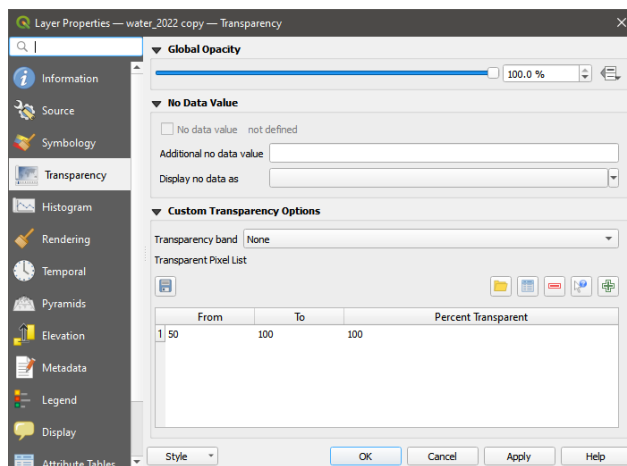
We recommend keeping both versions of the result's visualization in the QGIS project to simplify map checking and adding training data.

6.5. Iterating Classification

The machine learning classification results may contain errors due to insufficient training data. The model can be improved through the iterative process of adding training data, known as **active learning**. The active learning method consists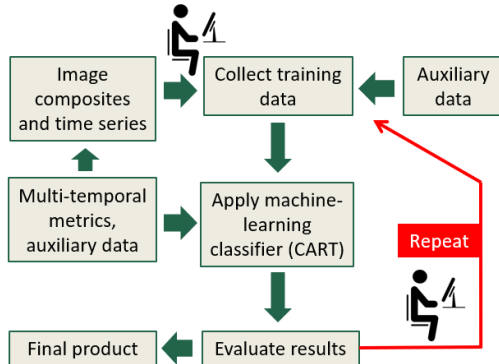 of iterations of model construction, application, evaluation of results, and adding new training data until the desired map quality is achieved.

Active learning focuses on the interaction between the analyst and the classifier. After applying the classification, the results are checked by the analyst. The analyst assigns correct land cover labels to pixels classified incorrectly and adds these training pixels to the training set to improve the model. In this way, the model is optimized on well-chosen difficult examples, maximizing its generalization capabilities.

Below we describe an efficient workflow for map iteration and improvement.

1. Open the saved QGIS project and load the classification results (results will be updated automatically only if the QGIS project is closed during the classification application).

2. Check the map over the entire area of analysis and **add training** within omission and commission errors. We recommend using grid (**section 6.2**) to check the entire image at a high zoom level. Use the grid to navigate and check image row by row.
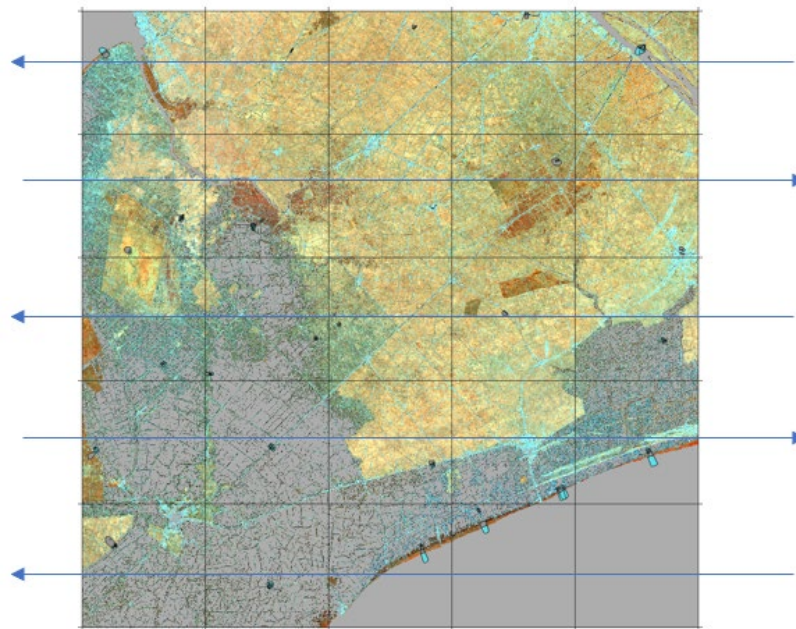
3. Save all edits and **close** the QGIS project.

4. Run the classification code in CMD (section 6.3). There is no need to close the CMD after each iteration. All results and model files will be overwritten.

5. Repeat the process until the map is of sufficient quality.

7. Hierarchical Classification

Hierarchical classification allows mapping multiple land cover classes sequentially. Each new class is mapped outside classes already included in the map. Another application of this tool is the reduction of the area of analysis, e.g., the classification of a land cover type within administrative boundaries.

The hierarchical classification is controlled by a **classification mask**. The mask is represented by a GeoTIFF file (8-bit) which has values 0 and 1. The classification is applied only within pixels with value 1.

The hierarchical classification template workspace is in **C:\GLAD_Tools\Templates\06_Hierachical_classification**. The workspace and commands are like the ones used for simple land cover classification (**section 6**), with several changes.

Before the classification starts, the user must create the classification mask. The classification mask is based on earlier results. In the template, we assume the user completed water extent classification and the resulting water map is **C:\GLAD_Workspace\01_Water_2022\water_2022.tif**. The mask is created using the Image Modeler tool that requires the model parameter file, **C:\GLAD_Tools\Templates\06_Hierachical_classification\mask_model.txt**. The **model parameter file** has the following content:

```
INPUT
%1=<path to earlier classification results>\<classification output name>.tif
END
MODEL
if (%1<50) {%0=1;}
END
```

To create the mask, open the CMD in the workspace and run the command:

```
perl C:/GLAD_Tools/raster_model.pl mask_model.txt mask.tif
```

The mask can be used during image analysis and training data collection. First, load the **mask.tif** to the QGIS project, set the legend to "Paletted/Unique values", and add only the value "0" as a black color.
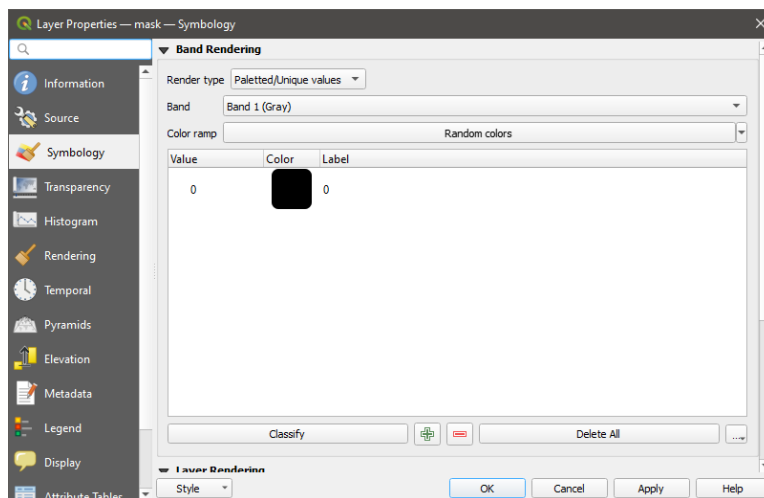
Training data collection is performed the same way as for the simple classification (**section 6.2**). The difference is that areas within mask value 0 will be ignored for both training data collection and classification. You may draw both target and background polygons over the masked areas (areas with **mask.tif** values of "0"). These pixels will be ignored during classification.

To use the mask for classification, change the classification parameter file: replace "**mask=none**" with "**mask=mask.tif**". See an example of the parameter file below. The training data collection and classification will be restricted to the areas with the value "1" in the **mask.tif** file.

Example	
Changes to the parameter file from section 6.3: **mask=mask.tif**	classification_pheno_2022.txt - Notepad File Edit Format View Help mettype=pheno metrics=C:/GLAD_Workspace/Metrics_Pheno_2022 dem=C:/GLAD_Workspace/DEM year=2022 target_shp=target.shp bkgr_shp=background.shp tilelist=tiles.txt outname=builtup_2022 mask=mask.tif maxtrees=25 sampling=10 mindev=0.0001 threads=4 treethreads=25 reuse_model=none Ln 1, Col 1 100% Unix (LF) UTF-8

Classification results (target class probability) are provided only within the mask=1 values. Outside the mask, the classification output TIF file has value=0. We recommend keeping the **mask.tif** file in the QGIS project during the classification iteration to avoid confusion between unclassified areas and the background class.

Classification mask (value 0 shown in black) → ☑ **mask** Band 1 (Gray) ▪ 0

Training shape files → ☑ **target** ☑ **background**

Classification output: Target class (transparency 0-49) → ☐ *builtup_2022* Band 1 (Gray) ▪ 0

Background class (transparency 50-100) → ☑ **builtup_2022 copy** Band 1 (Gray) ▪ 0

Landsat mosaic → ☐ *MOSAIC_2022_median654* Band 1 Band 2 Band 3

Google Satellite QMS layer → ☑ **Google Satellite**
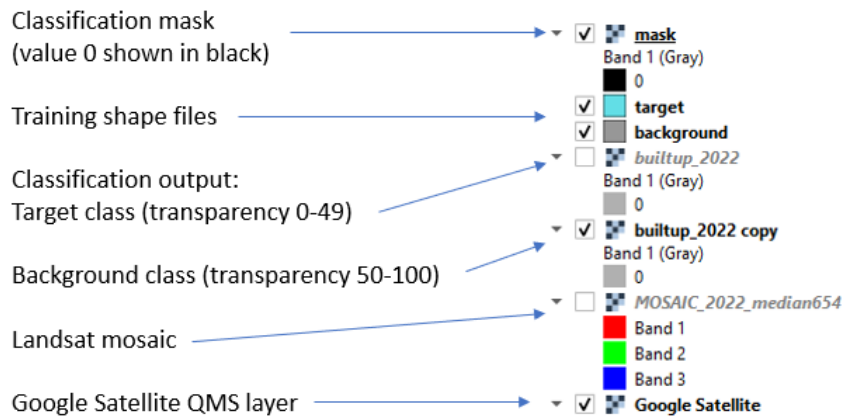
8. Multitemporal Classification
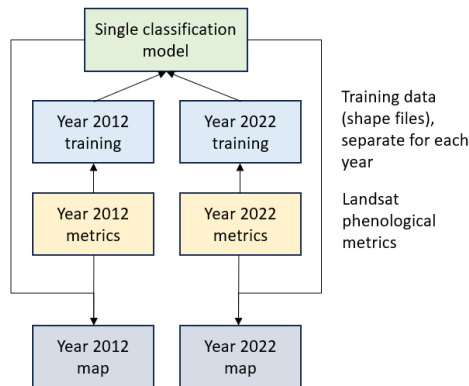
The multitemporal classification tool is designed for calibrating temporally stable land cover and land use mapping models. The multitemporal classification model uses training data from different years to create a single classification model. This way, the classification model is trained to ignore differences between Landsat sensors and inconsistencies in annual data frequency. The number of intervals for model calibration and application is not limited; two to four intervals are recommended. The following example shows the model calibration using two annual intervals.

A workspace template is provided here: **C:\GLAD_Tools\Templates\07_Multitemporal_classification**. This template illustrates a two-year (2012 and 2022) multitemporal classification.

The following **steps must be completed before** starting the multitemporal classification:

1. The ARD data for all target years of classification (2012 and 2022) should be downloaded. To produce multitemporal metrics, each year user should download ARD data for the five-year interval: the target year and four preceding years (**sections 3.3 and 3.5**).

2. DEM data should be downloaded (**section 3.5**).

3. The phenological metrics for both the years 2012 and 2020 should be created (**section 4.1**)

4. A set of image mosaics for both years 2012 and 2012 should be created (**section 5.1**).

8.1. Multitemporal Classification Workspace

To start a new classification, copy template **C:\GLAD_Tools\Templates\07_Multitemporal_classification** to your workspace.

The new classification workspace should include the following elements:

- List of tiles (**tiles.txt**)
- Two sets of training files (**target_year1.shp**, **background_year1.shp, target_year2.shp**, **background_year2.shp**). The **Year1** files used to make training for the year 2012, and the **Year2** for the year 2022. If more years are used for classification, they should have unique names.
- A set of parameter files to implement classification for each year (**classification_pheno_year1.txt** and **classification_pheno_year2.txt**).
- The **classification.bat** file contains commands to execute classification.

The **classification.bat** file simplifies the application of multitemporal classification. The file contains commands to (a) export training data for each year; (b) build a single model; and (c) apply the model to each year.

```
perl C:/GLAD_Tools/multitemporal_class_export.pl classification_pheno_year1.txt
perl C:/GLAD_Tools/multitemporal_class_export.pl classification_pheno_year2.txt
perl C:/GLAD_Tools/multitemporal_class_model.pl 25 25 0.0001
perl C:/GLAD_Tools/multitemporal_class_apply.pl classification_pheno_year1.txt
perl C:/GLAD_Tools/multitemporal_class_apply.pl classification_pheno_year2.txt
```

Extracting training for each year independently
Building a common classification model
Applying classification for each year

C:\GLAD_Tools\Templates\07_Multitemporal_classification folder content (before classification):


8.2. Collecting Training Data

To start the training collection, open a **new QGIS project**. Add to the project the following layers:

1. Training files **target_year1.shp**, **background_year1.shp, target_year2.shp**, **background_year2.shp,** from classification workspace.

The position of shapefiles is important. The "target' class layer should always be above the "background" layer.

2. Image mosaics for both years**.**

3. Use the QickMapServices toolbar to search for "Google" and add the **Google Satellite** layer.
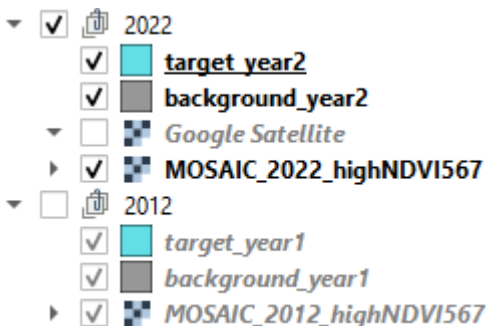
We suggest using groups to organize the layers in the project's TOC (table of contents). Use the "**Group selected**" option in QGIS TOC to group layer for each year and rename groups according to the year (**2012** and **2022**). The Google Satellite layer may be presented only in the 2022 group or outside the groups.

Only one group should be visible at a time during training data collection. Within the group, keep the "target" training layer above the "background" layer to avoid confusion.

To collect training data, visualize the group for the year 2022 to add training sites for the Year2 training files. Then, visualize the group for the year 2012 to collect data for the Year1 training files.

8.3. Applying Multitemporal Classification

Unlike single-year classification, multitemporal classification uses separate tools for sample extraction, model calibration, and model application. For each year, sample extraction and model application are implemented separately and are guided through parameter files. The parameter file structure is the same as described in **section 6.3**. For each year, the parameter files have different source metric sets, different training files, and different output file names (differences between these files are highlighted in color):

classification_pheno_year1.txt	classification_pheno_year2.txt
mettype=pheno	mettype=pheno
metrics=C:/GLAD_Workspace/Metrics_Pheno_2012	metrics=C:/GLAD_Workspace/Metrics_Pheno_2022
dem=C:/GLAD_Workspace/DEM	dem=C:/GLAD_Workspace/DEM
year=2012	year=2022
target_shp=target_year1.shp	target_shp=target_year2.shp
bkgr_shp=background_year1.shp	bkgr_shp=background_year2.shp
tilelist=tiles.txt	tilelist=tiles.txt
outname=water_2012	outname=water_2022
mask=none	mask=none
maxtrees=25	maxtrees=25
sampling=10	sampling=10
mindev=0.0001	mindev=0.0001
threads=1	threads=1
treethreads=25	treethreads=25
reuse_model=none	reuse_model=none

The **classification.bat** file simplifies the application of multitemporal classification (**section 8.1**.)

To implement multitemporal classification, follow these steps:

1. Save and stop editing all training files. Save and close the QGIS project.

2. Open CMD in the classification folder and use the command:

```
classification.bat
```

Wait for the process to complete, this will take time. The classification is complete when the new CMD prompt appears:

```
C:\GLAD_Workspace\06_classification_Water_2012-2022>
```

3. The classification outputs are stored in separate files for each year, e.g., **water_2012.tif** and **water_2022.tif**.

8.4. Classification Results and Iteration

To check the classification results, open the saved QGIS project and add results (water_2012.tif and water_2022.tif). Put each file in the corresponding group (**2012** and **2022**). Follow the instructions in **section 6.4**. to apply the legend to the output layers. I suggest keeping two copies of each results file: one for the target class visualization, and another for the background class.
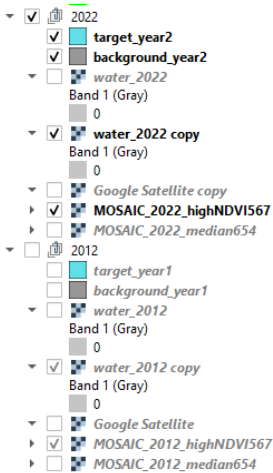
First, check the classification results for the year 2022. Make visible only data from the "**2022**" group. Check both water class commission and omission errors, add training if needed, and save the training files.

Next, check the results for the year 2012. To do this, make only the "**2012**" group visible, and check the results for the year 2012 using the corresponding Landsat image mosaic. Using the Google Satellites layer for the early year (before 2020) Is not recommended.

After adding training data, save edits and close the QGIS project. Run the **classiffaction.bat** batch file again to update the results.

The QGIS allows users to quickly check the target land cover class changes between years. To visualize change, follow these steps:

1. Make a copy of the change classification layers (water_2012 and water_2022) and merge them in a new group (named "**change**").

2. For the **water_2012** layer, use the "Singleband pseudocolor" symbology and leave only one value in the legend. Make the color of this class green. Set "Transparency" to the interval 0-49.

3. Use the same legend (single color, transparency 0-49) for the **water_2022** layer. Change the color to red and set "Blending mode" to "Darken".

4. Display results for both years 2012 and 2022. The resulting map shows stable water extent in black, water gain in red, and water reduction in green.

Water in both 2012 and 2022

Water 2012 only

Water 2022 only
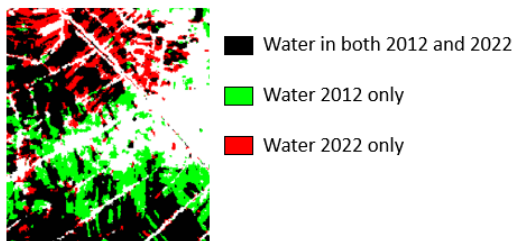
9. Change Detection

Change detection metrics and classification algorithms are designed specifically to map the abrupt loss of tree and shrub canopy cover, such as logging, windfalls, wildfires, and others.

9.1. Change Detection Workspace

The template **C:\GLAD_Tools\Templates\ 08_Change_detection** shows the example workspace to implement forest change detection. Each classification (each year of change detection) should have its own workspace. The main components of the classification workspace are the same as for the Land Cover Classification (**section 6.1.**)

9.2. Collecting Training Data

Similarly to land cover classification (**section 6.1**), training data represent two polygon shapefiles, one with areas marking change class pixels (**target**), and the other marking other pixels (**background**). The "change" is defined by the data analyst and represents a specific spectrally detectable land cover transition, such as forest to non-forest. The annual change detection metrics are not designed to map slow processes like tree regrowth. To examine land cover dynamics and to create a comprehensive training dataset, data analysts are encouraged to visualize a combination of different metrics (**section 5.2**).

To collect training data, follow the routine described below:

- Download required ARD data (**section 3.5**) and build change detection metrics (**section 4.2**).
- Copy the **C:\GLAD_Tools\Templates\ 08_Change_detection** to your work folder.
- Open QGIS (new project).
- Add raster layers (mosaics of selected metrics). Optionally: load the Google Satellite layer using the QMS plugin.
- Load **target.shp** and **background.shp** files. Put the target layer onto the top of the background layer in the Layer Panel.
- Start editing (Toggle Editing button) for both shapefiles.
- Use the "Add Polygon" tools to add training samples. Avoid creating large training polygons. Distribute samples over the entire area of the image.
- Save layers and project (periodically)

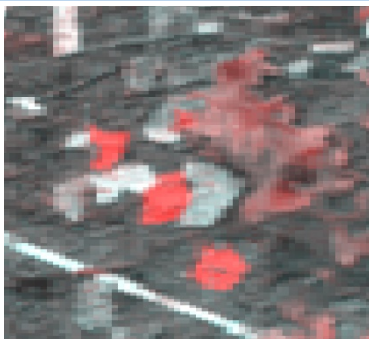Example of the training site for forest loss detection

| Image composite (**section 5.2**) | Target training | Background training (overlaid with target training) |
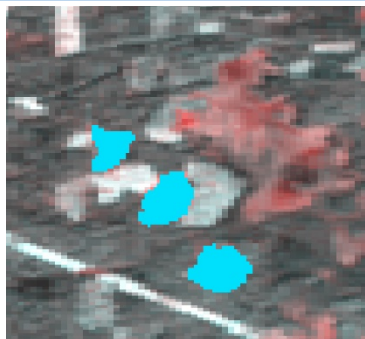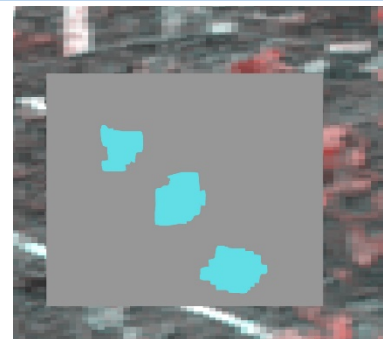
9.3. Applying Classification

The model application for change detection is the same as for the land cover classification. See **section 6.3** for parameter file structure (**classification_change_2022.txt**) and commands (**commands.txt**). Usually, the classification process must be iterated several times to obtain a sufficiently accurate map (**section 6.5**).

Example of forest change classification

1. Image composites (**section 5.2**) used as source data for change detection and drawing training sites in QGIS
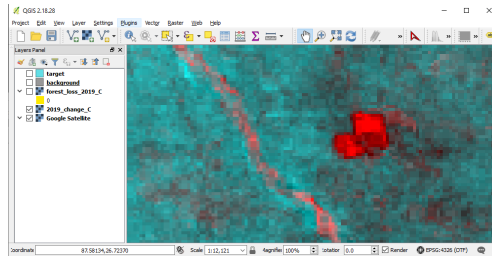
2. Drawing change and no change training sites.
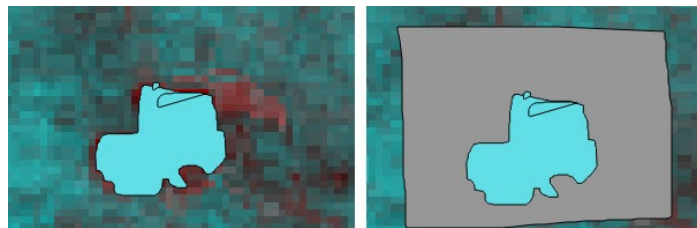
3. Classification parameter file.
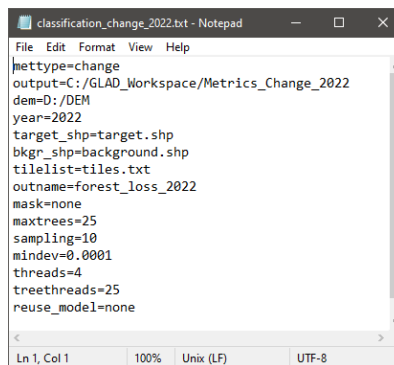
4. CMD command

```
>perl C:/GLAD_Tools/classification.pl classification_change_2022.txt
```

5. Open QGIS and load the classification result (TIF file). To visualize the target class, use transparency threshold 0-49. To show only the background class, apply transparency to the interval 50-100.
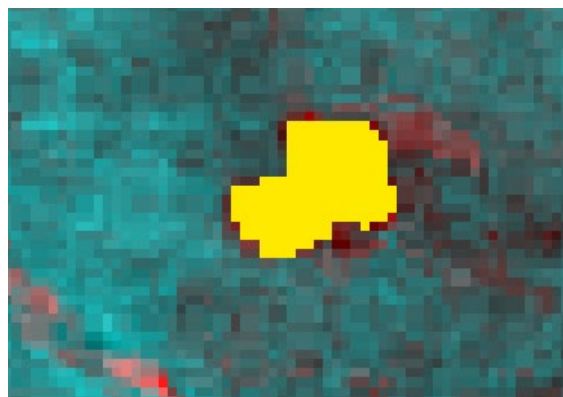
10. Aggregateing LCLU Maps

After classification for multiple land cover classes is complete, separate classes may be aggregated into a LCLU map using **Image Modeler**. The Image Modeler tool allows the application of simple mathematical functions and conditional statements for a set of raster layers. See the *User Manual* document for details on using Image Modeler and other raster image analysis tools.

To implement Image Modeler, all input images must be in geographic coordinates on the WGS84 datum. All input images must have the same extent (in pixels), same pixel size, and same UL coordinates. The 8-bit unsigned, 16-bit unsigned, and 16-bit signed data in GeoTIFF format are supported for input images. The output image is an 8-bit unsigned raster (uncompressed GeoTIFF). Input layers may include results of data analysis (focal average, recoding) and rasterized vector layers.

The workspace example for the LCLU map aggregation is provided here:
C:\GLAD_Tools\Templates\09_LCLUC_map. Copy this template to your workspace before running the commands.

The workspace includes the following input layers (land cover classification results):

|	
water_2022.tif	Permanent water bodies
settlements_2022.tif	Built up land and rural mosaics
treecover_2022.tif	Tree canopy cover
cropland_2022.tif	Annual cropland

Each map represents the likelihood of a target class, 0-100%.

The objective of the tool is to create a LCLU map with the following classes:

> 1 – Inland water
> 2 – Tree canopy cover
> 3 – Cropland
> 4 – Settlements
> 5 – Other lands

The model parameter (**lclu_model_2022.txt**) includes the path to the files and data processing rules. The parameter file includes two sections: **input data definition** and **model definition**. The sections are clearly marked (**INPUT/END** and **MODEL/END**). These marks are important for the code to interpret the parameters correctly. Each input file is associated with a variable denoted as **%n**, where n is the number of the input file. The output file is represented by variable **%0**. The Image Modeler supports **only one output layer**.

The format of the model section is similar to the C++ syntax for arithmetic and conditional statements. The following examples illustrate the most common operations:

Arithmetic operations:	%0= %1+%2; %0= (%1-%2)/(%1+%2)*100;

Conditional statements:	if (%1==1) {%0=10;}
	if (%1==1 or %2==1) {%0=1;}
	       else if (%3==2) {%0=2;}
	       else {%0=3;}

The following example shows the parameter file to create an LCLU map using conditional statements (**lclu_model_2022.txt**):

INPUT	The header of the input section (do not remove)
%1=water_2022.tif	First input file associated with the variable %1
%2=settlements_2022.tif	Second input file
%3=treecover_2022.tif	…
%4=cropland_2022.tif	
END	End of the input section (do not remove)
MODEL	The header of the model section (do not remove)
if (%1>=50) {%0=1;}	
else if (%3>=50) {%0=2;}	
else if (%4>=50) {%0=3;}	Conditional statements
else if (%2>=50) {%0=4;}	
else {%0=5;}	
END	End of the model section (do not remove)

To run the Image Modeler, open the CMD in the folder that contains the parameter file and execute the following command:

```
perl C:/GLAD_Tools/raster_model.pl lclu_model_2022.txt LCLU_2022.tif
```

The output file (**lclu_2022.tif**) has 5 LCLU classes that are consistent with the required legend.

1 – inland water
2 – tree canopy cover
3 – cropland
4 – settlements rural mosaics
5 – other lands
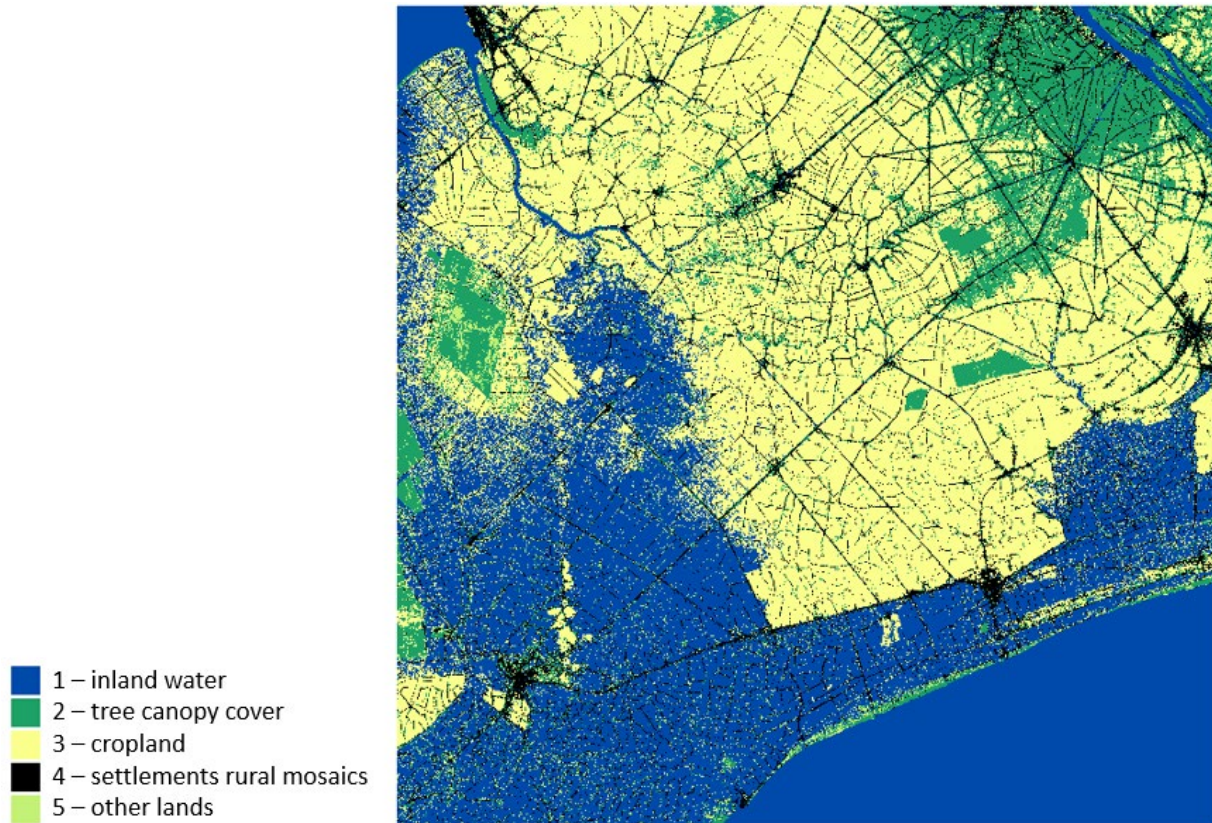
11. Raster Area

11.1. Area per LCLU class

The raster area tools use an equation for the area of a spherical trapezoid on the WGS84 ellipsoid to correctly calculate the area of each pixel of the raster data in ARD geographic format. The tool is required to estimate the area of classes from the output classification maps. The input file should be in geographic coordinates on WGS84. Only 8-bit unsigned GeoTIFF files are supported. The tool outputs the area in square meters and the number of pixels for each class present on the map.

To run the tool, use the following **CMD** command:

C:\GLAD_Tools\get_area.exe <input>.tif <output>.txt

<input> is the name of the map file; <output> is the name of the output area report text file.

The output file contains information on the area and pixel count of each data pixel value. The area is in square meters. The "i" column denotes pixel values.

i	area,m2	count,pixels
0	5646992607.4	13721122
1	20695919151.7	50278878

Example of the command implemented using template **C:\GLAD_Tools\Templates\09_LCLUC_map**:

C:\GLAD_Tools\get_area.exe lclu_2022.tif LCLU_2022_area.txt

11.2. Zonal Statistics

The tool calculates the area of target pixel classes within zones defined by another raster layer. Both input files should have the same extent, pixel size, and UL coordinate and both should be 8-bit unsigned GeoTIFF raster files. The input files should be in geographic coordinates on WGS84.

To run the tool, use the following **CMD** command:

C:\GLAD_Tools\zonal_stat.exe <input>.tif <zones>.tif <output>.txt

<input> is the name of the map file; <zones> is the name of the zones map file; <output> is the name of the output area report text file.

The output file contains information on the area and pixel count of each data pixel value. The area is in square meters. The "class" columns denote pixel values. The header row denotes columns corresponding to zones.

class	0	1	2	3	4
0	0.0	363759095.3	363481653.7	44602049.6	545098244.4
1	0.0	1338230723.6	2308410369.5	56567126.8	2228712247.0

12. Sample Analysis

12.1. Sampling Design

The GLAD Tools supports stratified random sampling design. Simple random sampling is also supported, in this case, just one stratum is considered.

Sample allocation requires a source raster layer that defines the sampling region (area of interest) that is subdivided into sampling strata. The raster should be in the same format as ARD data, with pixel size and pixel boundary matching the ARD dataset. Only 8-bit GeoTIFF files are suitable for the following process.

The workspace template **C:\GLAD_Tools\Templates\10_Sampling_Design** provides an example of the sampling design workflow. The workspace template includes **lclu_2022.tif** map layer that defines sampling strata and **provinces.shp** vector layer that defines AOI. To work with this template, copy it to your workspace.

12.1.1. Sampling AOI extent.

GDAL utilities (accessible through OSGeo4W) are used to create a raster file that represents the area of interest (AOI). Follow these steps to create the AOI raster layer:

1. Open the LCLU 2020 map (**lclu_2022.tif**) in a new QGIS project.

2. Open layer properties and copy the Extent data. The extent shows two pairs of X, Y coordinates for the **LL and UR corners**.
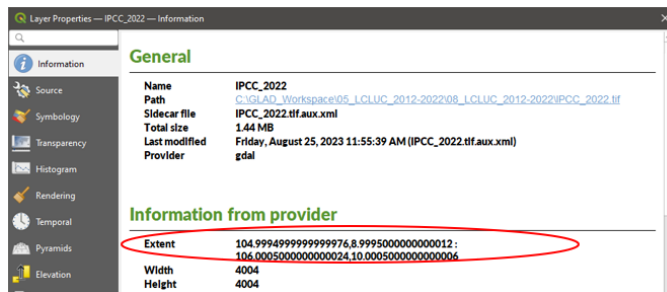
3. Convert the extent coordinates into format:

ULx LRy LRx Uly (or Left Bottom Right Top):

>   104.9994999999999976,8.9995000000000012 : 106.0005000000000024,10.0005000000000006

<div align="center">↓</div>

<div align="center">104.9995 8.9995 106.0005 10.0005</div>

4. Open the **OSGeo4W Shell** interface using the Windows Start menu. Navigate to the workspace folder.

5. Run the following command to rasterize the provinces layer (one line!)

```
gdal_rasterize -te 104.9995 8.9995 106.0005 10.0005 -tr 0.00025 0.00025 -ot Byte -of GTiff
-co COMPRESS=LZW -co BIGTIFF=IF_SAFER –burn 1 provinces_NRSD_training.shp provinces.tif
```

6. Load the output file **provinces.tif** to the QGIS project and check that both the LCLU map and the map of provinces have the same extent.

* If AOI is presented as a raster file, use GDAL command "**gdalwarp**" to resample (and reproject) it to the ARD pixel grid and the LCLU map extent.

12.1.2. Strata map.

The strata map is a raster layer that is usually created by the intersection of the LCLU map and the AOI extent raster. Each pixel of the strata layer should have a value corresponding to the stratum ID (1– N). Pixels with values of 0 are considered outside of the sampling region.

Follow these steps to create the **strata map**:

1. Check and modify (if needed) the Image Modeler parameter file **clip_strata.txt**. The file contains commands to clip the LCLU map using province boundaries:
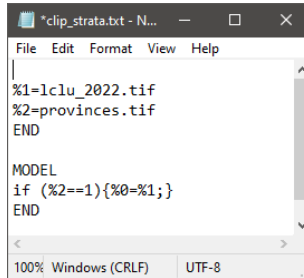
2. Open CMD in the workspace and execute the command:

```
perl C:/GLAD_Tools/raster_model.pl clip_strata.txt strata.tif
```

3. Open the **strata.tif** file in QGIS and check if it is correct.

12.1.3. Strata area.

To calculate the strata area and pixel counts, use the image area tool (**section 11.1**). Run the following command in the CMD:

```
C:\GLAD_Tools\get_area.exe strata.tif strata_area.txt
```

Open the file **strata_area.txt** and transfer its content to the Excel table **Sampling_Design.xlsx** (ignore value 0), columns A-C.

12.1.4. Sample allocation.

The sample allocation table is provided in Excel file **Sampling_Design.xlsx (**see "Table for parameter file" section below).

	Strata area (from strata_area.txt)			For sampling_parameters.txt				For area_parameters.txt		
	A	B	C	D	E	F	G	H	I	J	K
1	Stratum	area,m2	count,pixels		Table for the parameter file				Table for estimation parameter file, area in ha		
2	1	1193212165	1570955		1	1570955	20		1	119321.216	1570955
3	2	576733667	760344		2	760344	20		2	57673.367	760344
4	3	1806421051	2380432		3	2380432	20		3	180642.105	2380432
5	4	436530419.9	575173		4	575173	20		4	43653.042	575173
6	5	88239425.6	116186		5	116186	20		5	8823.943	116186

Here, we allocated 20 samples for each stratum for presentation purposes. Follow these **principles of sample allocation** in the actual work:

1. Allocate **100-1000 samples per stratum**.

2. Large strata with possible uncertainty (e.g., cropland areas that include mixed pixels) requires a **larger number of samples** compared to homogeneous and small strata.

3. Using optimal sample allocation for multiple classes is not useful and hard to implement.

4. Samples may be added **iteratively** after analyzing the results based on the initial sample set.

Copy the "Table for parameter file" section of the Excel table into **sampling_parameters.txt** file:

Strata – name of the sampling frame file.
First – number of the first sample.

SAMPLING (Header of the sampling frame)
Each line (tab-separated):
Strata_ID   N_pixels   n_samples

END (End of the sampling frame)

Open CMD in the workspace and perform the following command:

```
perl C:/GLAD_Tools/samples_generate.pl sampling_parameters.txt
```

The output file **sample_coordinates.txt** contains the list of selected samples. The columns are Sample ID; Stratum; Pixel center X; Pixel center Y.

sample_coordinates.txt

The sample table contains only the center point coordinates. To create a vector dataset with sample outlines, use **sample_kml.pl** tool. The tool can be executed using the following command:

```
perl C:/GLAD_Tools/samples_kml.pl sample_coordinates.txt
```

The output includes KML files of center points (**samples_points.kml**), sample outlines (**samples_polygons.kml**), and sample outlines with nine sub-plots for sub-pixel interpretation (**samples_polygons_subplots.kml**).

Example of sample represented by the center point, outline, and outline with subplots.

37
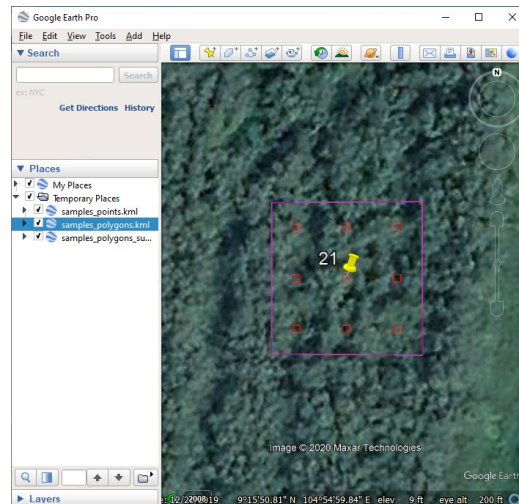
12.2. Extracting Sample Data

The following **data are required** for sample data extraction:

1. The ARD data for the entire analysis interval (e.g., from 2012 to 2022) should be downloaded.

2. The list of tiles (**tiles.txt**).

3. The list of sample coordinates (**sample_coordinates.txt**) from **section 12.1.4**.

Sample data extraction controlled by a parameter file that has the following structure:

Sample reference data extraction parameter file (sample_data_parameters.txt)

tile_list=tiles.txt	Name of the tile list file
sample_list=sample_coordinates.txt	Name of the sample coordinates file
start_year=2012	First year of reference data
end_year=2022	Last year of reference data
ARD=C:/GLAD_Workspace/ARD	Path to the ARD data folder
threads=5	Number of parallel processes

* The reference sample extraction will not work if the first and the last year of the reference data are the same.

* Unlike metric generation and classification processes, the reference sample extraction does not require large RAM. We recommend using all available cores for data processing.

See the example of sample data extraction parameter file here:
C:\GLAD_Tools\Templates\10_Sampling_Design\sample_data_parameters.txt

Open CMD in the workspace and run the following command:

```
perl C:/GLAD_Tools/samples_data.pl sample_data_parameters.txt
```

The command may take a long time to complete. If errors are found, remove all temporary folders and files before re-running the command.

The reference data are stored in a new subfolder **Sample_Data**. For each sample, there are annual image composites, bi-monthly composites, temporal profile, and KML file. A set of html pages is created for each sample to display annual and monthly data. The sample index page is located in the root folder (**image.html**). The index file allows navigation to any sample. Each sample page contains temporal profiles of NDVI, NDWI, and SWIR1 reflectance. The indices and reflectance scaling were selected to simplify visual analysis: NDVI and NDWI values scaled to range 0-2; and SWIR2 reflectance scaled to range 1-3. The monthly composites open in a pop-up window when clicking on the image composite area. The KML may be opened automatically in Google Earth (check browser settings). We recommend using the Chrome browser to work with sample data.

The link to Google Earth simplifies VHR data analysis

index.html allows navigation between samples.

Sample page provides navigation links, reflectance temporal profile and annual average reflectance image composites (SWIR1-NIR-red band combination).

Bi-monthly data provides average reflectance for every two month.

12.3. Sample Interpretation

Sample interpretation is the most important part of the sample analysis. Interpretation should be performed by an analyst that knows regional land cover and the relationship between land cover types and their spectral properties. If several interpreters are working independently, **quality control** and **disagreement resolution** protocols should be implemented. The **classification legend** should be selected considering the ability of analysts to discriminate classes. The protocol for dealing with **mixed pixels** should be implemented to avoid interpretation errors.

We recommend using a spreadsheet editor, e.g., MS Excel or Google Sheets, to record sample interpretation results. The sample interpretation table should have sample IDs, but no information about strata as it may cause interpretation bias. The land cover or change classes may be recorded using different formats (see examples below).

### Examples of sample data recording.

Simple land cover classes. The entire pixel is assigned to the majority land cover class. Class abbreviations were used to simplify data entry.

	A	B	C	D
1	ID	X	Y	catergory
2	1	130.6144	-23.4136	SNC
3	2	139.5139	-27.4239	SNC
4	3	143.6766	-37.8344	SNC
5	4	145.7281	-34.0169	CG
6	5	177.3454	-38.0679	SNC
7	6	146.7619	-31.9441	cl
8	7	130.8436	-28.8001	SNC
9	8	141.7591	-17.9009	SNC
10	9	115.5271	-28.3911	SNC
11	10	131.0376	-18.6859	SNC
12	11	148.3834	-26.6294	SNC
13	12	147.6321	-22.4626	CG

Sample data recorded as proportion of land cover classes within each pixel. The fractions of 0, 0.25, 0.5, 0.75, and 1 are recorded.

ID	Cropland	Rural Mosaic	Wetland	Aquaculture	Mangroves	Build-up
1	1					
2					1	
3	0.5	0.5				
4			1			
5	1					
6		0.5		0.5		
7				0.5	0.5	
8	0.5				0.5	
9			1			
10		0.5				0.5
11		0.5		0.5		

Example of sample interpretation (tree canopy cover) using sub-plots
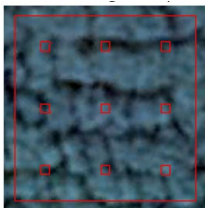
9 sub-plots intersect canopy
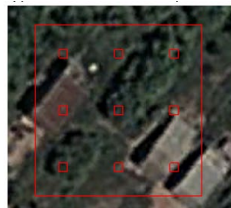
3 sub-plots intersect canopy

0 sub-plots intersect canopy. Even there is trees within this sample none of the sub-plots intersect it.

The area of a class is recorded as the count of sub-plots that intersect with the class. We used 9 subplots. The percent can be calculated from the count (n) using the equation:

% tree cover = n/9×100

12.4. Sample-based Area Estimation

The following example shows the case of a sample interpretation protocol where the entire sample area is allocated to only one land cover class. For fractional land cover class interpretation refer to **User's Manual**.

The workspace template **C:\GLAD_Tools\Templates\11_Sample_Estimation** provides an example of sample interpretation data and sample analysis results. All data provided in this section is generated for training purposes and does not represent the actual sample interpretation output.

The source sample interpretation data is provided in the table **Sample_Analysis.xlsx**, page **sample_reference_data**:

	A	B	C	D	E
1	ID	Stratum	X	Y	Interpreted Type
2	1	1	105.4174	9.225625	water
3	2	1	105.5416	9.258125	water
4	3	1	105.3299	9.415625	water
5	4	2	105.6906	9.851875	trees
6	5	3	105.7196	9.812375	cropland
7	6	5	105.7874	9.283875	other
8	7	1	105.3496	9.094875	water
9	8	5	105.6904	9.242875	settlement
10	9	2	105.8186	9.924625	trees
11	10	2	105.4226	9.697375	trees

Columns A-D are from the sample allocation table sample_coordinates.txt (**section 12.1.4**)

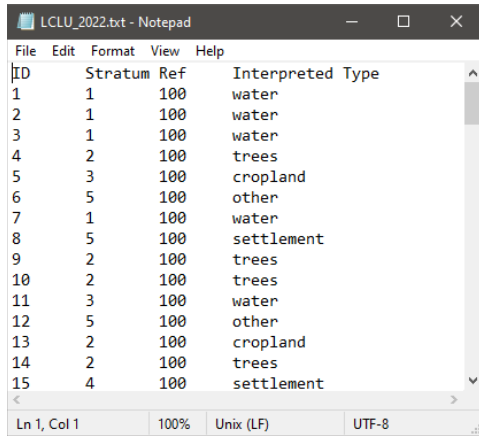Column E contains sample interpretation results

Convert the interpretation results into the data input table as shown below. The input table should have the following columns (please strictly follow the instructions otherwise the results will be incorrect):

- Sample ID (link to column A)
- Sample strata (link to column B)
- Reference (value = 100, indicating the entire sample area)
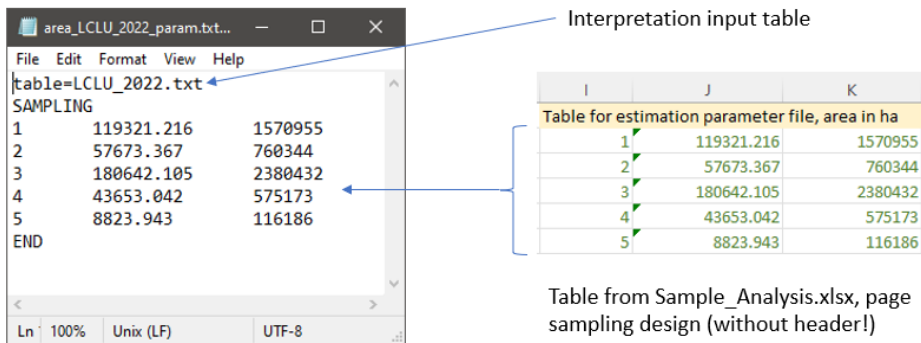- Type (link to the corresponding reference data column)

The example is provided in the table **Sample_Analysis.xlsx**, page **area_estimation_input**:

	A	B	C	D	E	F	G	H	I	J
1	ID	Stratum	X	Y	Interpreted Type		ID	Stratum	Ref	Interpreted Type
2	1	1	105.4174	9.225625	water		1	1	100	water
3	2	1	105.5416	9.258125	water		2	1	100	water
4	3	1	105.3299	9.415625	water		3	1	100	water
5	4	2	105.6906	9.851875	trees		4	2	100	trees
6	5	3	105.7196	9.812375	cropland		5	3	100	cropland
7	6	5	105.7874	9.283875	other		6	5	100	other
8	7	1	105.3496	9.094875	water		7	1	100	water
9	8	5	105.6904	9.242875	settlement		8	5	100	settlement
10	9	2	105.8186	9.924625	trees		9	2	100	trees
11	10	2	105.4226	9.697375	trees		10	2	100	trees
12	11	3	105.6546	9.658125	water		11	3	100	water
13	12	5	105.3241	9.345875	other		12	5	100	other
14	13	2	105.8289	9.782125	cropland		13	2	100	cropland
15	14	2	105.7889	9.716875	trees		14	2	100	trees

Copy the input data table (columns G-J) into a new text file **LCLU_2022.txt**. Check that the file has no empty lines at the beginning or the end of the file.

The file **area_LCLU_2022_param.txt** contains parameters for the area estimation.

Interpretation input table

Table from Sample_Analysis.xlsx, page sampling design (without header!)

Open CMD in the workspace and run the CMD command:

```
perl C:/GLAD_Tools/samples_estimate_area.pl area_LCLU_2022_param.txt
```

The output file **Area_report_LCLU_2022.txt** contains sample-based area estimation report:
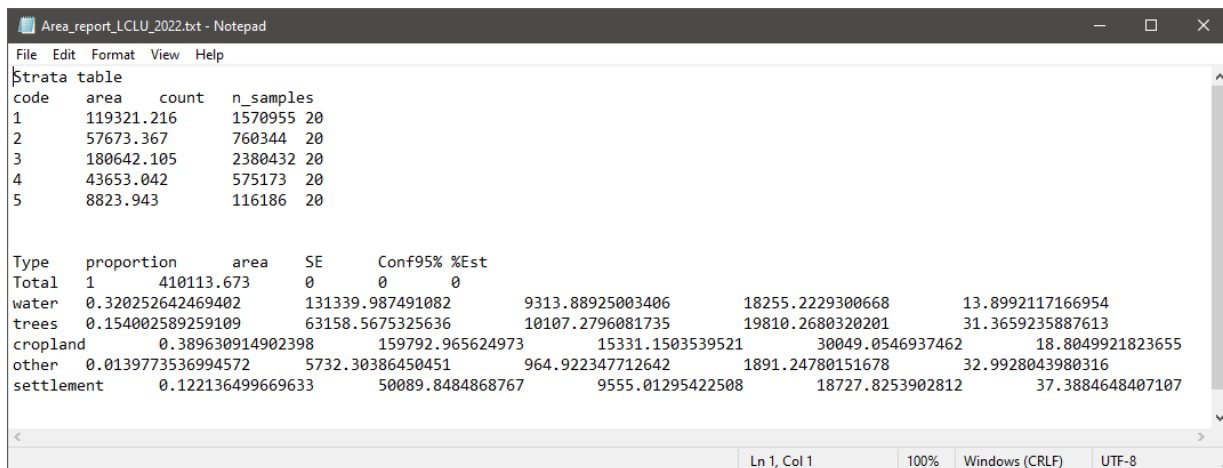
Sample area report structure

Strata table			
code	area	count	n_samples
1	119321.216	1570955	20
2	57673.367	760344	20
3	180642.105	2380432	20
4	43653.042	575173	20
5	8823.943	116186	20

Strata table is the copy of the table provided by the parameter file. The number of samples is calculated from the sample table. If this number is not correct, it is due to a possible error in sample table.

The area estimation table shows the following statistics (for each target estimation class)

- Proportion of the class within AOI
- Class area (in area units used in the parameter file)
- Standard error of area estimation (area units)
- 95% confidence interval of area estimation (area units)
- Percent 95% confidence interval of estimated class area (percent)

The "Total" row shows the total AOI area if multiple classes were used for the analysis.

Type	proportion	area	SE	Conf95%	%Est
Total	1	410113.673	0	0	0
water	0.32	131339.99	9313.89	18255.22	13.90
trees	0.15	63158.57	10107.28	19810.27	31.37
cropland	0.39	159792.97	15331.15	30049.05	18.80
other	0.01	5732.30	964.92	1891.25	32.99
settlement	0.12	50089.85	9555.01	18727.83	37.39

12.5. Map Accuracy Estimation

The map accuracy estimation requires (1) the map land cover class information and (2) the reference interpreted land cover information for each sample. Both datasets should use the same land cover definitions that define the class.

In case the map data do not match the strata, use the following workflow to extract map labels for each sample:

1. Make a parameter file **sample_2022_map_data_param.txt**

sample_list=sample_coordinates.txt	Path to sample coordinate table
map_file=lclu_2022.tif	Path to map file (tif)
threads=5	Number of parallel processes

2. Open CMD in the workspace and run the command:

```
perl C:/GLAD_Tools/samples_map_data.pl sample_2022_map_data_param.txt
```

3. Open the output text file **sample_map_data_report_lclu_2022.txt** and transfer the values (last column, denoted "lclu_2022") to the Excel table **Sample_Analysis.xlsx**, page **map accuracy input** (column F):


The map accuracy tool uses numeric class names, so the reference class names must be converted into numeric codes that match map land cover legend. In the example above, we translated map LCLU classes to column J and converted text LCLU class labels from column E to numeric values in column K.

The resulting input table has the following columns (please strictly follow the instructions otherwise the results will be incorrect):
- Sample ID (link to column A)
- Sample strata (link to column B)
- Map LCLU type
- Reference LCLU type

Copy the input data table (columns H-K) into a new text file **LCLU_2022_accuracy.txt**. Check that the file has no empty lines at the beginning or the end of the file.

GLAD Tools V2.0 Quick Start Guide
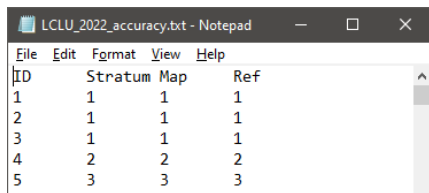
The file **area_LCLU_2022_param.txt** contains parameters for the area estimation.

Input table for accuracy estimation

```
accuracy_2022_param.txt - ...      —    □    ×
File  Edit  Format  View  Help
table=LCLU_2022_accuracy.txt
SAMPLING
1        119321.216      1570955
2        57673.367       760344
3        180642.105      2380432
4        43653.042       575173
5        8823.943        116186
END

Ln 1,  100%    Unix (LF)        UTF-8
```

	I	J	K
	Table for estimation parameter file, area in ha		
1		119321.216	1570955
2		57673.367	760344
3		180642.105	2380432
4		43653.042	575173
5		8823.943	116186

Table from Sample_Analysis.xlsx, page sampling design (without header!)

Open CMD in the workspace and run CMD commands...

... to estimate the overall accuracy of the map:

perl C:/GLAD_Tools/samples_estimate_oa.pl accuracy_2022_param.txt

... to estimate the accuracy of individual LCLU classes:

perl C:/GLAD_Tools/samples_estimate_accuracy_by_class.pl accuracy_2022_param.txt

The output file **Overall_accuracy_report_LCLU_2022_accuracy.txt** contains map overall accuracy report. The map's overall accuracy and its standard error (SE) is shown on the last line of the report.

N samples with map and reference agreement per stratum

N samples with map and reference disagreement per stratum

```
Overall_accuracy_report_LCLU_2022_accuracy.t...   —   □   ×
File  Edit  Format  View  Help
Confusion matrix
Strata  Map==Ref        Map=/=Ref
1        20      0
2        18      2
3        17      3
4        18      2
5        13      7
Accuracy
OA        SE
90.1670784680618         3.81513021989581

Ln 1, Col 1           100%    Windows (CRLF)    UTF-8
```

Overall accuracy and SE

The output file **Accuracy_report_LCLU_2022_accuracy.txt** contains accuracy reports for each LCLU class. The report for each class is separated by a header. Below we present the structure of the report of the first LCLU class:

Report header

Confusion matrix per stratum

Overall accuracy and SE

User's accuracy and SE

Producer's accuracy and SE

User's and Producer's accuracies for the background class (target class absence)

```
Accuracy_report_LCLU_2022_accuracy.txt - ...   —   □   ×
File  Edit  Format  View  Help
Accuracy report for class 1

Confusion matrix
Strata  Map/Ref Map/0   0/Ref   0/0
1        20      0       0       0
2        0       0       0       20
3        0       0       1       19
4        0       0       1       19
5        0       0       2       18
Accuracy
OA        SE
97.049857581495  2.27105065332315
UA_class1        SE
100      0
PA_class1        SE
90.7880778258158         6.4381771176165
UA_class0        SE
95.8404688248196         3.20205083444706
PA_class0        SE
100      0

Ln 1, Col 1           100%    Windows (CRLF)    UTF-8
```

Licenses and Redistribution

The GLAD Tools and Landsat ARD data are available with no charges and no restrictions on subsequent redistribution or use, as long as the proper citation is provided as specified by the Creative Commons Attribution License (CC BY). The toolbox includes libraries and codes that were shared by other open-source software projects:

- MinGW - C++ compiler, GNU C Library (Open-source software; Copyright © Free Software Foundation)
- gdal - GDAL Core (Open-source software; Copyright © Frank Warmerdam and others)
- tree.exe – CART model (Open-source software; Copyright © B. D. Ripley and J. Ju)
- Other utilities – GLAD ARD Tools (Freeware; Copyright © GLAD UMD)

Copyright © Global Land Analysis and Discovery Team, University of Maryland

Suggested citation: Potapov, P., Hansen, M.C., Kommareddy, I., Kommareddy, A., Turubanova, S., Pickens, A., Adusei, B., Tyukavina A., and Ying, Q., 2020. Landsat analysis ready data for global land cover and land cover change mapping. Remote Sens. 2020, 12, 426; doi:10.3390/rs12030426 *https://www.mdpi.com/2072-4292/12/3/426*

GLAD Tools consists of compiled packages for use as described in the Tools manual. The GLAD team does not make any warranty, express or implied, including the warranties of merchantability and fitness for a particular purpose; nor assumes any legal liability or responsibility for the applications of the GLAD Tools.

GLAD Tools depend on several open-source third-party packages (PERL, curl, wget, OSGeo4W, R). The GLAD team is not responsible for supporting these packages. Users should refer to the licenses of the individual packages for redistribution policies.