



Reference documentation

Sample-based national tree canopy cover and change assessment

Peter Potapov and Alexandra Tyukavina

Global Land Analysis and Discovery Lab
University of Maryland
Department of Geographical Sciences
4321 Hartwick Rd, Suite 400
College Park, MD 20740
United States

National assessment of tree cover extent and change

1. Overview of project activities

National information on tree cover extent and change is required for effective land use planning and management, Forest Resources Assessment reporting to the Food and Agriculture Organization of United Nations, national reporting on greenhouse gas emissions according to the Intergovernmental Panel on Climate Change guidelines and in context of REDD+ agreements, and nature conservation strategy development. Satellite imagery, especially freely available medium spatial resolution data such as delivered by the Landsat program, provides practical means for national tree canopy cover monitoring. Available high spatial resolution from Google Earth is sufficient to interpret land cover type and tree canopy cover proportion. The overarching goal of GLAD training is to support ongoing national efforts for forest mapping and monitoring.

The system developed by the UMD Global Land Analysis and Discovery (GLAD) Lab is used by a number of countries for annual tree cover change monitoring. The system includes two components (stages). The first stage consists of wall-to-wall Landsat-based tree cover extent and change mapping. Wall-to-wall mapping may be performed at the national as well as the regional or global scale. The tree cover extent and change maps are further used to stratify country in the implementation of a stratified random sampling protocol. The second stage of the analysis consists of tree cover and change area estimation based on a sample interpretation using multi-resolution time-series data.

As part of the SilvaCarbon program, Cambodia, Lao PDR and Thailand will start implementing the GLAD System in 2017. This first training workshop is aimed to initiate the capacity building process. The workshop objective is to provide a brief introduction on GLAD System, its methodology and tools, with a particular focus on sample-based methods of forest cover area and change estimation and map accuracy assessment. The sample data presented during the first workshop will be interpreted by the national specialists and will be used to report national tree cover extent and change from the year 2000 to 2016. The in-depth training on national data management and characterization will be performed during the following training workshops.

2. Objectives of the first training workshop

The first task of the national forest monitoring capacity building project will consist of sample-based analysis of tree cover extent for the year 2016 and gross loss from the year 2001 to 2016. The national samples were selected using a subset of the regional tree cover extent and change products developed by the GLAD tem as an input to the Regional Land Cover Monitoring System. Using reference data collected for samples, we are going to estimate the following parameters at the national scale:

(1) tree cover in the year 2016;
(2) tree cover net change area 2001-2016;
(3) gross tree cover loss and gain area;
(4) tree cover loss area by forest type/category in the year 2000;
(5) proximate causes of tree cover loss (area proportion for each cause);
(6) land-cover outcomes of tree cover conversion (area proportion for each outcome).

Every parameter (area, proportion) will be estimated at the national level with known uncertainty (95% confidence interval). Regional total for the Mekong region will be estimated from the national sample interpretation results by the SERVIR-Mekong team.

3. Stratified sampling design

Sampling unit is a single Landsat pixel (~30x30 meters) selected using stratified random sampling design. Stratification is based on the GLAD regional annual tree canopy cover products, specifically, annual tree canopy cover, annual tree cover loss, and tree cover gain 2000-2016. Country boundaries (obtained from the open data source GADM V2) were used as additional strata for sampling design to ensure integration of national products in regional context. Strata within each country represent the following classes:

- "Stable tree cover": pixels with tree canopy cover of at least 10% for any year;
- "Dynamic tree cover": pixels where tree cover loss or gain was detected;
- "No tree cover": pixels where tree canopy cover was always lower than 10%.

The strata codes and their explanation are provided in the table
http://glad.geog.umd.edu/Potapov/SERVIR_Mekong/Regional_workshop_August_2017/Sampling_frame/Sampling_frame.xlsx
(page "sampling frame"). The strata map is available here:
http://glad.geog.umd.edu/Potapov/SERVIR_Mekong/Regional_workshop_August_2017/Sampling_frame/sampling_frame.tif
The raster data are in geographic coordinates, geodetic model WGS84, pixel size 0.00025 degrees.

The total of 5,000 samples was allocated among the countries equally (1,000 samples per country) and among three strata according to the anticipated strata variance. The sample database can be obtained here:
http://glad.geog.umd.edu/Potapov/SERVIR_Mekong/Regional_workshop_August_2017/Sampling_frame/Sampling_frame.xlsx
(page "samples all"). See database fields explanation in the next section.
Reference data for all samples are available from this index file:
http://glad.geog.umd.edu/Potapov/SERVIR_Mekong/Regional_workshop_August_2017/Samples/all_samples.html

For the first training, we selected the first 100 samples in each country for interpretation. Our goal is to interpret these samples during the training workshop and compute preliminary estimates. The final results may be calculated after the entire sample set (1,000 samples per country) is analyzed.

4. Response design (image interpretation)

Image interpretation is based on the Landsat image time-series and available VHR data from Google Earth. Each sampled pixel has its own page containing reference data and assessable from the following index page via sample ID:
http://glad.geog.umd.edu/Potapov/SERVIR_Mekong/Regional_workshop_August_2017/Samples/main_index.html
Each page displays (see figure below):

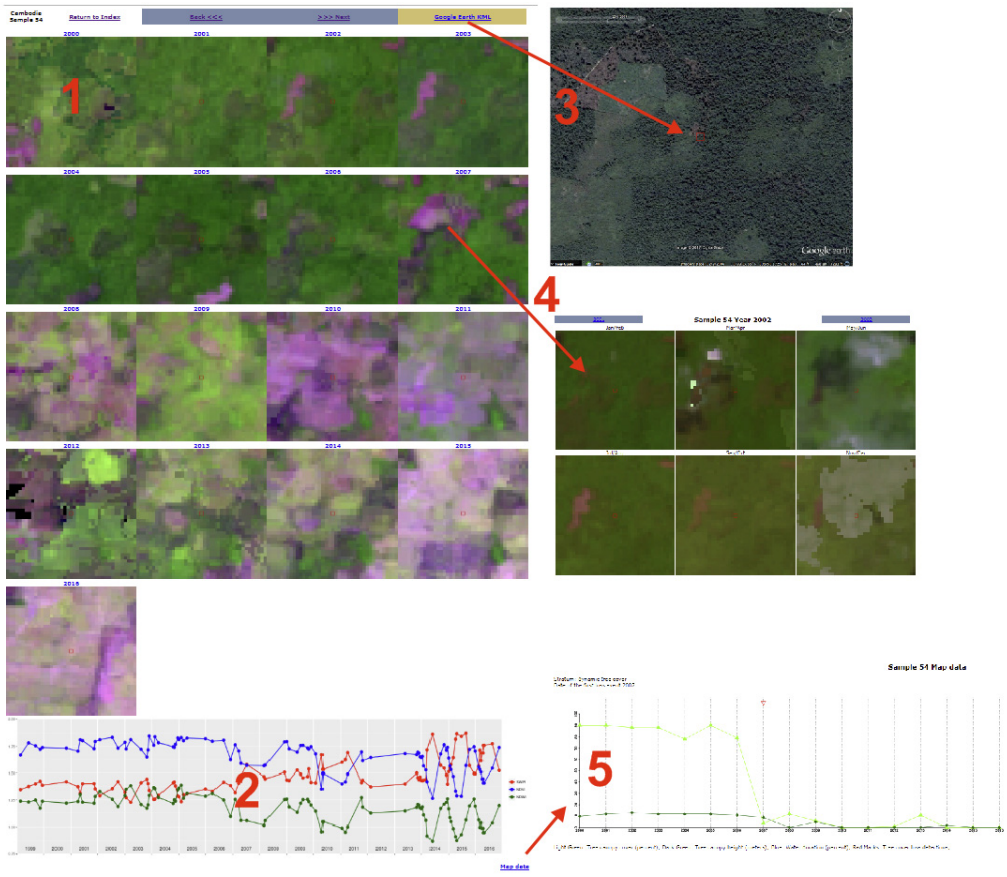
1. A set of annual Landsat image composites (SWIR-NIR-red band combination). Annual mean reflectance value is shown. Black areas have no cloud-free data during selected year. Red outline highlights the sampled pixel (the rest of the image is provided for the context).

2. Temporal profiles of SWIR reflectance, NDVI (normalized ratio of red and NIR bands) and NDWI (normalized ratio of SWIR and NIR bands). Profiles are extracted from 16-day Landsat cloud-free composites.

3. Google Earth KML file with sample outline (link that opens Google Earth and loads KML). To initialize the link in Google Chrome, select “always open this file types” on the downloaded KML file. Google Earth should be installed.

4. Time-series of bi-monthly composites for the sample area (opens in separate window using the annual subset as a link). Sampled pixel is highlighted in red. Remaining clouds are not masked out.

5. Map data, including annual tree canopy cover, tree height, water duration, and change detection provided in a separate window.



Interpretation results should be recorded in a separate table (MS Office, Open office, Google docs, etc.). Interpretation table templates (EXCEL) can be downloaded here:
http://glad.geog.umd.edu/Potapov/SERVIR_Mekong/Regional_workshop_August_2017/Interpretation_tables
Use "Test_interpretation.xlsx" file first to do the training sample interpretation. The interpretation guide and the list of samples are provided as a separate document entitled "Sample interpretation guide".
The Database columns are listed below. During the training, fill only columns marked with asterisk (*). Double asterisk (**) marked columns that should be filled only if a sample has change

	Column name	*Description*	*Technical fields*
	ID	**Sample ID**	
	X	X coordinate of sample center, decimal degrees	Hidden
	Y	Y coordinate of sample center, decimal degrees	Hidden
	STRATUM	Stratum ID	Map data - Hidden
	COUNTRY	Country	Map data - Hidden
	TYPE	Stratum type	Map data - Hidden
	Map_loss	Mapped loss (1/0)	Map data - Hidden
	Map_loss_year	Year of the first detected loss	Map data - Hidden
	Map_gain	Mapped gain (1/0)	Map data - Hidden
	Map_TC2000	Mapped year 2000 tree cover %	Map data - Hidden
	Map_TC2016	Mapped year 2016 tree cover %	Map data - Hidden
**	**Y2000_canopy**	**Year 2000 tree canopy detection (point count)**	Default value = Y2016_canopy
	Y2000_TC	Year 2000 tree cover (%)	Auto calc - Hidden
**	**Y2000_type**	**Year 2016 land cover type (N,S,P,O,NF)**	Default value = Y2016_type
	Y2000_subtype	Year 2000 land cover subtype (CODE)	Hidden
*	**Y2016_canopy**	**Year 2016 tree canopy detection (point count)**	
	Y2016_TC	Year 2016 tree cover (%)	Auto calc - Hidden
*	**Y2016_type**	**Year 2016 land cover type (N,S,P,O,NF)**	
	Y2016_subtype	Year 2016 land cover subtype (CODE)	Hidden
**	**MIN_canopy**	**Minimal tree canopy detection (point count)**	Default value = 2000/2016 min
	MIN_TC	Minimal tree cover (%)	Auto calc - Hidden
**	**MAX_canopy**	**Maximal tree canopy detection (point count)**	Default value = 2000/2016 max
	MAX_TC	Maximal tree cover (%)	Auto calc - Hidden
	LOSS_year	Year of the first tree cover loss (YYYY)	Hidden
*	**CHANGE_type**	**Tree cover change direction (N,L,G,R)**	
	CHANGE_subtype	Change subtype (CODE)	Hidden
*	**VHR1**	**Date of the first VHR image (YYYY)**	
*	**VHR2**	**Date of the last VHR image (YYYY)**	Default value = 2016
*	**CERTAINTY**	**Certainty of sample interpretation (H/M/L)**	Default value = H
	NOTES	*Comments*	*Optional*

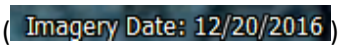
Below are the proposed land cover types and subtypes (based on GLAD work in other countries). These types may be amended or changed during the workshop. For the workshop exercise, use types only.

Type		*Subtype*	
Natural primary and mature secondary forests	**N**	Deciduous/semi-deciduous	ND
		Evergreen	NE
		Mangroves	NM
		Woodlands	NW
		Woody wetlands	NWW
Young secondary forest	**S**		
Tree plantation	**P**		
Trees outside forests	**O**	Industrial orchards, palm plantations, agroforestry, tea gardens, other	AF
		Homestead trees, trees within urban areas, village woodlots, smallholder tree plantations	RF
Treeless areas	**NF**		

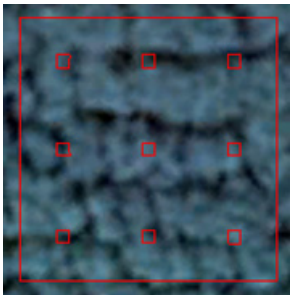
The following codes are proposed for change type and subtype:

Change type		*Change subtype*	
Tree cover rotation	**R**	Logging in natural forests	RL
		Logging in tree plantations	RT
		Non-timber planted trees rotation	RA
		Shifting cultivation	RS
		Charcoal production	RC
		Natural disturbances	ND
Tree cover removal	**L**	Natural forest to tree plantation	LT
		Forest to agriculture (cropland or pasture)	LA
		Forest to construction (settlement, infrastructure, roads)	LC
		Forest to mining area	LM
		Forest flooding (due to water reservoir construction)	LW
		Unknown/other	LU
Tree cover establishment	**G**	Natural forest restoration	GN
		Tree plantation establishment	GP
		Tree cover increase within trees outside forests	GO
No change	**N**		

The following sequence is recommended for sample interpretation:

1. Open the sample interpretation page and check annual composites and spectral profiles to determine land cover type and change dynamics
2. Open Google Earth KML. Use time slider () to navigate between available image dates. Check the dates of the image closest to the end of the year 2016 and the earliest available image. Only consider image dates that are displayed at the bottom of the screen (

). Record the year of the first Very high resolution (VHR) and the last (closest to 31/12/2016) image on Google Earth (columns **VHR1** and **VHR2**).
3. Unless obvious change happened after the date of the last VHR image, use it to estimate tree canopy cover for the year 2016. Count the number of sub-plots hitting tree canopy (tree is defined as woody vegetation above 5m tall). Ignore sub-plots that are within canopy shadow on the ground (if shadow is between tree canopies in a dense tree stand, count it as canopy). Be careful in assessing tree canopy during the dry season. Record the number of sub-plots within the tree canopy in the column **Y2016_canopy** (the value is between 0 and 9).
   The following examples show typical cases and output values:



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.

   If only Landsat data are available for a sample, use the following rules:
   - if sample is considered covered with trees, use value 9
   - if sample is treeless, use value 0
   - if sample is covered with trees but it located immediately on the boundary with the non-forested pixel, use value 5.
4. Record land-cover type for the year 2016 in the column **Y2016_type.**
5. Check temporal profiles of Landsat and VHR data. If no tree canopy change is detected for a sample, put code “N” in the column **CHANGE** and go the step 7.
6. If change is detected, check if it is repeated change or change within regularly rotated landscape to set up the **CHANGE_type** code.
   - If change is unidirectional (tree cover loss or gain), only estimate tree cover for the year 2000 and fill the **Y2000_canopy** and **Y2000_type** columns.
   - If the sample experienced tree cover rotation (consequent loss and gain events), estimate maximal and minimal tree cover (**MIN_canopy** and **MAX_canopy**).
7. Fill **CERTAINTY** of sample interpretation (high, moderate, low: H, M, L). High certainty is set as the default value.

5. Accuracy and Area Estimation

Software requirements

Sample interpretation may be performed on any computer connected to the Internet. All sample data are stored on-line. Required software for image interpretation:

(1) Web browser (recommended browser Google Chrome);
(2) Desktop version of Google Earth or Google Earth Pro;
(3) Spreadsheet editor (MS EXCEL recommended);

For statistical data analysis, the following software required:

(1) ActiveState PERL (community edition)
http://www.activestate.com/activeperl/downloads
(2) Notepad++ (for editing PERL scripts)
http://notepad-plus-plus.org/

In case of slow Internet connection, we recommend to use the local version of the sample reference data (the analyst will still require to have Google Earth and Internet access). A copy of local sample reference data may be obtained from the GLAD team by request.

Data export

In this section we will export source data for running statistical analysis. First, we need to make a workfolder for running a script. Make an empty folder in your user directory, and copy all scripts and text files from
http://glad.geog.umd.edu/Potapov/SERVIR_Mekong/Regional_workshop_August_2017/Sample_analysis/

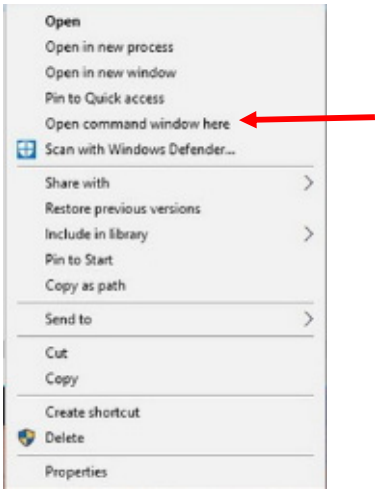
Second, in the completed interpretation table (EXCEL file), select all columns from A to AD, and using context menu from right-click on the column header, select "Unhide". Select all columns again and copy them. In the sample analysis workfolder create empty text file (name it “input_table.txt”) and copy there the table from EXCEL. Make sure that there are no empty lines at the end of the table.

Area estimation

To run the area calculation script, we will need to start the command prompt. You can use any of the two ways:

a. Launch The Command Prompt From The “Run” Window. Launch the “Run” window (press Win+R on your keyboard to open it). Then, type cmd or cmd.exe and press Enter or click OK. Next, you will need to navigate to the estimation workfolder using DOS “cd” command (e.g. “cd D:\myfolder”).

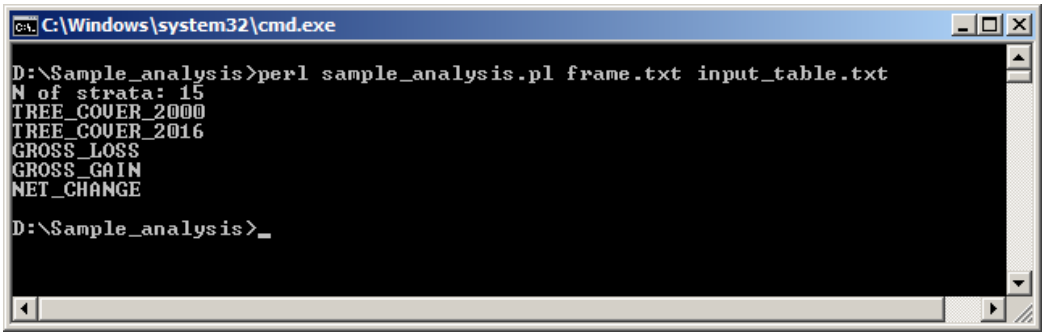
b. Use right-click context menu inside the workfolder to call command prompt:



In the command window, type the following command:

perl sample_analysis.pl frame.txt input_table.txt

If there are no errors, you will see the following messages:



Open the file input_table_report.txt, and copy its contents into the new sheet in EXCEL file. The upper part of the table entitled “Strata table” is a reproduction of the sampling frame table. The N_samples column shows how many samples from each strata were found within the input table. If only one country was processed, you will only see samples for this country.

Each table below shows area estimation for categories: tree canopy cover for the year 2000, tree canopy cover for the year 2016, gross tree cover loss and gain, and net tree cover change. All values are shown in thousands hectares. “Total” line shows total area for a country, and following lines reflect types:

For TCC2000 and NET change – land cover type for the year 2000

For TCC2016 and Gross gain – land cover type for the year 2016

For Gross loss – Change type

Each parameter has area, standard error (SE), and the 95% confidence interval (Conf95%)

Accuracy estimation

Two different scripts are provided to estimate accuracy of the tree canopy cover map for the year 2016 (estimate_accuracy_TCC2016.pl) and tree cover loss map (estimate_accuracy_loss.pl). To run each script, use the following commands:

perl estimate_accuracy_TCC2016.pl frame.txt input_table.txt

perl estimate_accuracy_loss.pl frame.txt input_table.txt

The output is stored in files “input_table_TCC2016_accuracy_report.txt” and “input_table_LOSS_accuracy_report.txt”. In each report, the “Confusion matrix” section shows proportion of each agreement type per strata, and “Accuracy” section shows Overall accuracy “OA”, User’s accuracy “UA” and Producer’s accuracy “PA” for each class (tree cover or tree cover loss, respectively, denoted as class “1” and the remaining as class “0”). Accuracy estimations are supplemented with their standard errors (SE).