REDD+ TRAINING OF TRAINERS

Satellite Land Monitoring System (Pakistan)

Satellite Land Monitoring System (SLMS)

Organized by

IUCN Pakistan &

National REDD+ Office (Ministry of Climate Change)

Venue: Margallah Hotel *Date: 28-30 July 2021*

Abbreviations

AD	Activity data
AJK	Azad Jammu & Kashmir (autonomous territory)
BRDF	Bidirectional Reflectance Distribution Function
DOS	Dark Object Subtraction
FREL	Forest Reference Emission Level
FRL	Forest Reference Level
GIS	Geographic Information System
GPS	Global Positioning System
ha	Hectare (1 ha = 10,000 m^2)
HR	High Resolution
IPCC	Intergovernmental Panel on Climate Change
km^2	square kilometre
KP	Khyber Pakhtunkhwa (Province)
L5 TM	Landsat 5 Thematic Mapper
L8 OLI	Landsat 8 Operational Land Imager
LULC	Land Use Land Cover
LULUC	Land Use Land Use Change
MRV	Monitoring, Reporting and Verification
NFI	National Forest Inventory
NFMS	National Forest Monitoring System
REDD	Reducing Emission from Deforestation and Forest Degradation
RS	Remote Sensing
SLMS	Satellite Land Monitoring System
TOA	Top of Atmospheric
UTM	Universal Transverse Mercator
WGS	World Geodetic System

Contents

1.	**Introduction**	**1**
1.1.	Background	1
1.2.	Objective of the Traning Module	1
1.3.	Using this Manual	1
1.4.	Traning Datasets	3
1.5.	Approach for SLMS – Land Use Land Cover Mapping	4
2.	**Satellite Imagery Acquisition and Processing**	**5**
2.1.	Exercise 1A Image Exploration and Downloading	6
2.1.1.	Background	6
2.1.2.	Exploring and Downloading Landsat Images	6
2.2.	Exercise 1B Atmospheric and BRDF Corrections	10
2.2.1.	Atmospheric Correction and Topographic Correction	10
3.	**Systematic Sampling and LULC Interpretation**	**22**
3.1.	Exercise 2A Sample Design for Visual Interpretation of Systematic Grids	23
3.1.1.	Background	23
3.1.2.	Generation of Sampling Grids/Plots	23
3.1.3.	Assignment 1 Densification of Sampling Plots	32
3.1.4.	Home Province Assignment	32
3.2.	Exercise 2B Interpretation of LULC Classes using OpenForis CollectEarth	33
3.2.1.	Background	33
4.	**Sample Set for Image Training**	**38**
4.1.	Exercise 3 Creating Training and Validation Set	39
4.1.1.	Background	39
4.1.2.	Generating Region of Interest (ROI)	40
4.1.3.	Training and Validation Subsets	56
5.	**Image Classification using Random Forests**	**60**
5.1.	Exercise 4 Image Classification Using Random Forests	61
5.1.1.	Background	61
5.1.2.	Image Classification	62
6.	**Post Classification Processing**	**73**
6.1.	Exercise 5 Post Classification Processing	74
6.1.1.	Background	74
6.1.2.	Classification Post Processing	75
6.1.3.	Assignment	77
7.	**Accuracy Assessment and Area Estimation**	**78**
7.1.	Exercise 6 Accuracy Assessment and Area Estimation	79
7.1.1.	Background	79
7.1.2.	Exercise 6A Designing Reference Sample (Optional)	80

7.1.3. Exercise 6B Construction Error Matrix 85
7.1.4. Calculating User's, Producer's, Overall Accuracies, Variances and Confidence Interval 89
7.1.5. Calculating Standard Error 90
7.1.6. Calculating Area Estimation and Confidence Interval 91
7.1.7. Summary Result 92

List of Figures

Figure 1 Training data files structure 3
Figure 2 SLMS Land Use Land Cover Mapping Flowchart 4
Figure 3 Landsat 8 scene (a) before topographic correction (b) after topographic correction 16
Figure 4 Band stacked L8 Image (a) B432 NIR composite (b) B543 SWIR composite 18
Figure 5 Individual atmospherically corrected image scenes 20
Figure 6 Clipped image scenes 20
Figure 7 Mosaiced image scenes 21
Figure 8 System overview of Collect Earth (http://www.openforis.org/tools/collect-earth.html) 33
Figure 9 Mosaic L8 image with AJK clipping boundary vector data and clipped L8 image 40
Figure 10 Confusion matrix text output 67
Figure 11 Random Forest training model output 67
Figure 12 Classified raster before and after running Sieve algorithm 76

1. Introduction

1.1. Background

Satellite Land Monitoring System (SLMS) integrates remote sensing method to assess activity data related to land use/land use change and forest degradation processes over large areas. SLMS is one of the three functions of Monitoring, Reporting and Verification (MRV) to continuously monitoring land by establishing National Forest Monitoring System (NFMS).

Under the "Develop Forest Reference Emission Levels/Forest Reference Level and National Forest Monitoring System (NFMS)-Measurement and Reporting and Verification (MRV) System for REDD+", a robust SLMS approach based on open data and Free and Open Source Software (FOSS) is being designed and tested. The SLMS is designed to use both desktop based processing platforms as well as cloud based system for accessing, processing, analyzing and reporting activity data. The desktop based processing approach uses various FOSS tools to process, analyze and report remote sensing based spatial data products especially land use, land use changes and degradation.

Cloud based computation systems to support SLMS, such as FAO's SEPAL[1], European Space Agency supported Forestry TEP[2] will are now available. However, these platforms are currently under Beta testing phase and will be accessible in near future.

This short training addresses the core process in the application of desktop based processing and analysis platform for land use and land use change assessment.

1.2. Objective of the Traning Module

The primary objective of this Module 1 Satellite Land Monitoring System is to learn various technical processes of SLMS Land Use Land Cover Mapping using open source tools. Other objectives of this module are:

- Image sources and tools to download open satellite imagery data
- Satellite image processing
- Design of systematic sampling for land use/land use change mapping
- Generating of data for image analysis and classification
- Image classification using advanced machine learning algorithm – Random Forests
- Accuracy assessment, uncertainty calculation and area estimation for activity data reporting
- Change analysis

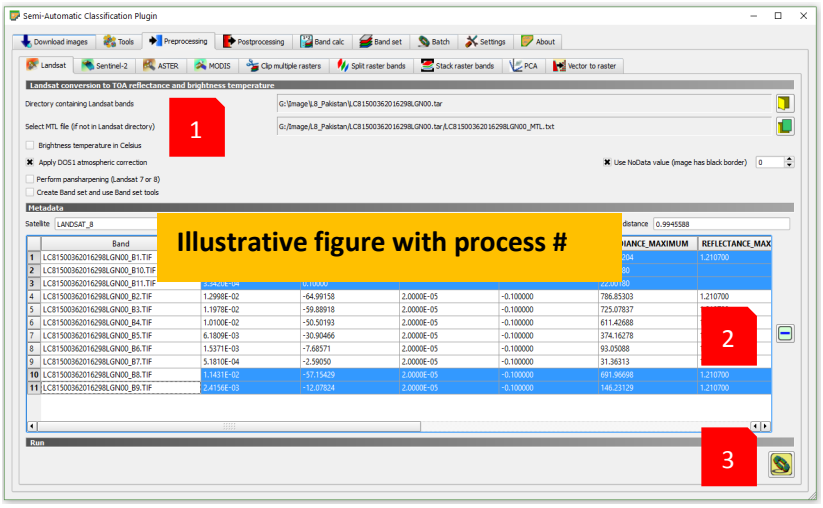
1.3. Using this Manual

- General flowchart of the **Steps** is shown at the start of the exercise.
- Data required is listed in the table **Data Used.**
- Software required is listed in **Software Used**.
- Each exercise has key **Steps** with specific **Processes** to follow and illustrative figure indicating process #.

[1] https://sepal.io/
[2] https://forestry-tep.eo.esa.int/app/#/

Step 1

1. Process 1
2. Process 2
3. Process 3
4. Process …….



- GIS/Software commands/menus to be used are *Italicized and highlighted with Grey*
- Important notes/tips are given in *[Note:……..]*
- Filenames are '*Italicized and single quoted*'
- Exercises are followed by **Assignments,** wherever needed and must be completed to go to the next exercise.
- *Messages and instructions are highlighted are in red and highlighted in yellow*
- Cautions are indicated in red within a red box
- Besides the steps and processes in the manual, there might be several different ways to do certain steps/process in the exercises, participants are encouraged to try out such different processes
- Optional Exercise or Step are highlighted in blue and indicated as (Optional). These are advance exercises/steps for which simple or alternative exercise is given. However, the participants are encouraged to attempt these in their available time.

1.4. Traning Datasets

The training references and datasets are structured under different Exercises Folders under Module2_SLMS as shown. It is advised to follow the data structure to properly use and manage the training outputs.

1. Data and references for each exercise are in their respective folders Ex1, Ex2……
2. Under each exercise there are four (4) sub-folders
   - <ExerciseData> folder contains necessary data for doing the exercise
   - <ReferenceData> folder contains the completed data after running each steps of the exercises developed by Arbonaut Team. This data should be used as reference/input wherever indicated in the exercises.
   - <AdditionalReading> folder contains reference materials, papers, manuals related to the exercises.
   - <Output> folder is blank and is used to save all the outputs generated in each of the exercises.
3. <AdditionalReading> folder contains manuals and references of software and other related documents.
4. <Software> folder contains necessary Free and Open Source Software (FOSS) to be used in this module.

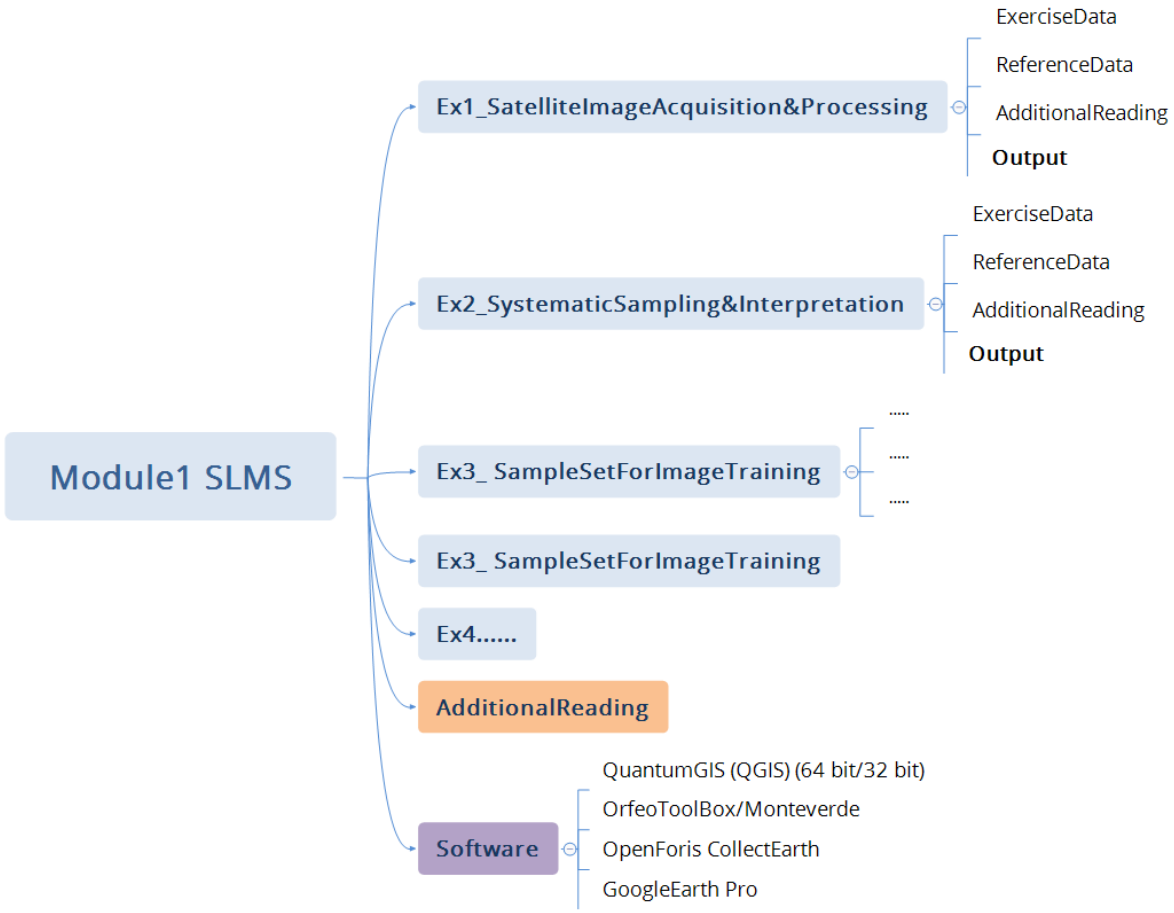


Figure 1 Training data files structure

1.5. Approach for SLMS – Land Use Land Cover Mapping

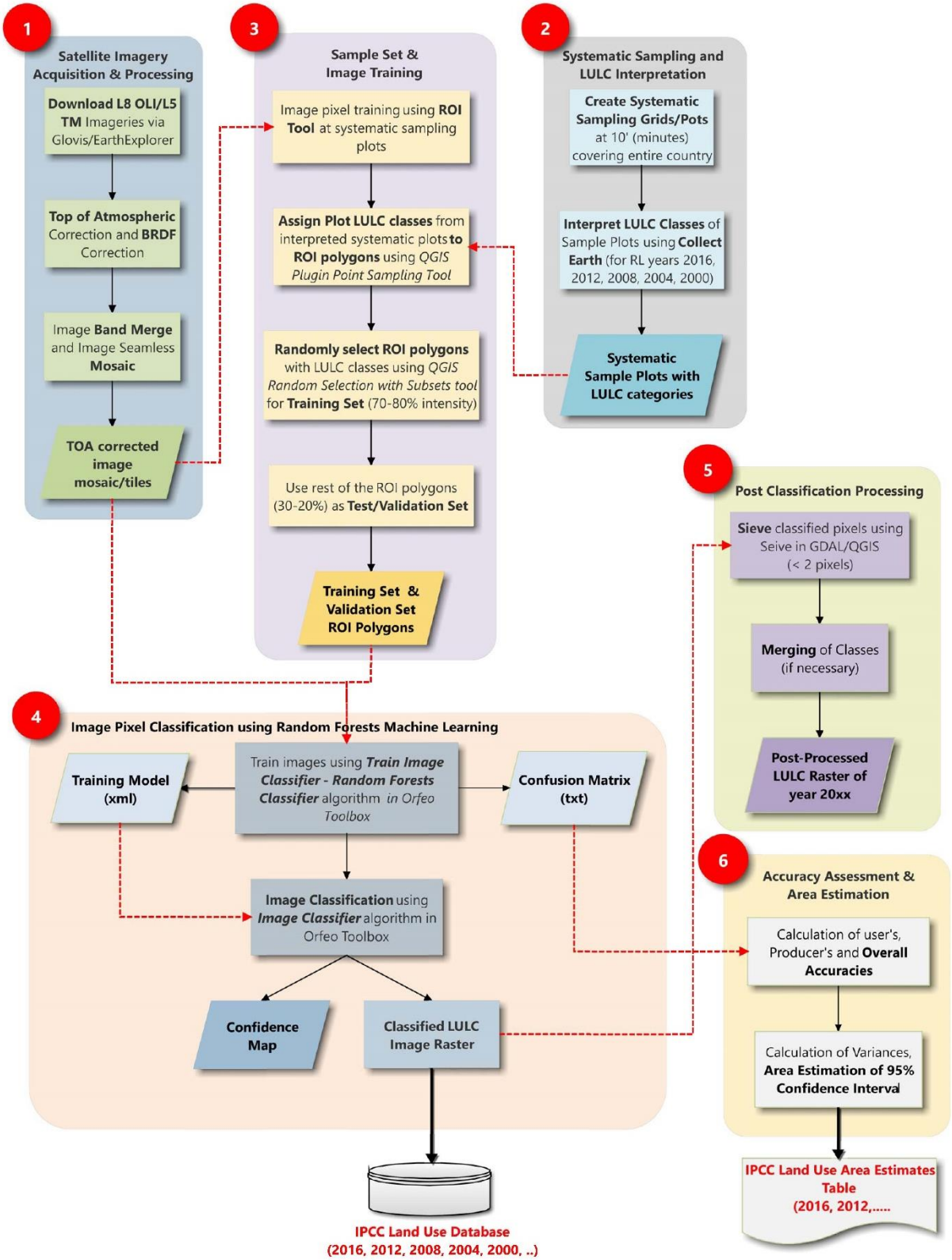
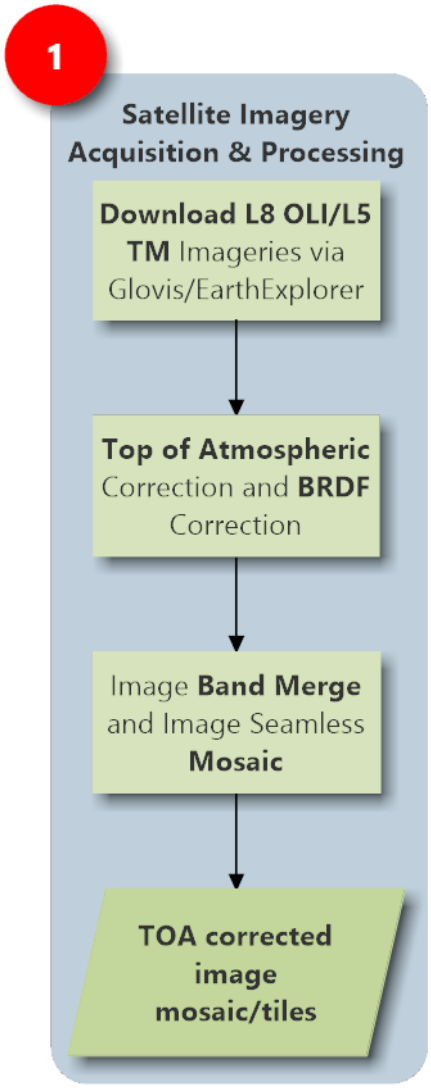


Figure 2 SLMS Land Use Land Cover Mapping Flowchart

2. Satellite Imagery Acquisition and Processing



Archived Landsat 5 TM/Landsat 8 OLI imageries can be downloaded from various sources, such as

- USGS Glovis (https://glovis.usgs.gov/next/#) OR
- EarthExplorer (https://earthexplorer.usgs.gov/)OR
- USGS Landsat Look (https://landsatlook.usgs.gov/viewer.html)
- EOS Land Viewer (https://lv.eosda.com)

For FRL/FREL in Pakistan, reference years are taken as 1996, 2000, 2004, 2008, 2012 and 2016

Top of Atmospheric (TOA) correction is done using Dark Object Subtraction (DOS) algorithm.

Bidirectional Reflectance Distribution Function (BRDF) topographic correction is done to remove terrain shadows using topographic normalization algorithm

Atmospherically (and topographically) corrected individual image bands need to be merged to get a single image scene (with Bands B2-B7) for L8 image

For multiple scenes covering certain area, individual band merged scenes need to be mosaicked with adjacent scenes after atmospheric (and topographic) corrections.

2.1. Exercise 1A Image Exploration and Downloading

2.1.1. Background

Availability of reference year archived Landsat 5 TM/Landsat 8 OLI images can be explored and downloaded using several free online tools such as

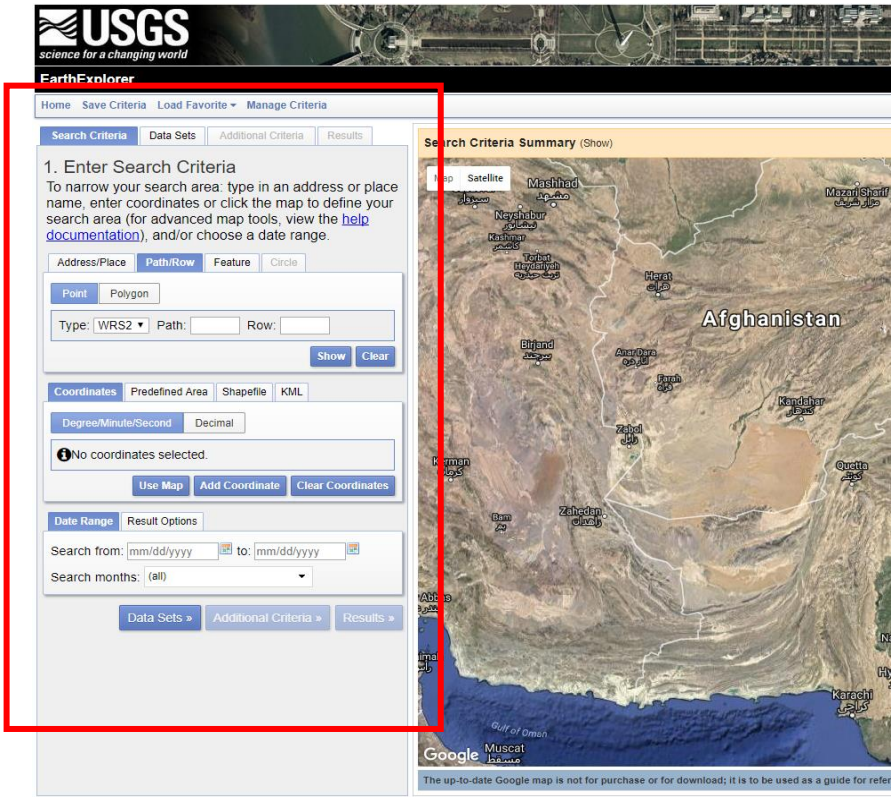
- USGS Glovis (https://glovis.usgs.gov/next/#)
- EarthExplorer (https://earthexplorer.usgs.gov/)
- USGS Landsat Look (https://landsatlook.usgs.gov/viewer.html)
- EOS Land Viewer (https://lv.eosda.com)

Among these, we will be using USGS Earth Explorer tool for exploring the availability of the imageries and subsequent downloading. Alternatively, any other tools mentioned above can be used. These tools have different interfaces for exploring and downloading the images. EarthExplorer allows to select and download the images in bulk. EarthExplorer requires user registration and authentication for downloading the imageries. User registration is free.

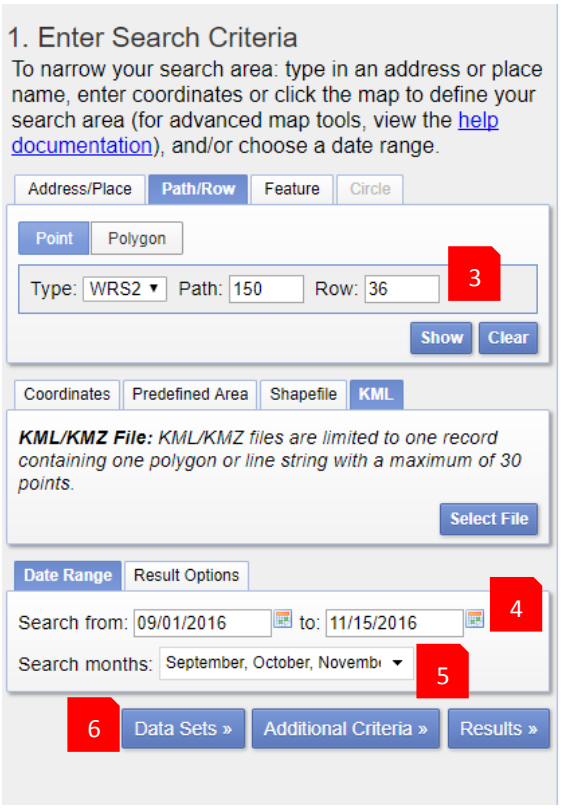
2.1.2. Exploring and Downloading Landsat Images

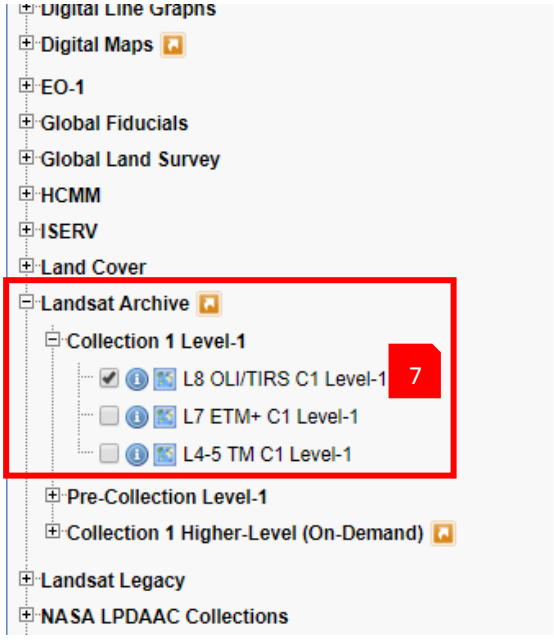
Step 1 Explore and Browse Landsat 8 OLI Images in EarthExplorer

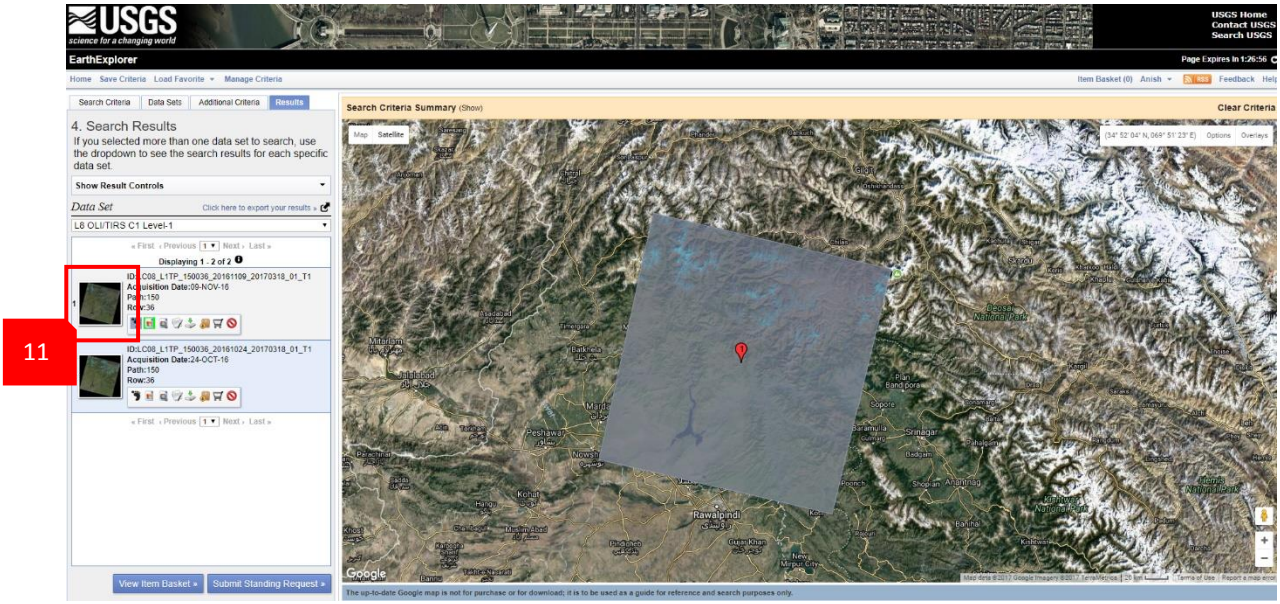
1. Go to USGS EarthExplorer web portal using https://earthexplorer.usgs.gov/ in your browser.
2. You can choose to browse all the images within a defined extent by polygon (shape file or kml file with maximum of 30 nodes that can be uploaded OR user drawn polygon on the map interface).
3. In this exercise, we will define specific Path and Row of Landsat image to be downloaded. Go to Path/Row tab and specify **Path: 150 and Row 36** covering parts of KP and AJK and use **Show** button to navigate to the location.

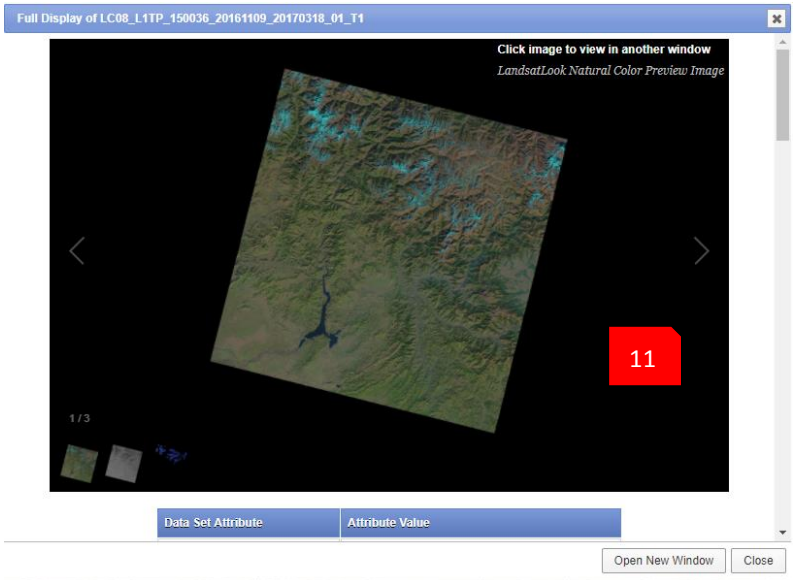


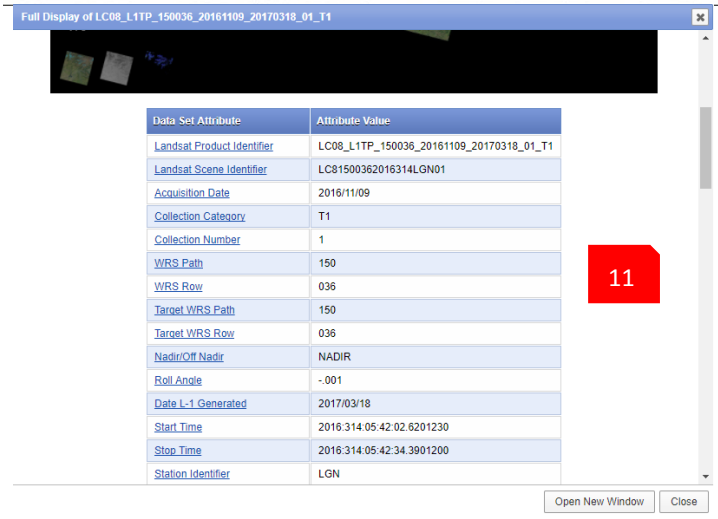
4. Define the search criteria using *Date Range*. Here we select date range from beginning of September to end of November (for all reference years), considering that these are post monsoon months with least cloud cover as well as leaf-on season for deciduous tree species in the forests.
5. You can define *Search months* as September, October and November.
6. Once search criteria is defined, go to *Data Sets* to select Landsat 8 OLI/Landsat 5 TM dataset to download.
7. Browse to *Landsat Archive > Collection 1 Level-1* and select ***L8 OLI/TIRS C1 Level-1***
8. You can choose to define *Additional Criteria* for cloud cover (<10% preferably), Data Type (Precision Terrain Correction Level-1TP) and other parameters.
9. Go to *Results* to see the results of the search.
10. You will see two images displayed which fulfills the criteria for Landsat 8 OLI scene 150 and path 36.
11. Preview of the image and its metadata can be viewed by clicking on the Image Icon



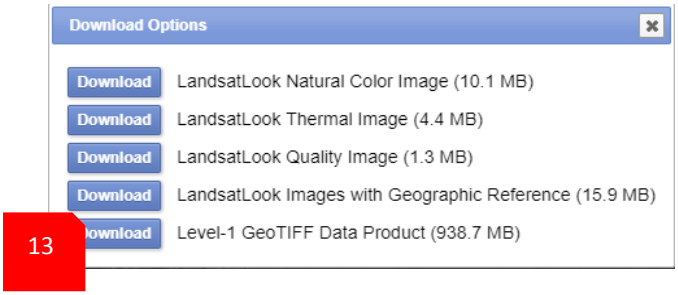


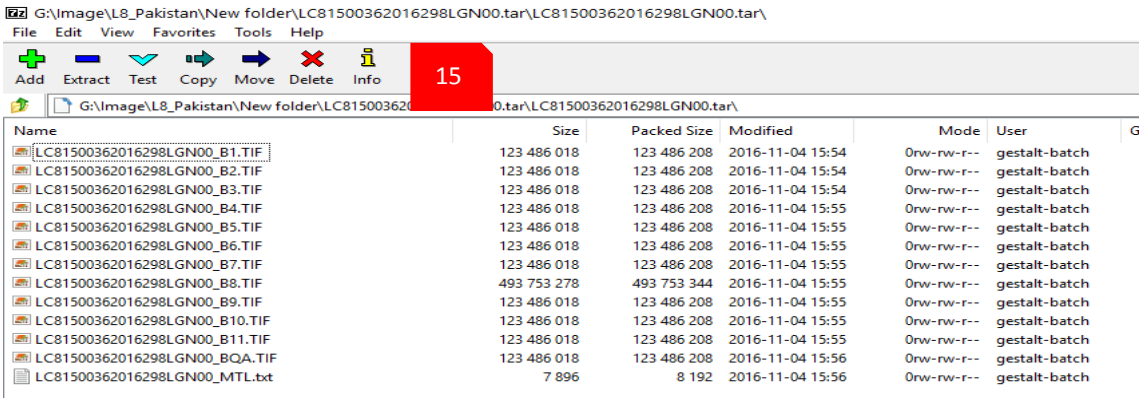

USGS
science for a changing world
EarthExplorer
Search Criteria
Data Sets
Additional Criteria
Results
4. Search Results
If you selected more than one data set to search, use the dropdown to see the search results for each specific data set.
Show Result Controls
Data Set
L8 OLI/TIRS C1 Level-1
Displaying 1 - 2 of 2
ID:LC08_L1TP_150036_20161109_20170318_01_T1
Acquisition Date:09-NOV-16
Path:150
Row:36
ID:LC08_L1TP_150036_20161024_20170318_01_T1
Acquisition Date:24-OCT-16
Path:150
Row:36
View Item Basket »
Submit Standing Request »
Search Criteria Summary (Show)
Clear Criteria
11


Full Display of LC08_L1TP_150036_20161109_20170318_01_T1
Click image to view in another window
LandsatLook Natural Color Preview Image
1 / 3
11
Data Set Attribute
Attribute Value
Open New Window
Close


Full Display of LC08_L1TP_150036_20161109_20170318_01_T1
Data Set Attribute
Attribute Value
Landsat Product Identifier
LC08_L1TP_150036_20161109_20170318_01_T1
Landsat Scene Identifier
LC81500362016314LGN01
Acquisition Date
2016/11/09
Collection Category
T1
Collection Number
1
WRS Path
150
WRS Row
036
Target WRS Path
150
Target WRS Row
036
Nadir/Off Nadir
NADIR
Roll Angle
-.001
Date L-1 Generated
2017/03/18
Start Time
2016:314:05:42:02.6201230
Stop Time
2016:314:05:42:34.3901200
Station Identifier
LGN
11
Open New Window
Close

12. Selected image can be downloaded individually or multiple images can be downloaded in bulk using USGS Bulk Download Application -BDA (https://earthexplorer.usgs.gov/bulk/)
13. For downloading individual scenes, use *Download Options* icon, Download Options window pops up with options to download various products for the selected image scene. Select **Level-1 GeoTIFF Data Product** to download the full scene. The data is downloaded in TAR format.
14. Compressed TAR file can be extracted using any of compression applications such as 7-Zip (http://www.7-zip.org)
15. The downloaded TAR file need to be unzipped twice. In the first unzip, the file is extracted as "tar" file. This file needs to be unzipped again to extract all the 11 bands, quality assessment raster and the metadata MTL file as shown.





For this exercise Landsat 8 OLI image 150-36 is already downloaded and available in the folder //..L8_Image/150-36/

2.2. Exercise 1B Atmospheric and BRDF Corrections

Atmospheric correction is done to remove haze and other anomalies using Dark Object Subtraction (DOS) method and convert Surface Radiance to Top of Atmosphere (TOA) Reflectance in each image pixel. This process, however, do not remove the effects of terrain illuminations and shadows in the higher elevation regions. Bidirectional Reflectance Distribution Function (BRDF) corrections is applied to minimize/remove the effects of terrain on the pixels.

Data Used

SN	Datasets	Sources	Descriptions
1	L8 OLI image scene 150-36 *…//L8_Image/150-36/*	USGS EarthExplorer https://earthexplorer.usgs.gov/	Landsat 8 OLI image scene covering part of AJK and KP
2	Digital Terrain Model (DTM) 1 arc-second (≈30m) clipped using L8 image scene footprint vector file (DEM_150-36.tif)	Shuttle Radar Topography Mission (SRTM) 1 Arc-Second Global DEM https://lta.cr.usgs.gov/SRTM1Arc	Digital Terrain Model (DTM)

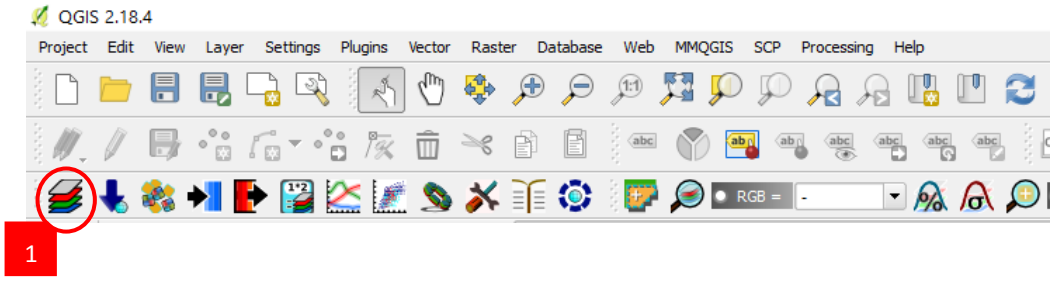
Software Used

- Quantum GIS (QGIS) Ver 2.18 (or older) - Open source GIS software
- Semi-Automatic Classification Plugin in QGIS
- SAGA Processing Tool in QGIS

2.2.1. Atmospheric Correction and Topographic Correction

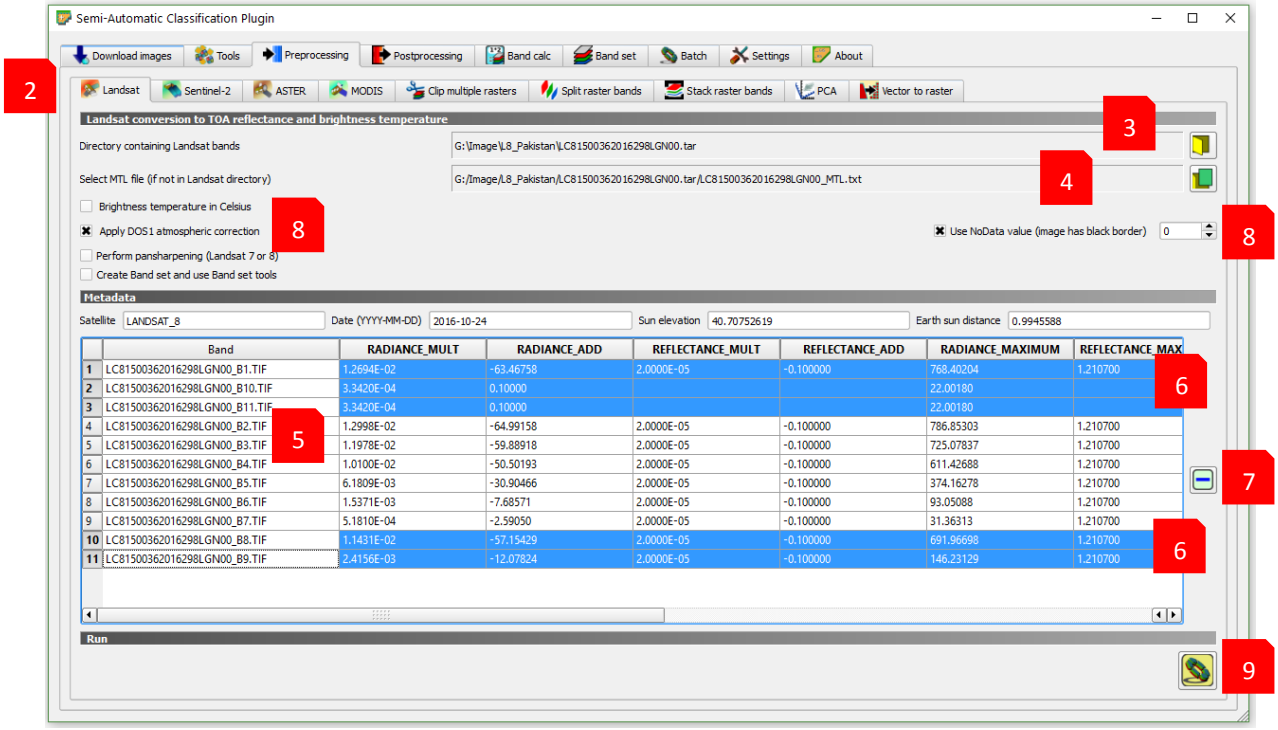
Step 1 Converting Landsat Surface Radiance to TOA Reflectance

1. Use *Semi-Automatic Classification Plugin > Band Set Tool* icon to open the plugin window interface. *[Note: Make sure that the latest version of Semi-Automatic Classification Plugin tool is installed using Plugin in QGIS]*

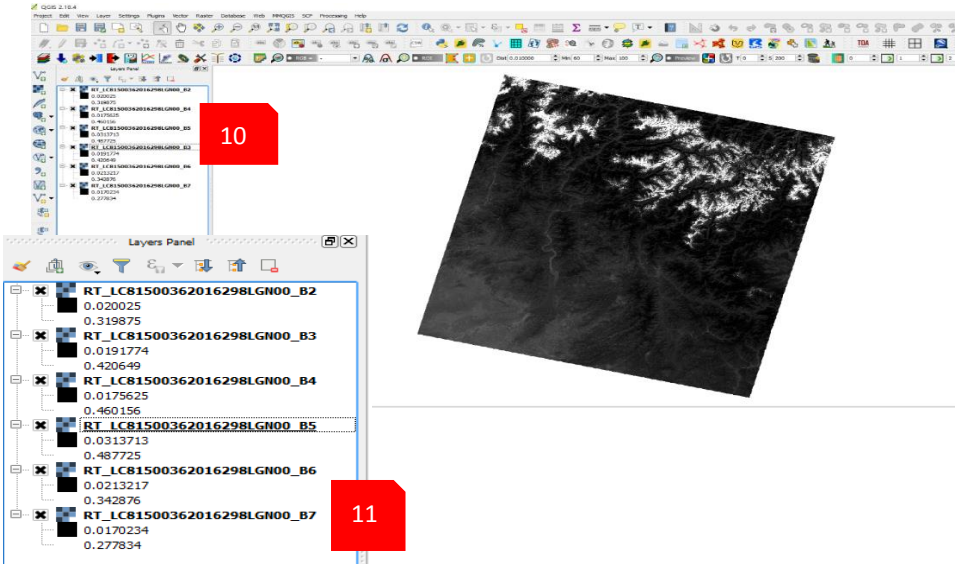


2. Go to *Preprocessing tab* and *Landsat* sub-group.
3. Browse the directory where Landsat 8 OLI images have been downloaded and unzipped.
4. Select the MTL (metadata) text file accompanying the Landsat image scene. This file contains the metadata of the image scene and can be browsed using any text editor.
5. The list of L8 OLI bands appears with their parameters read from the MTL file along with the date of image elevation, sun azimuth and earth sun distance.

6. Select bands B1, B8, B9, B10 and B11 to be removed from further processing. *[Note: Refer Landsat 8 OLI documentation for the characteristics of L8 Bands]*
7. Use *Delete Row* button to remote these bands from processing
8. Select *Apply DOS1 atmospheric correction* and *Use NoData value entering 0 (or -1)*
9. Once bands B1, B8, B9, B10 and B11 are removed, use *Run* located at the bottom right of the interface. Select a folder to create and save atmospherically corrected image bands. This might take few minutes depending on the processing capacity of your computer.
10. Once done, the atmospherically corrected image band are loaded in QGIS.

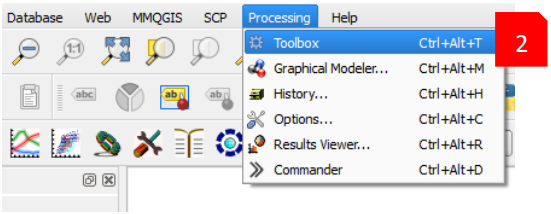


11. Rearrange the sequence of bands by selecting the image band layer and 'drag and drop' sequentially as B2, B3, B4, B5, b6 and B7 as shown. This is to make sure that the image bands are correctly sequenced for stacking as a band interleaved file.

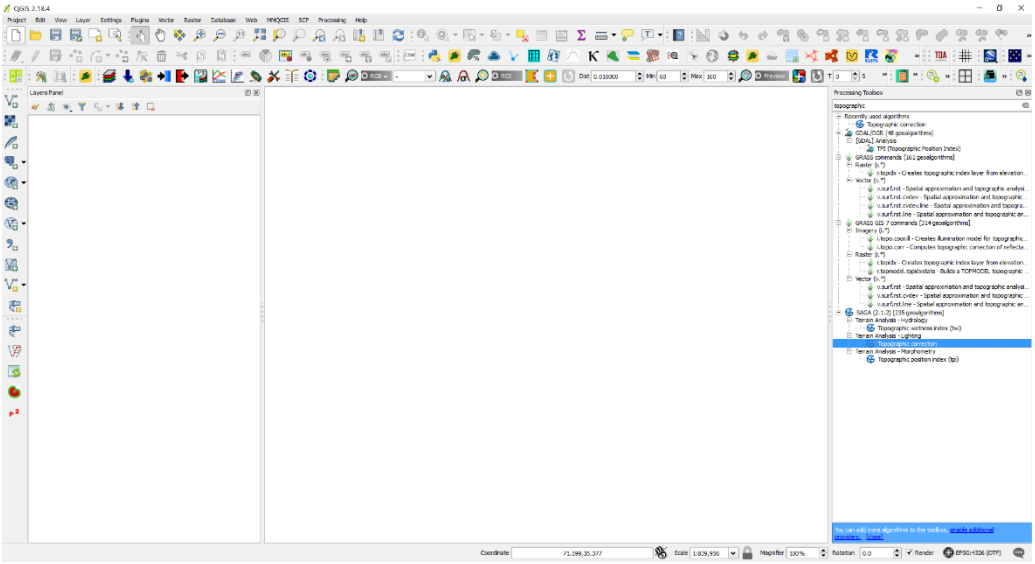


Step 2 Topographic (BRDF) Correction

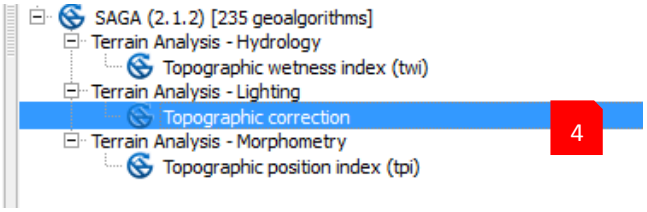
1. Prior to Topographic Correction for BRDF, make sure the Coordinate Reference System, extent and pixel sizes for image scenes and DEM are the same. In case, if they differ, ***Clip*** the rasters using image footprint *[Note: Clipping can be done using Semi-Automatic Classification Plugin>Preprocessing>Clip multiple rasters tool OR Raster>Extraction>Clipper tool OR GDAL Clip Raster by Mask Layer].*
2. Go to *QGIS>Processing Menu* and select *Toolbox.*

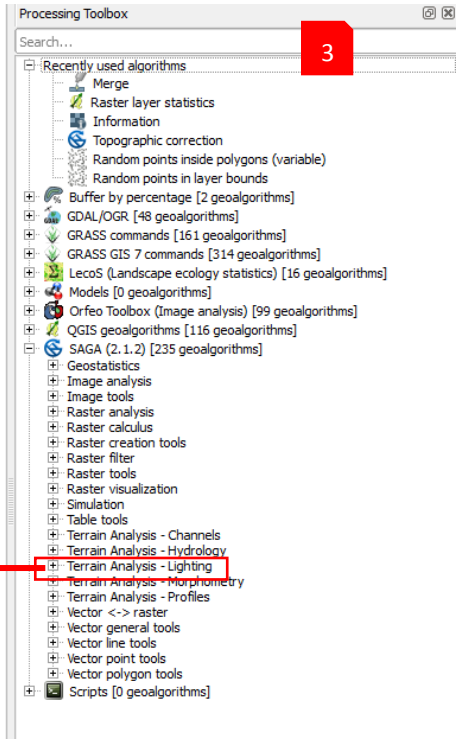


3. *Processing Toolbox* panel is activated and docked on the right panel of QGIS interface

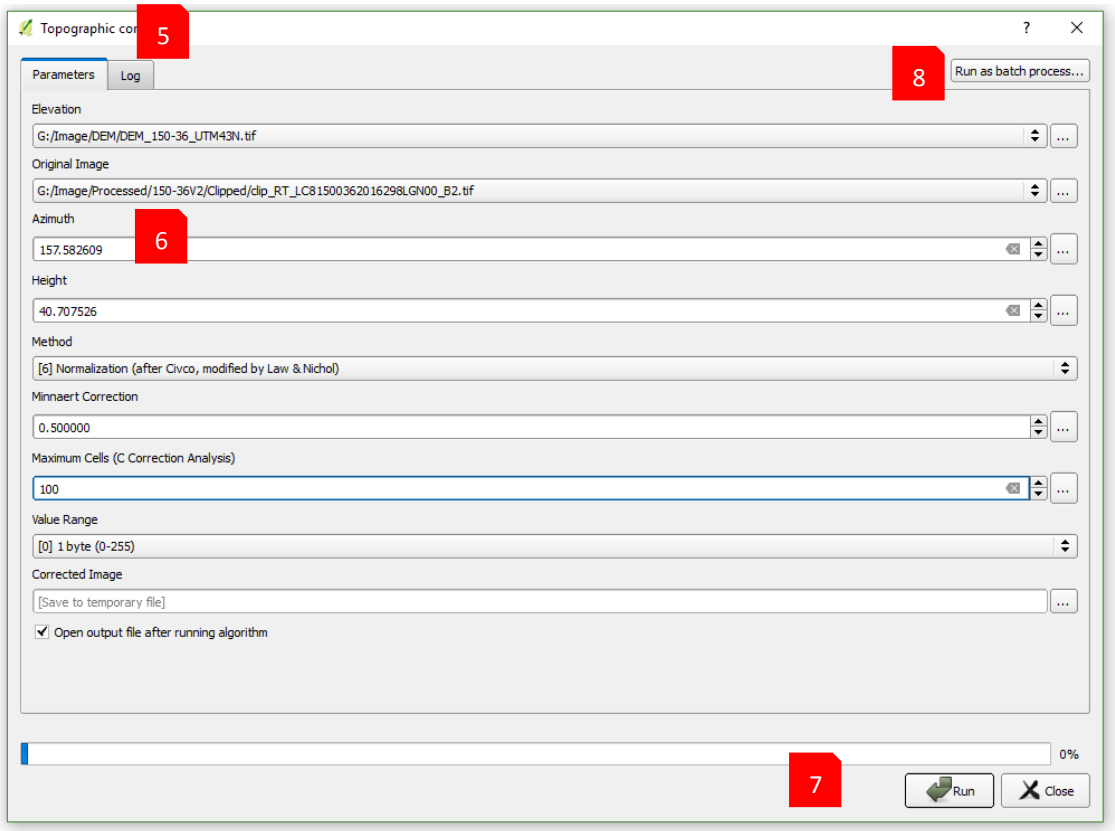


4. Browse to SAGA Geoalgorithms and under Terrain Analysis – Lighting, select *Topographic correction*
5. *Topographic Correction* window interface appears as shown. You can choose to apply topographic corrections individually to all the bands OR use *Run as batch process* to run the process in patch processing mode.





6. Following inputs are required:

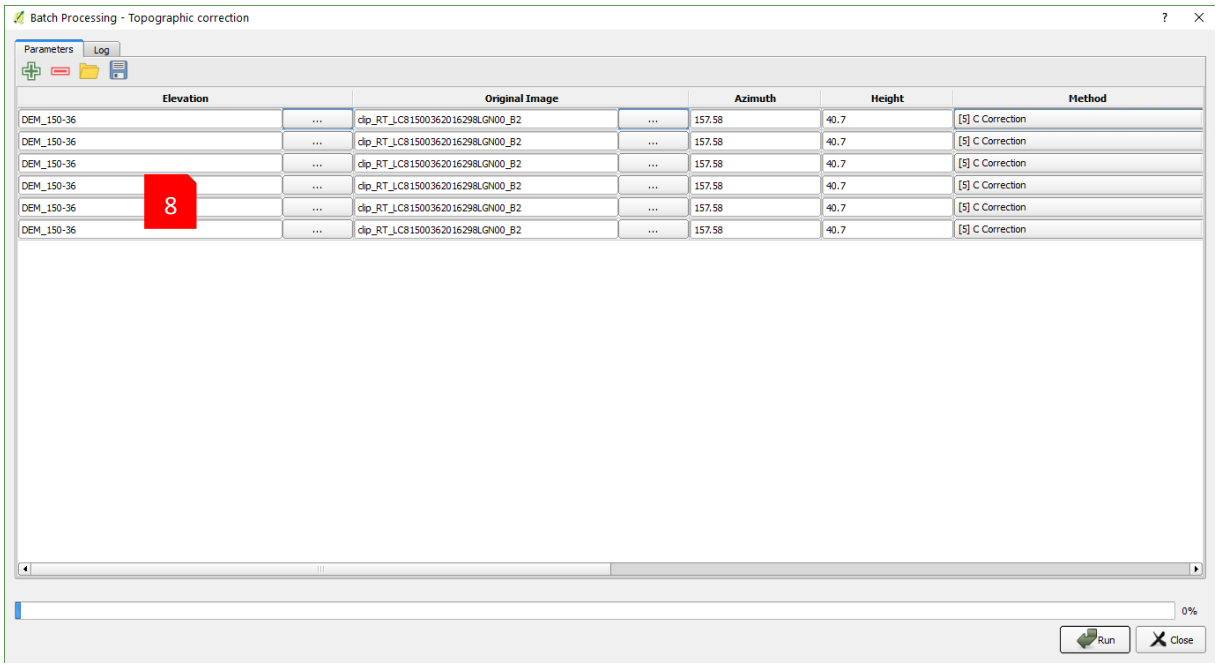


- Elevation: DEM (clipped to the boundary of image)
- Original Image: L8 image band to be corrected

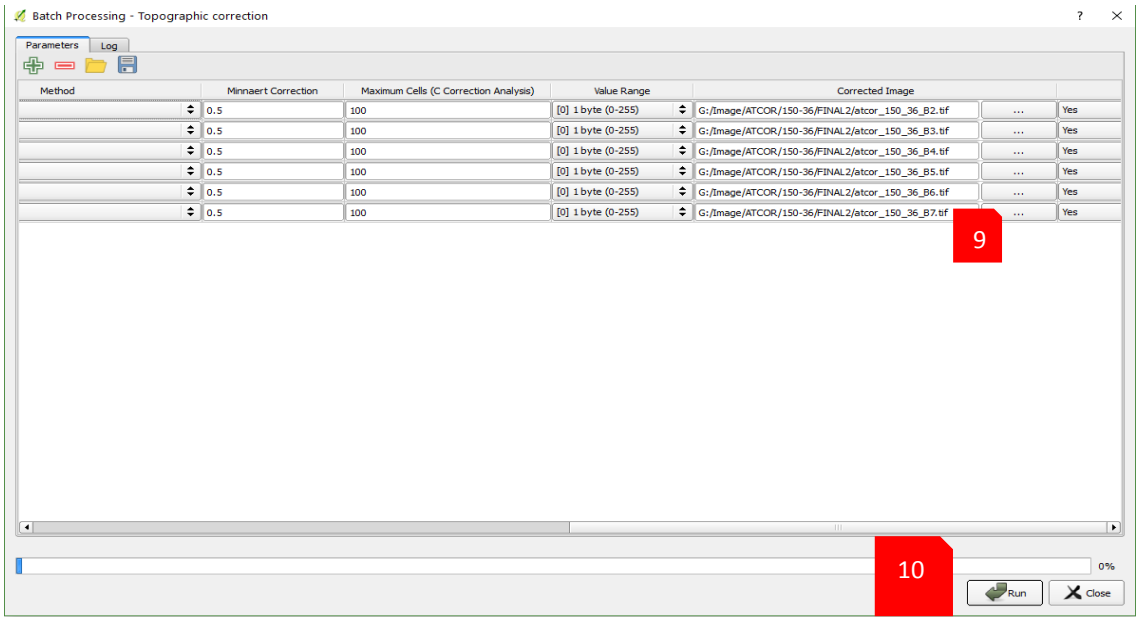
- Azimuth: Sun azimuth angle of the image (obtained from image metadata MTL file, use text editor to browse the metadata file as shown), in this case **157.582609** degrees for image scene 150-36
- Height: Sun elevation angle (obtained from image metadata MTL file, use text editor to browse the metadata file as shown), in this case **40.7075** degrees for image scene 150-36
- Method: *Normalization*
- Minnaert Correction: 0.5
- Maximum Cells: 100
- Value Range: [0] 1 byte (0-255)
- Corrected Image: Browse to the location to save the corrected image and the provide appropriate filename, in this case ***atcor_150-36_b2.tif***

```
CLOUD_COVER = 4.58
CLOUD_COVER_LAND = 4.58
IMAGE_QUALITY_OLI = 9
IMAGE_QUALITY_TIRS = 9
TIRS_SSM_MODEL = "FINAL"
TIRS_SSM_POSITION_STATUS = "ESTIMATED"
ROLL_ANGLE = -0.001
SUN_AZIMUTH = 157.58260856
SUN_ELEVATION = 40.70752619
EARTH_SUN_DISTANCE = 0.9945588
GROUND_CONTROL_POINTS_VERSION = 4
GROUND_CONTROL_POINTS_MODEL = 464
GEOMETRIC_RMSE_MODEL = 6.369
GEOMETRIC_RMSE_MODEL_Y = 4.694
GEOMETRIC_RMSE_MODEL_X = 4.305
GROUND_CONTROL_POINTS_VERIFY = 152
GEOMETRIC_RMSE_VERIFY = 3.080
```

7. *Run* the algorithm to process topographic correction for a single image band.
8. All the bands can be corrected simultaneously using *Run as batch process..* entering necessary parameters in the inputs.



9. Output corrected file names can be automatically generated. However, file names need to be renamed appropriately to reflect the bands 2-7. For example, “atcor_150-36_B1” generated automatically should be renames to “atcor_150-36_B2” to indicate the image band 2. Likewise, other corrected images need to be renamed from B3-B7 accordingly to indicate the appropriate the bands.
10. Once all the parameters are set, *Run* the algorithm.



Caution!

Due to memory resource needed for atmospheric correction, QGIS/SAGA might crash during processing. Therefore, Topographic BRDF Correction may not function as desired for C Correction. Alternatively, other algorithm may be chosen instead of C Correction, for e.g. Normalization

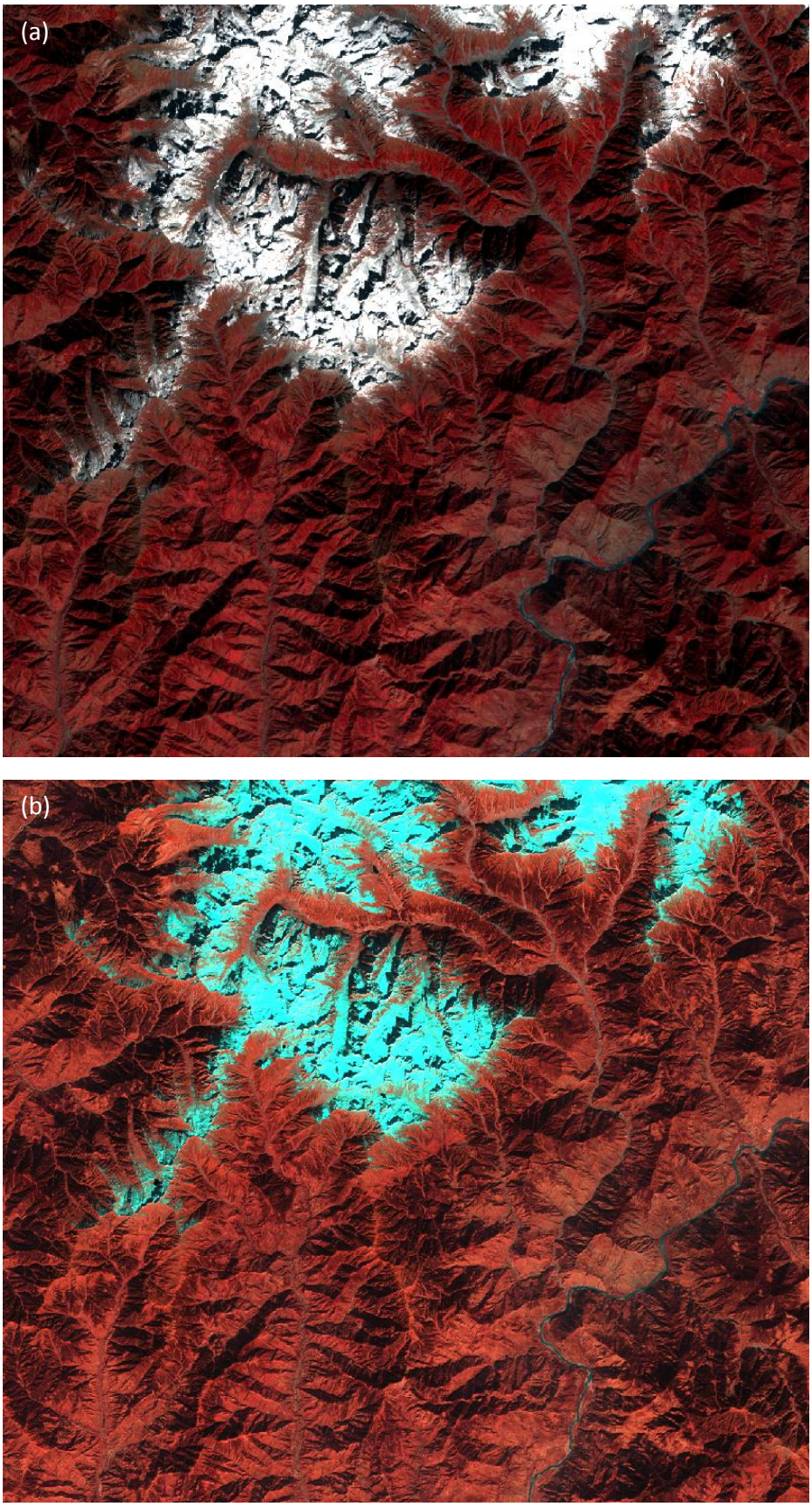


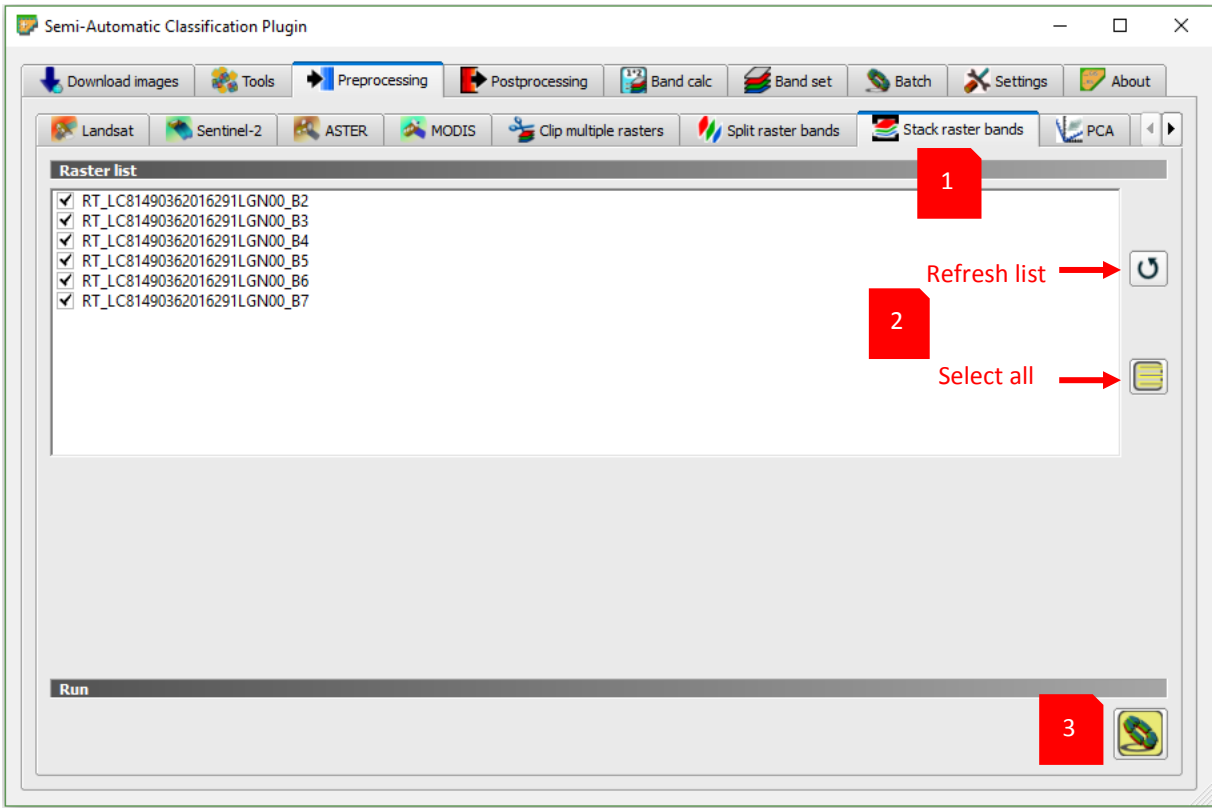
Figure 3 Landsat 8 scene (a) before topographic correction (b) after topographic correction

Step 3 Band Merge/Stacking

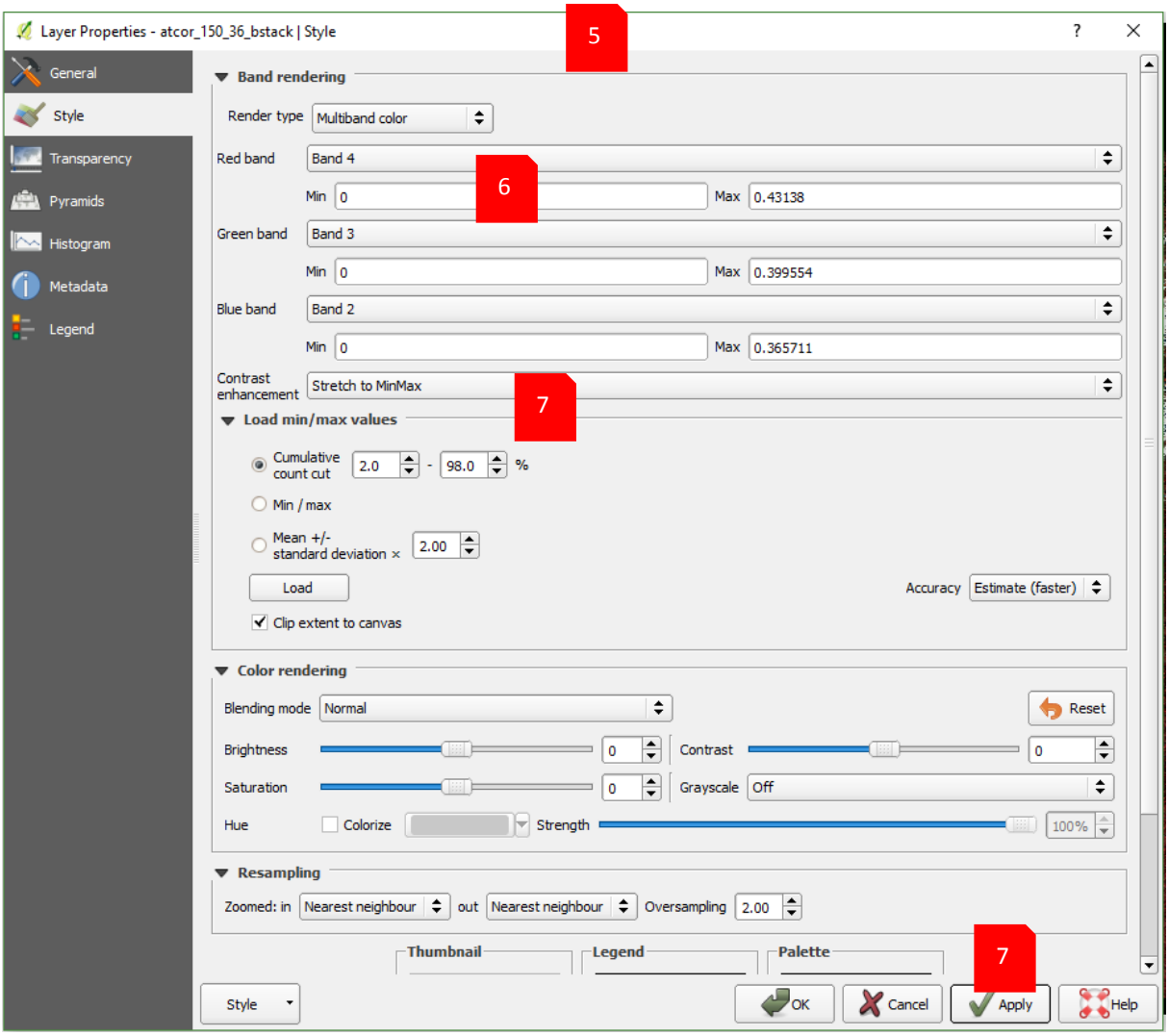
1. To merge the atmospherically corrected individual bands, use *Semi-Automatic Classification>Preprocessing>Stack raster bands*. *[Note: Before stacking the bands, make sure the band raster layers are sequentially arranged, if needed, drag the layers in the Table of Contents to arrange the band 2, 3, 4, 5, ,6 and 7 sequentially to ensure proper band stacking of L8 OLI i.e.*

   - Band 2 (0.45–0.51)
   - Band 3 (0.53–0.59)
   - Band 4 (0.64–0.67)
   - Band 5 (0.85–0.88)
   - Band 6 (1.57–1.65)
   - Band 7 (2.11–2.29)

2. Use *Refresh list* button to refresh the list of layers and *Select all* to select all the layers.
3. *Run* the process, browse and save the stacked image bands with appropriate names for e.g. “atcor_150_36_bstack.tif”, “bstack” indicating band stacked.



4. Add the band stacked image in QGIS using *Add Raster Layer* icon/menu.
5. Go to *Layer Properties by double clicking* the stacked layer or *Right Mouse button>Properties* context menu.
6. Select *Multiband color* and change the band sequence B432 for VNIR composition, B543 for SWIR composition, B321 for natural color composite as appropriate.
7. Use *Stretch to MinMax* contrast enhancement option, selecting any pf the three methods to stretch the histogram of the image and *Apply.*



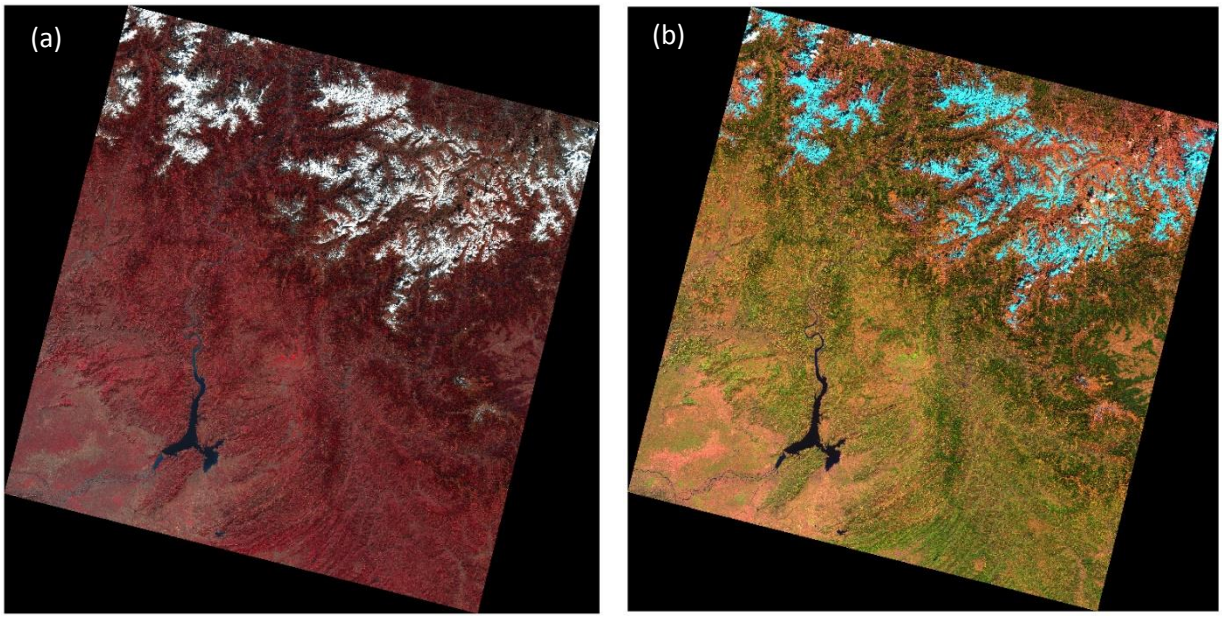
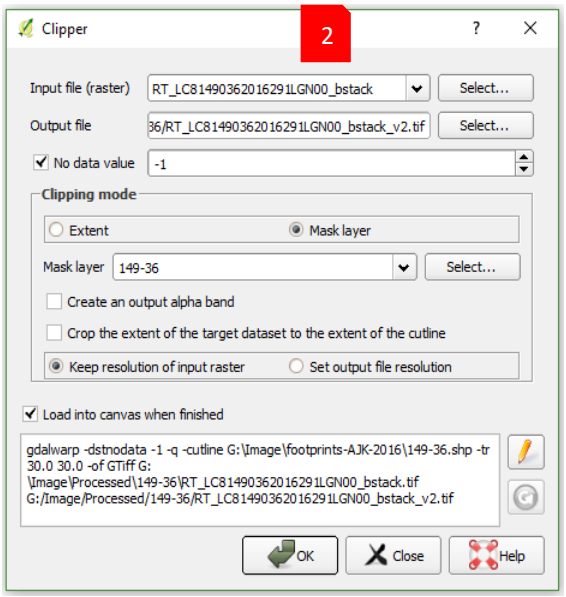
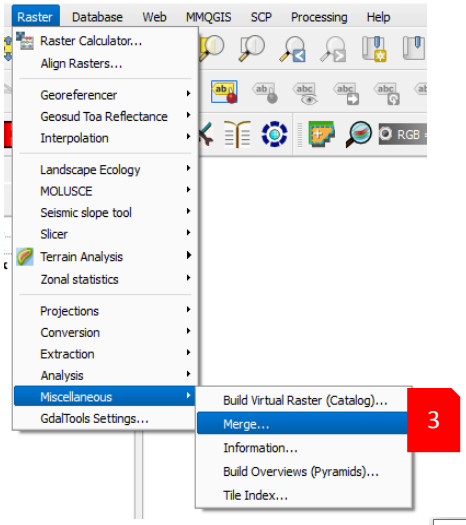


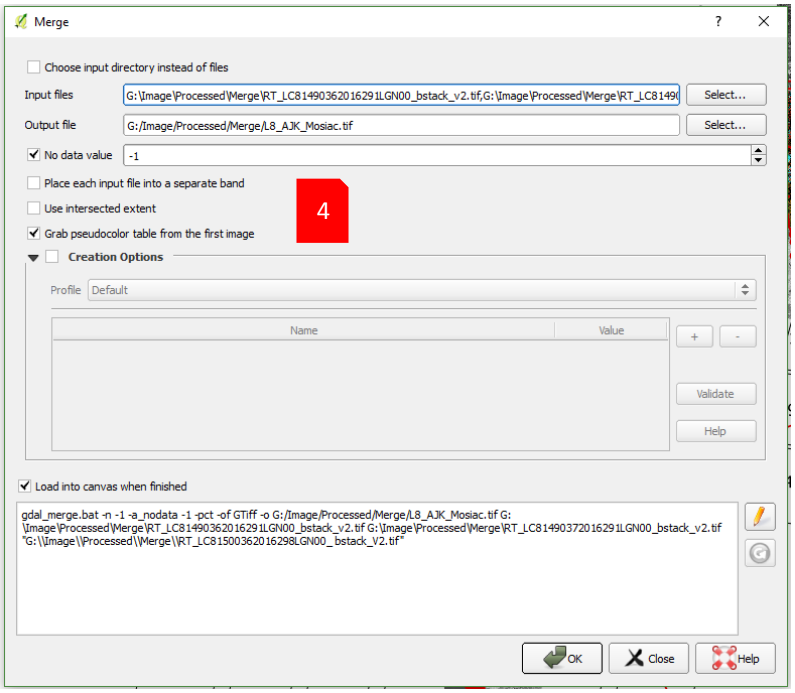
Figure 4 Band stacked L8 Image (a) B432 NIR composite (b) B543 SWIR composite

Step 4 Image Mosaicing

1. Add the band stacked image scenes in QGIS using *Add Raster Layer*. (Refer Figure 5Figure 4)
2. Prior to mosaicing the images, clip the images using their respective footprint vector data to remove the 'no-data' pixels (black pixels in the extent) using *Raster>Extraction>Clipper* menu with parameters as shown. (Refer Figure 6)
3. Use Raster>Miscellaneous>Merge the clipped image scenes.
4. Browse the images to be mosaiced, use -1 for *No data value* and select *Grab pseudocolor table from the first image. Browse and provide the filename* for mosaic and run the process pressing *OK* button.







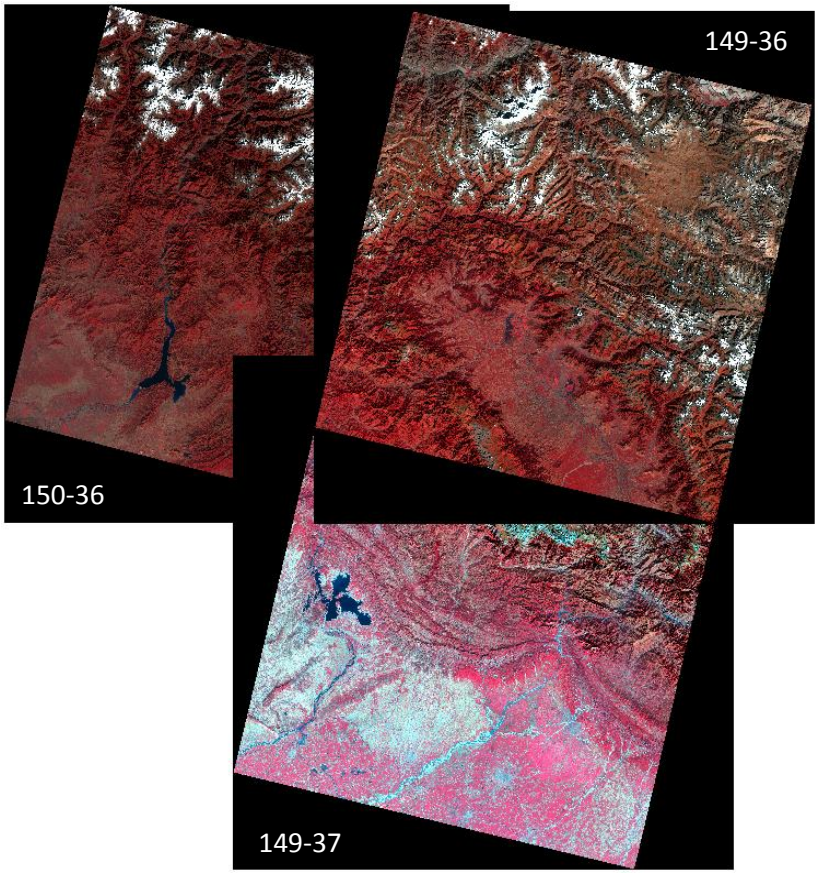


Figure 5 Individual atmospherically corrected image scenes

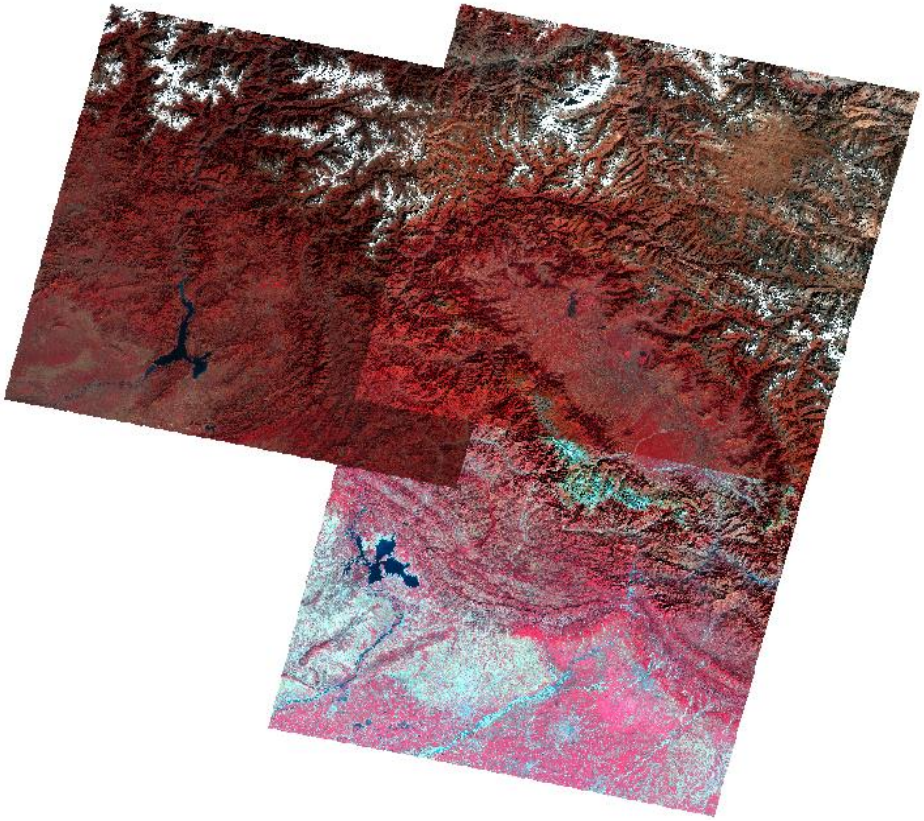


Figure 6 Clipped image scenes

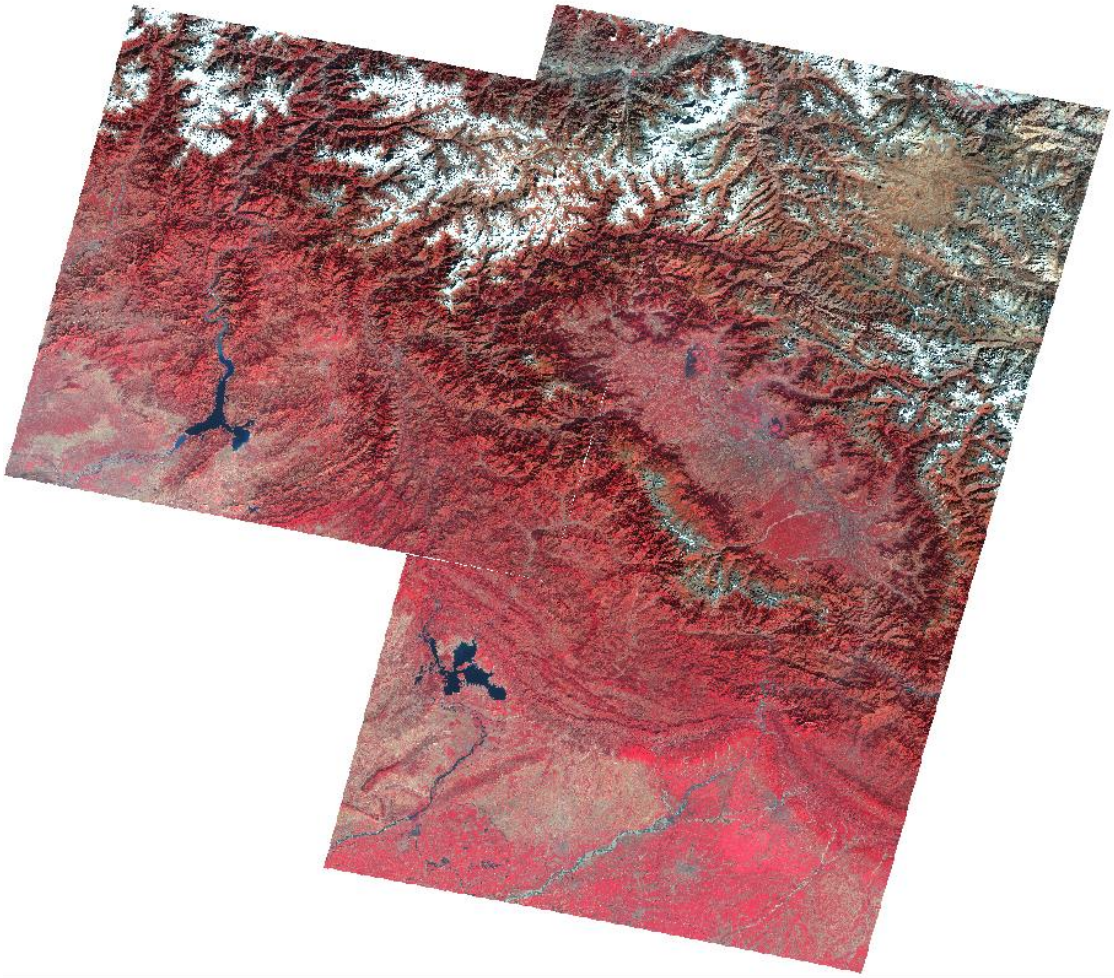
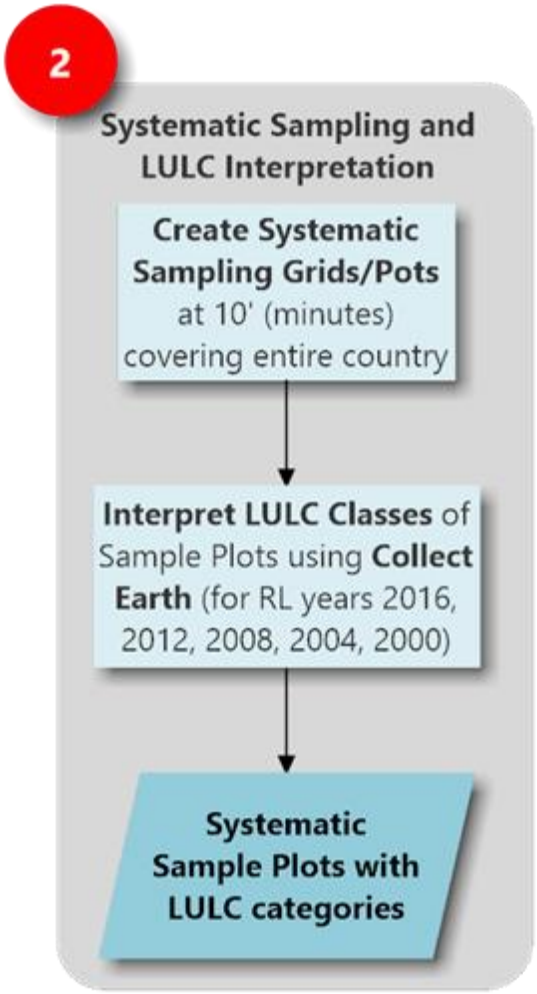


Figure 7 Mosaiced image scenes

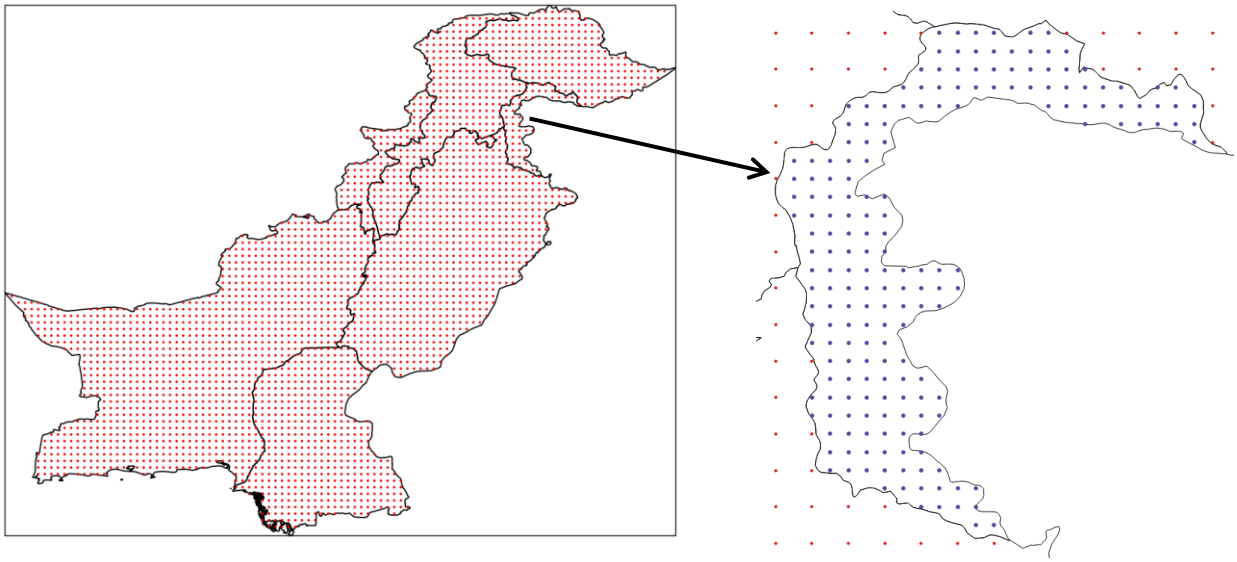
3. Systematic Sampling and LULC Interpretation



Systematic Sampling Plots are generated at 10'x10' within a defined Extent bounding the national boundary. Plots are generated in QGIS Research Tools, Regular Points

Densify the plot, if the sampling plots are not sufficient for provinces. Densification can be done at 5' x5' or 2.5' x 2.5' within the same reference

Generated Sampling Plots are interpreted for LULC classes using customized data form in Collect Earth with very high-resolution satellite imagery available in Google Earth/Bing Map platform for different reference years



3.1. Exercise 2A Sample Design for Visual Interpretation of Systematic Grids

3.1.1. Background

Systematic sampling grids are generated to cover the territory of the Islamic Republic of Pakistan for first phase (Phase I) sampling for IPCC Land Use classes to be used for preliminary wall-to-wall land use map preparation and forest inventory design. The systematic grids and the sample plots are created using Open Data and Open Source Tools and is a part of desktop based Satellite Land Monitoring System (SLMS) workflow to be implemented for REDD+ FRL/FREL and MRV assignment in Pakistan.

3.1.2. Generation of Sampling Grids/Plots[1]

Due to the non-availability (restrictions of use) of the recent land cover/land use spatial data for designing the sample set, independent systematic grids are generated. The systematic sample grids are initiated from 10' x 10' grids (≈16 km x 18 km) to cover the entire territory of Pakistan. The grids can be densified to 5' x 5' (≈8 km x 9 km) or 2.5' x 2.5' (≈4 km x 4.6 km) to represent the subnational levels i.e. at provincial level.

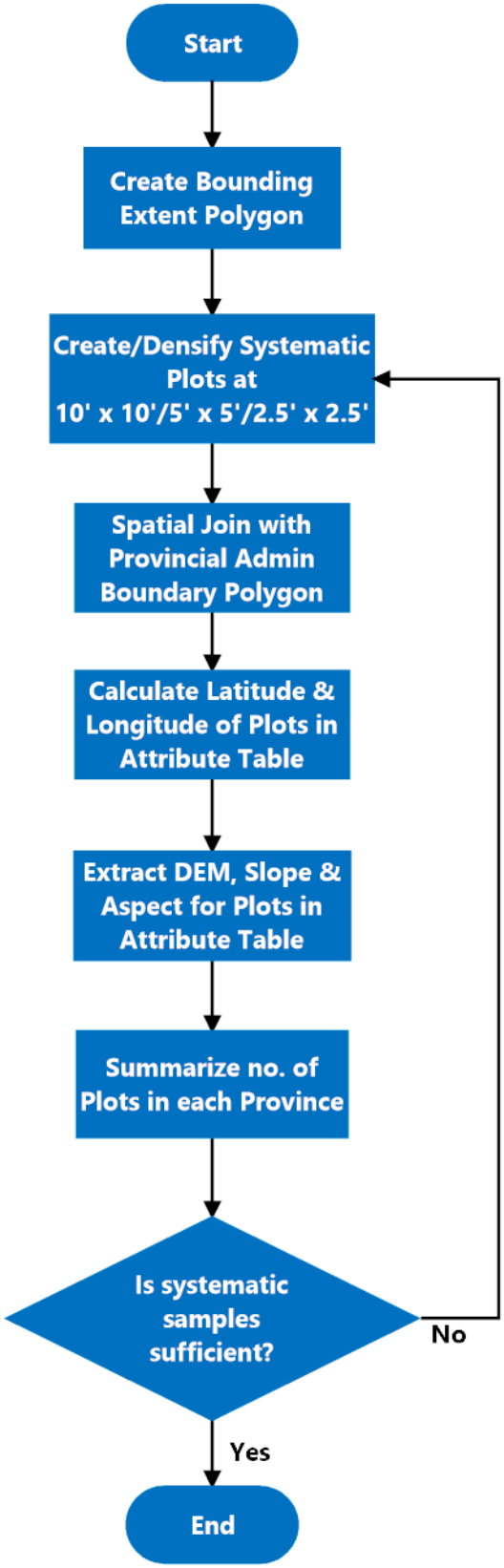
Data Used

SN	Datasets	Sources	Descriptions
1	Administrative Boundary Level 1 (PAK_adm1.shp) Administrative Boundary Level 2 (PAK_adm2.shp)	Global Administrative Areas http://www.gadm.org/	administrative areas (administrative boundaries)
2	Digital Terrain Model (DTM) 1 arc-second (≈30m) (PAK_DEM.tif)	Shuttle Radar Topography Mission (SRTM) 1 Arc-Second Global DEM https://lta.cr.usgs.gov/SRTM1Arc	Digital Terrain Model (DTM)

Software Used

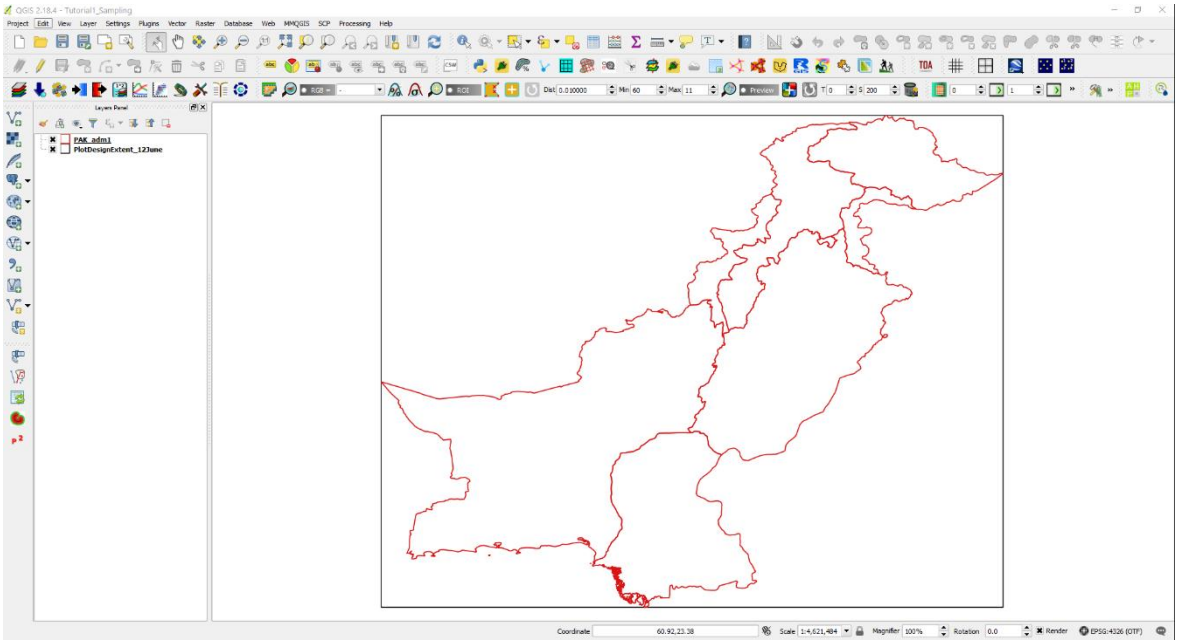
- Quantum GIS (QGIS) [2] Ver 2.18 (or older) - Open source GIS software

[1] Grids/Plots are generated in World Geodetic System 1984 (WGS84, EPSG:4326) coordinate reference system
[2] Latest version of QGIS can be downloaded from http://www.qgis.org/


Start
Create Bounding Extent Polygon
Create/Densify Systematic Plots at 10' x 10'/5' x 5'/2.5' x 2.5'
Spatial Join with Provincial Admin Boundary Polygon
Calculate Latitude & Longitude of Plots in Attribute Table
Extract DEM, Slope & Aspect for Plots in Attribute Table
Summarize no. of Plots in each Province
Is systematic samples sufficient?
No
Yes
End

Step 1 Creating a Bounding Extent Polygon

1. Create a boundary extent rectangle using the national administrative boundary data (data source 1) in QGIS using menu *Vector > Research Tools> Polygon from Layer Extent. [Note: Make sure that the SRS is in WGS84 (EPSG:4326) prior creating the extent]*



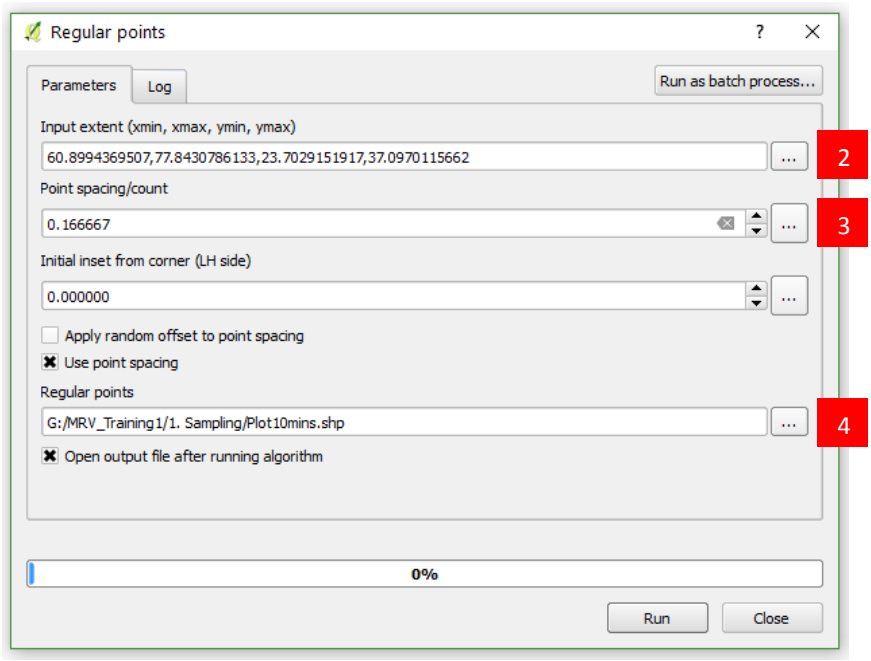
2. Save the extent file as **PlotDesignExtent.shp** in an appropriate directory.
3. A polygon is created with the following extent and coverage

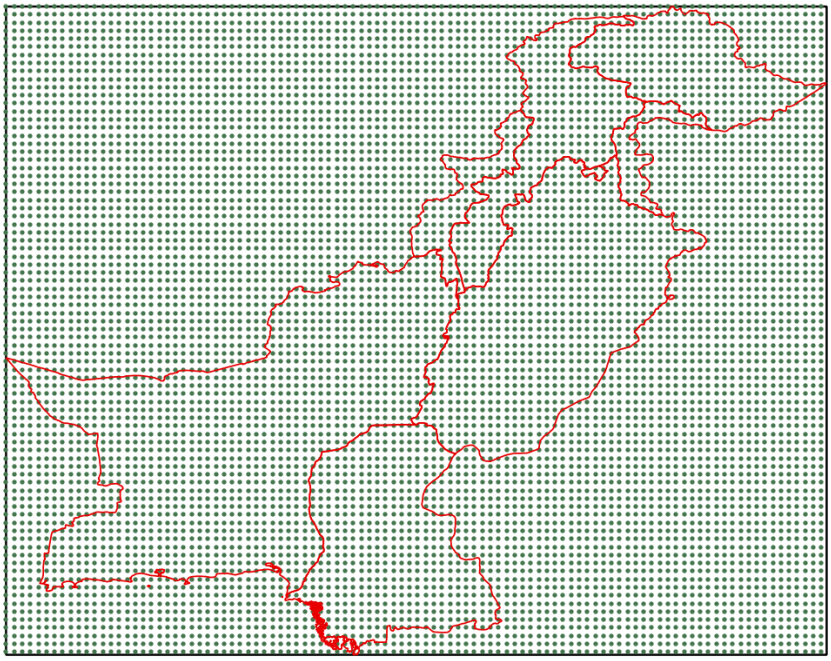
MinX	60.89944 degrees
MinY	23.70292 degrees
MaxX	77.84308 degrees
MaxY	37.09701 degrees
Area	2,428,061 sq.km

This extent polygon is used as the sample design extent for wall-to-wall grid.

Step 2 Creating Systematic Grids

1. Create systematic grid at 10' x 10' using menu *Vector>Research Tools> Regular points*.
2. Use the above created extent **PlotDesignExtent.shp** as an Input Extent
3. Input point spacing. For 10', the spacing is 10'/60 deg = *0.16666667*. Leave the *Initial Inset from the Corner as 0.000* for avoiding any offset.
4. Save the plot file as **Plot10mins.shp.**



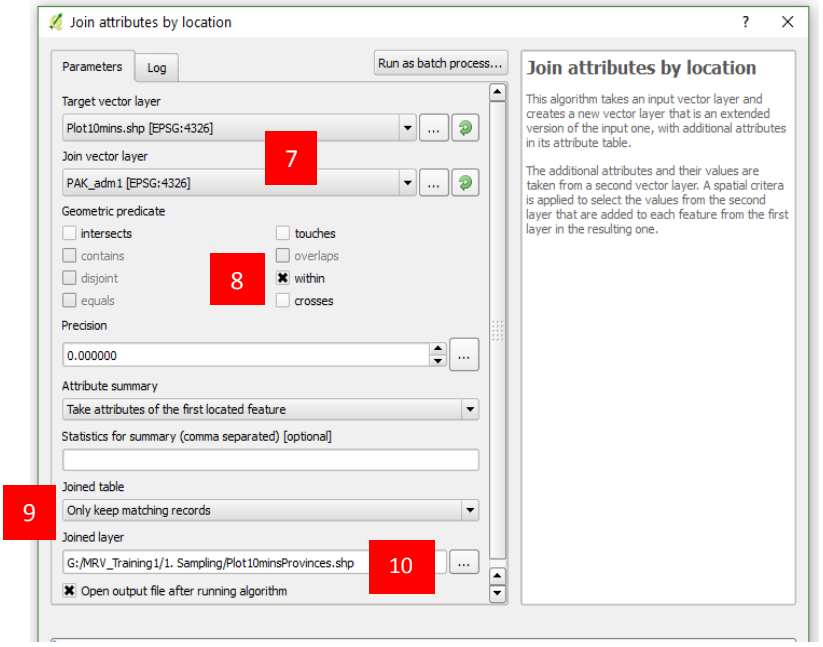


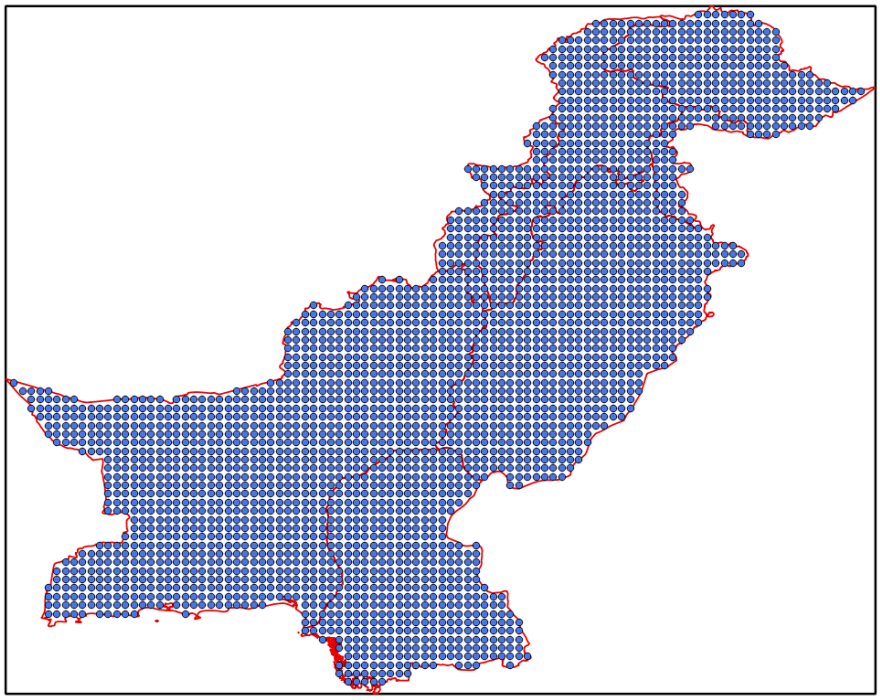
5. Regular grids at 10' x 10' (≈16 km x 18 km) are created within the defined extent as shown.
6. Select the plots within the territory of Pakistan using *Vector>Spatial Query>Spatial Query* (Spatial Query Plugin needs to be installed in QGIS) [Refer QGIS Manual]

OR

7. Alternatively, Spatial Join tool can be used to join the attributes of Provincial Administrative Boundary **(PAK_adm1.shp)** *with* plots (**Plot10mins.shp)** using *Vector>Data Management Tools>Join attributes by location*.

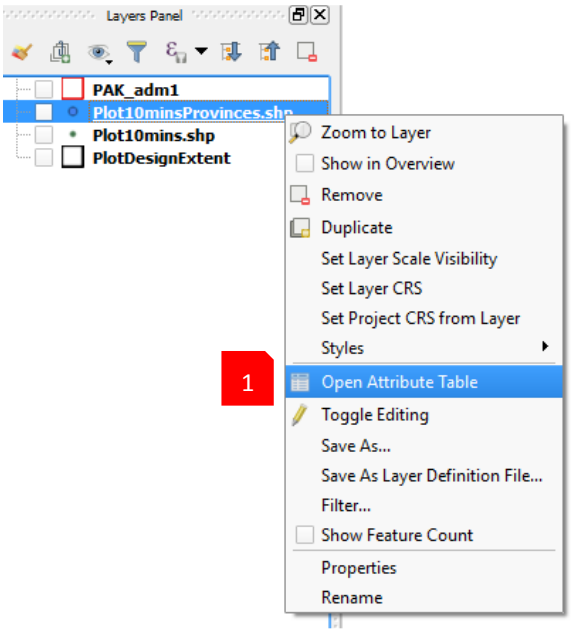
8. Select **Plot10mins.shp** as Target vector layer and Pak_adm1.shp as Join vector layer.
9. Select *within* Geometric predicate to joint all the plots falling within each of the administrative boundary unit in **PAK_adm1.shp** polygons.
10. In *Joined table,* select *Only keep matching records* to keep the plots that are within the Pakistan territory and omit the rest of the plots falling outside.
11. Save the joined layer as **Plot10minProvinces.shp.**
12. Plots wthin the territory of Pakistan (within **PAK_adm1.shp)** are queried along with the attributes of **PAK_adm1.shp** joined to the plot points in **Plot10minProvinces.shp.**
13. Go to the attribute table of the newly created **Plot10minProvinces.shp** and review. You can choose to delete fields that are not necessary. For interpretation of plots in CollectEarth, fields [plot_id, latitude, longitude, elevation, slope, aspect, region] are required.

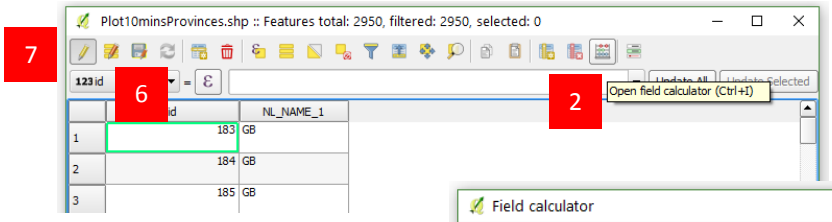


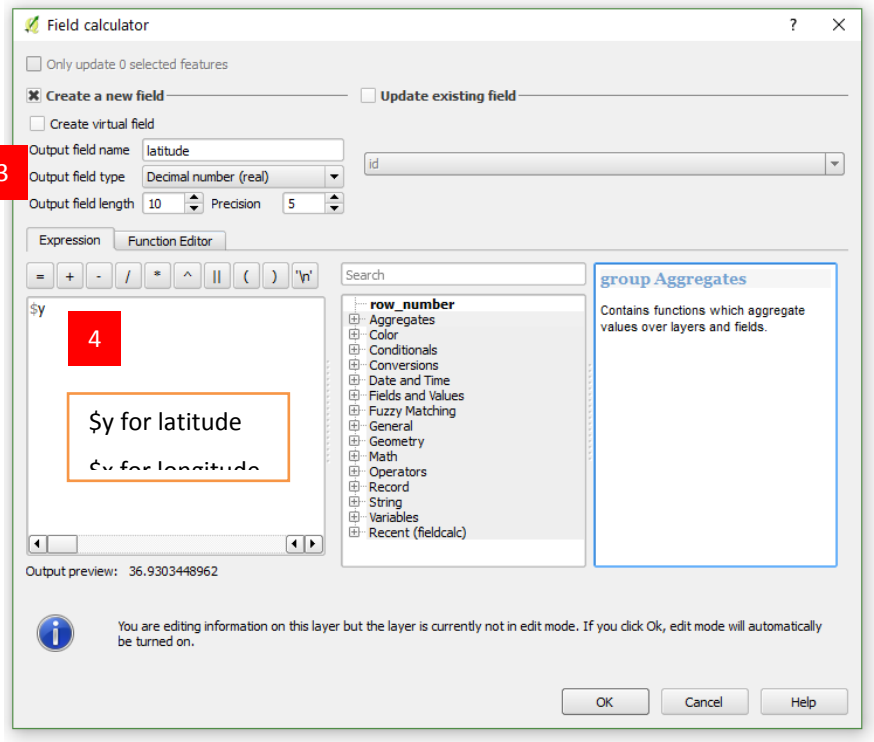


Step 3 Calculating Latitude & Longitude of Plot Locations

1. Latitude and Longitude of the plots can be calculated in the attribute table. Open the attribute table of **Plot10minProvinces.shp** by right clicking the mouse on the layer and selecting *Open Attribute Table.*
2. Open Field Calculator using icon on the top of Attribute Window or pressing *Ctrl+I keys.*
3. In the Field calculator, create a new field **latitude** (decimal number with length 10 and precision 5) as shown.
4. In the *Expression* type in ***$y*** for latitude
5. Repeat process #3 to create a new field ***longitude*** and type in ***$x*** in expression for longitude of each plots.
6. Save the table/datasets after this editing using the Save Edits icon of *Ctrl+S keys*.
7. Toggle the edits off using Toggle Edit icon or pressing *Ctrl+E keys* to stop the editing and save the edits.







Step 4 Extracting Terrain Parameters to Plots

Terrain parameters Digital Elevation Model (DEM), Slope and Aspect can be derived from freely available SRTM 1 arcsecond[1] (≈30m) or ASTER Global DEM[2] (30m) accessible to download from https://earthexplorer.usgs.gov/. Terrain parameter raster cell values can be extracted for stratification of sample plots.

Terrain Analysis

1. Merge individual SRTM/GDEM DEM tiles using *Raster>Miscellaneous>Merge.*
2. Generate Slope using *Raster>Terrain Analysis>Slope.*
3. Generate Aspect using *Raster>Terrain Analysis>Aspect.*

[Note: Make sure that the SRS of DEM is in UTM projection (UTM 42, EPSG:32642) prior to generating Slope and Aspect. Generation of Slope and Aspect might take longer due to large file size]

1. Add merged/mosaicked DEM, generated Slope, Aspect raster layers and design sample plots **Plot10minProvinces.shp** vector layers.

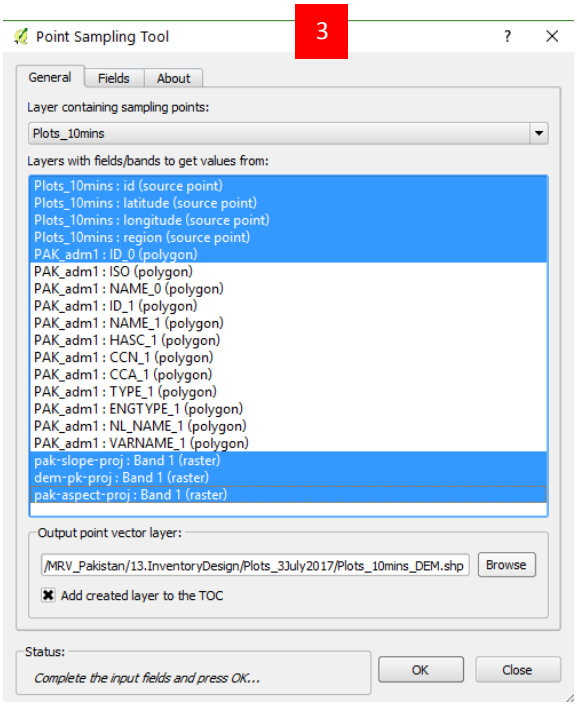
[Note: SRS of sample plots should be in the same projections as that of DEM, Slope and Aspect,

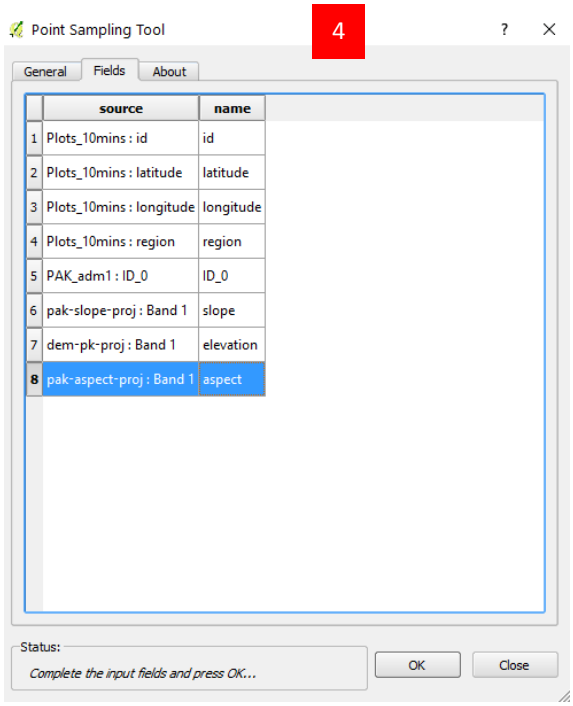
[1] NASA JPL. (2013). NASA Shuttle Radar Topography Mission Global 1 arc second [Data set]. NASA LP DAAC. https://doi.org/10.5067/measures/srtm/srtmgl1.003

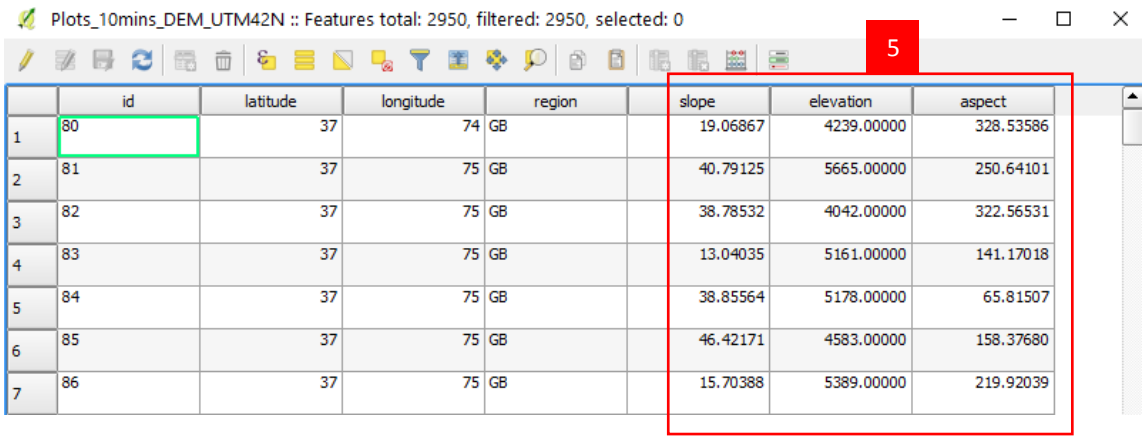
[2] NASA JPL. (2009). ASTER Global Digital Elevation Model [Data set]. NASA JPL. https://doi.org/10.5067/aster/astgtm.002

preferably in UTM 42N, reproject the Plot10minProvinces.shp to UTM42N, if needed]

2. Use *Plugin>Analysis>Point Sampling Tool* to extract raster pixel values as attribute fields in the **Plot10minProvinces.shp.**
3. In *General tab*, select the attribute fields of Sample Plots and raster layers for DEM, Slope and Aspect for e.g. *dem-pak-proj: Band 1 (raster)* as shown.
4. You may choose to rename the attributes of the Sample Plots upon extracting the raster values of DEM, Slope and Aspect as shown, for e.g. values for *pak-slope-proj raster* is named as slope in attribute data.
5. Click *OK* to extract the raster values of these raster layers in each of the sample plots points.





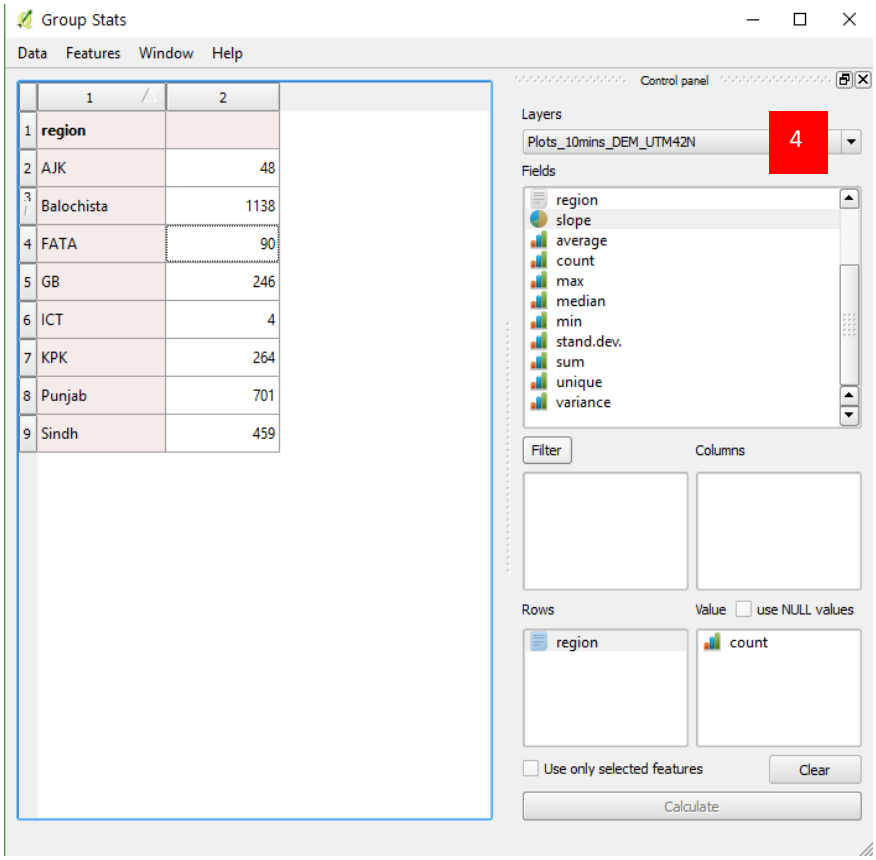


Similar exercise can be done to extract land cover/land use classes from existing data to the designed sample plots for further stratification.

Step 5 Summarizing Plots Statistics

Sample Plots statistics can be summarized based on their attribute data, for example, summarizing number of 10'x10' sample plots in each province to assess adequacy of designed plots at provincial levels.

1. Install *Group Stats plugin* from *Plugins>Manage Plugins* menu.
2. Use *Vector>Group Stats* tool to generate a pivot table for sample plots.
3. Select the sample plot layer in the *Layers* list.
4. Drag 'region' attribute field to Rows and 'count' calculation to Value and select *Calculate* to generate statistics table as shown.



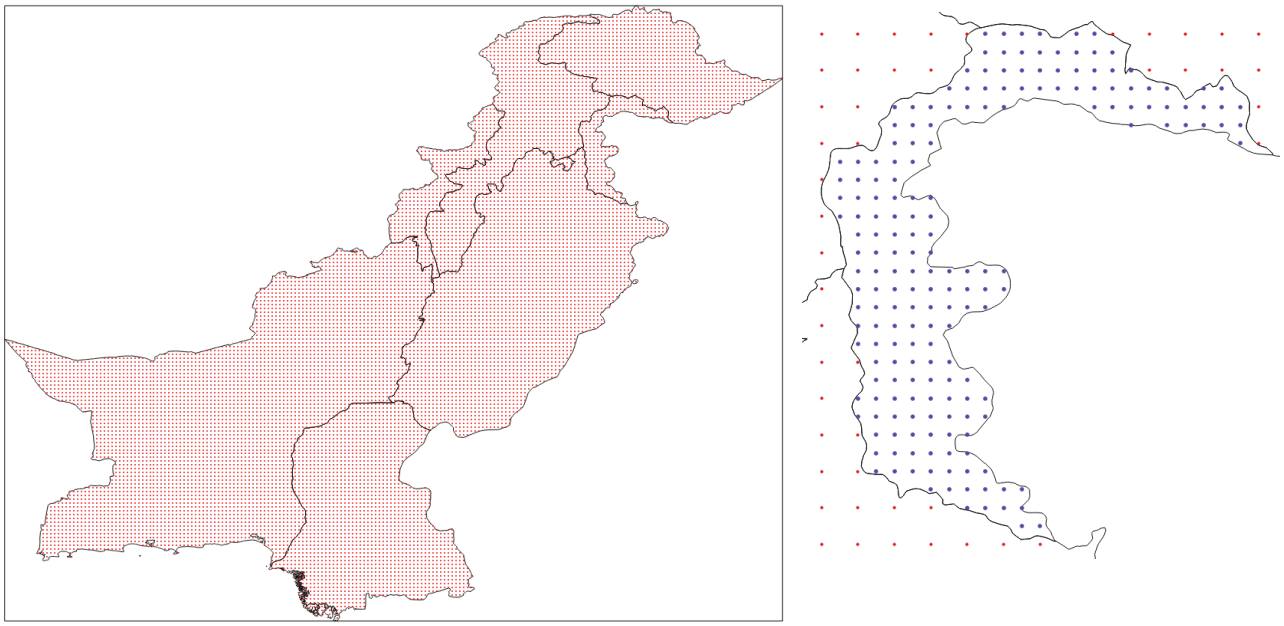
3.1.3. Assignment 1 Densification of Sampling Plots

Review the summary statistics of number of plots per province. Notice AJK has only 48 plots within its boundary, which may not be enough for producing land use/land cover and forest cover maps with sufficient accuracy. The plots may need to be densified for AJK and other smaller provinces (e.g. KP, FATA or ICT). In this assignment, generate densified plots in AJK @ 5' x 5' to supplement the 10' x 10' plots. In general, steps to be followed are:

1. Using the same **Bounding Extent** as created in **Step 1**, generate systematic plots at **5' x 5' (0.08333333 degrees)** following the process in **Step 2**.
2. Once 5' x 5' grids are created with in the extent, select the plots within AJK boundary using *Spatial Query* tool. Save the selected plots as **Plot5mins_AJK.shp**
3. Repeat **Step 3** to calculate latitude and longitude of the new 5' plots
4. Repeat **Step 4** to extract terrain parameters to the plots
5. Answer, how many plots are generated for 5'x 5' sampling design in AJK.

3.1.4. Home Province Assignment

To achieve higher accuracy levels, the sampling density may need to be densified, if the accuracy of subsequent classification and mapping do not comply with the required accuracy levels of overall and per class 80% or higher (using Landsat imagery and classification method). For this, densify the sampling densities at 2.5' for your home provinces and interpret the plots (in Exercise 2B) to be used for classification.



3.2. Exercise 2B Interpretation of LULC Classes using OpenForis CollectEarth

3.2.1. Background

Systematic sampling grids generated in the previous process is visually interpreted for IPCC Land Use Classes (Forest, Cropland, Grassland, Wetland, Settlement, Other land) using very high-resolution Google Earth imageries using customized OpenForis Collect Graphical User Interface (GUI) data collection form. Land use classes with various other parameters for Forest (canopy cover, disturbances etc.) are collected for two reference periods 2016 and 2012 in this exercise. These visually interpreted plots are also used as training sets for image classification.

Data Used

SN	Datasets	Sources	Descriptions
1	Plot5Mins_AJK.shp	Plots generated in Exercise 2A	Systematic sample points @ 5' x 5' covering AJK
2	pakistan_lulc_visual_interpretation.cep	Collect Earth data collection interface developed using OpenForis Collect	Collect Earth Project GUI

Software Used

- Google Earth (https://www.google.com/earth/desktop/)
- OpenForis Collect Earth (http://www.openforis.org/tools/collect-earth.html)

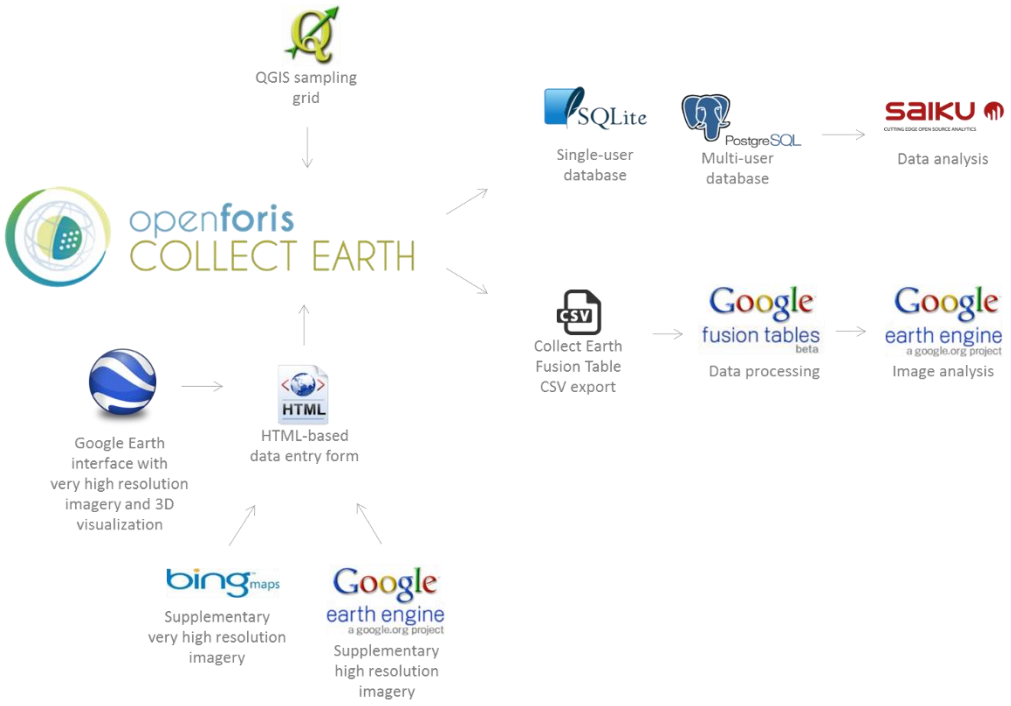
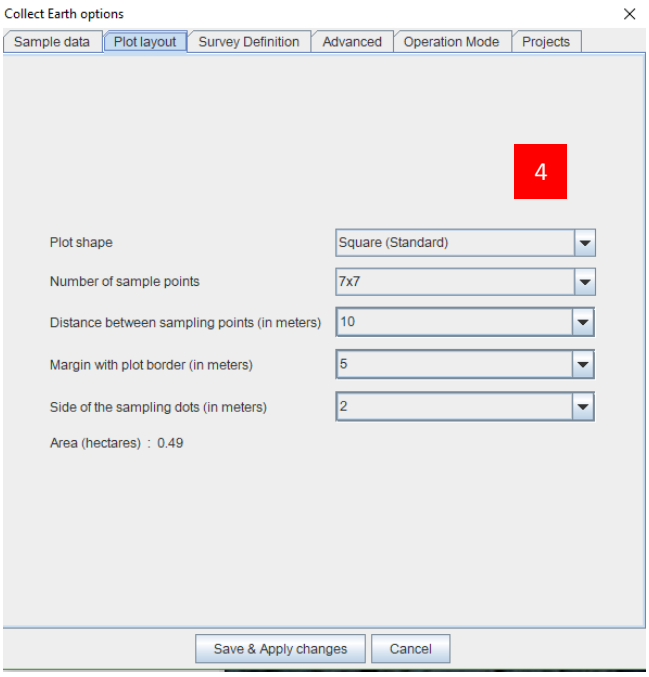


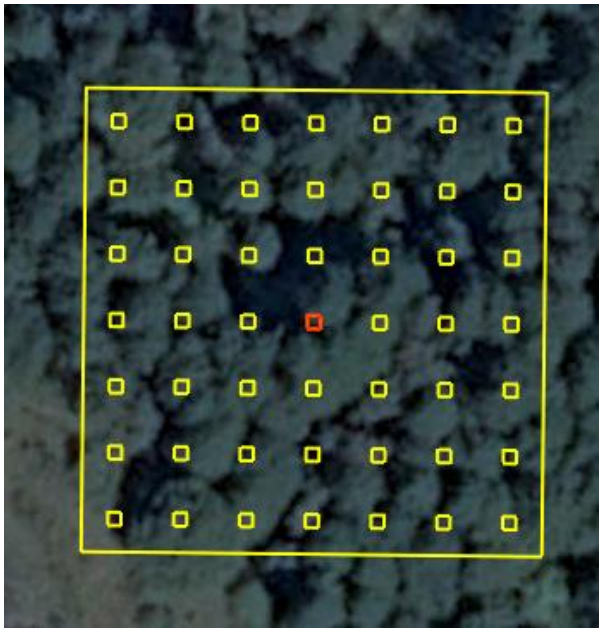
Figure 8 System overview of Collect Earth (http://www.openforis.org/tools/collect-earth.html)

Step 1 Visual Interpretation of Sample Plots

1. Download OpenForis Collect Earth from the site http://openforis.org/tools/collect-earth. There are also tutorials and case study you can follow. But before installing Collect Earth, you will need to install Google Earth. After Collect Earth has been installed, Launch Collect Earth.
2. **Right click** and choose ***Open*** the Collect Earth project file '*pakistan_lulc_visual_interpretation.cep*'. It should directly open in CollectEarth, it will prompt for the Operator's name. You will see the list of plots designed (194 points)
3. Enter your name in small letters (should be more than 8 characters for example "anishjoshi") as shown

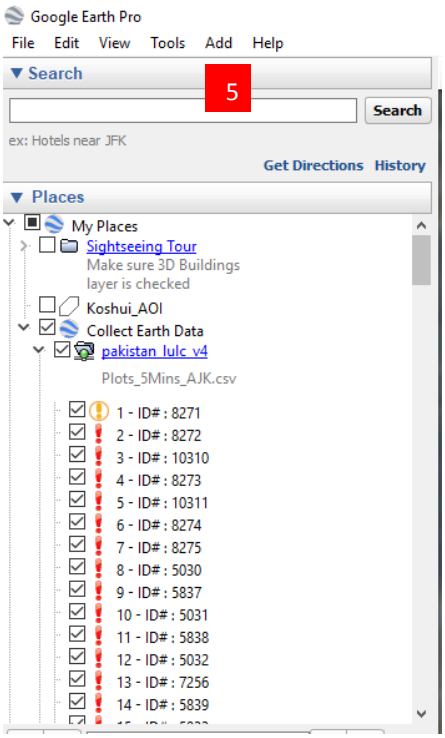
4. Go to *Tools and Properties of CollectEarth* snippet, then go to *Plot Layout Tab***.** Change the plot layout to **Square (Standard)**. Other parameters are used as shown. **Save and Apply Changes.** With these parameters interpretation plots of size 70 m x 70 m (0.49 ha)[1] are generated.
   - Plot shape: Square (Standard)
   - Number of sample plots: 7 x 7
   - Distance between sampling plots (in meters): 10
   - Margin with plot boarder (in meters): 5
   - Side of the sample dots (in meters): 2

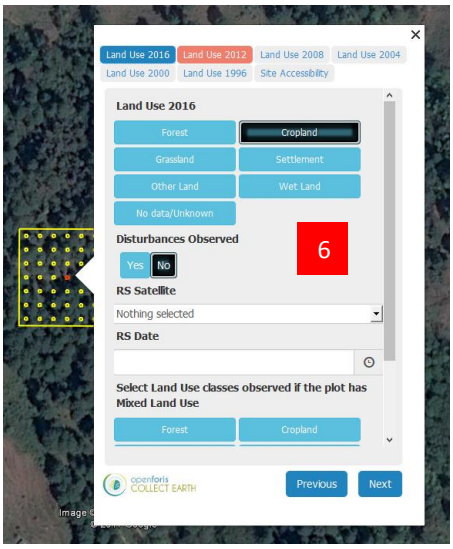




[1] This is visual interpretation plot and the size is recommended to correlate with the minimum area of forest (i.e. 0.5 ha) as well as to comply with the pixel size of image to be used for wall-to-wall mapping i.e. in this case Landsat (30m). Therefore, it would be logical to consider 30 m x 30 m (0.09 ha) to correlate with 1 pixels of Landsat image if the plot is to be used for image interpretation; or 70 m x 70 m (0.49 ha) to comply with the requirement of min. area of 0.5 ha for forest.

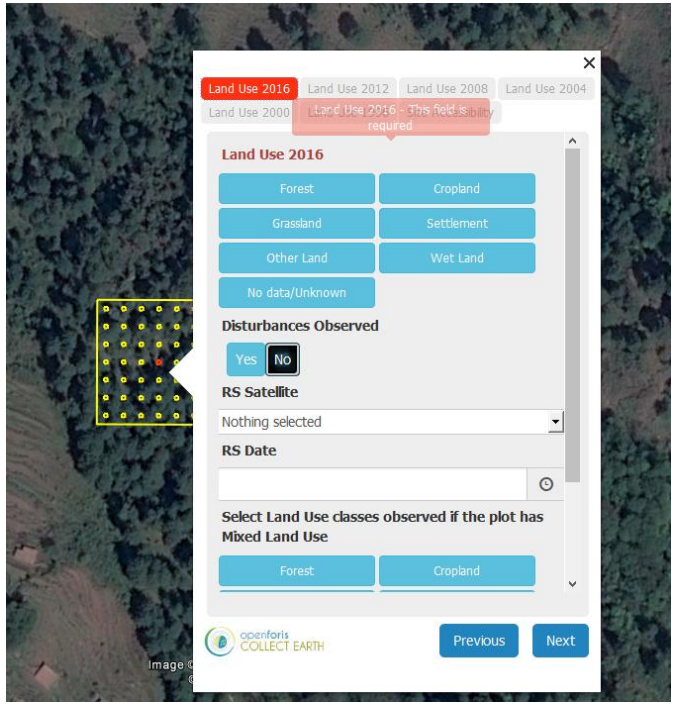
5. Now in Google Earth environment, you could see a square plot layout. Select one of the plots in the contents of Google Earth for e.g. plot ID#8271. Google Earth will zoom in to the location of the plot.
6. Click anywhere within the yellow square, a HTMP pop-up interface will appear with a data entry form as shown. Select *Next* to proceed further.
7. In this exercise, we will interpret LULC for the years 2016 and 2012 only.
8. Select observed Land Use Class interpreted in the image. *[Note: Make sure you select/browse the appropriate date of the Google Earth image. Adjust the year of the image as necessary i.e. 2016 and 2012]*
9. If Forest land use is selected, *Plots Covered in Canopy (7 x7)* option appears. Count the number of square sub-plots (1-25), which are covered by tree canopy. The proportion (in percentage) of crown cover in the 70x70 m plot is calculated for e.g. if 7 plots have crown cover, the proportion is calculated as 14/49 = 29%.
10. Switch the years of images to previous years, if any disturbances are observed select *Yes* for disturbances and select the observed disturbances (multiple choices).
11. Select the image source of interpretation and its date/year.
12. Select Land Use classes if the plot has mixed land use classes.
13. Once completed, select *Next* at the bottom of interface to go to Land Use 2012 interpretation form. Complete 2012 interpretation similarly making sure the image year corresponds to 2012.
14. Once completed, skip other reference years 2008-1996 and go to *Site Accessibility* form.
15. Select accessibility of the plots, exploring the site location, terrain and other restrictions.
16. Once done, select Send, the point is stored in the SQLLite Database and the GoogleEarth moves to the next point. On the left Table of Contents of GoogleEarth, you will notice the interpreted plot is not marked with green tick mark ✓
17. Export interpreted plots to 'csv' text file using *CollectEarth>Tools>Export Data to CSV* to save as ***'ajk_luse_intrepretation.csv'*** data file for further use in image analysis.

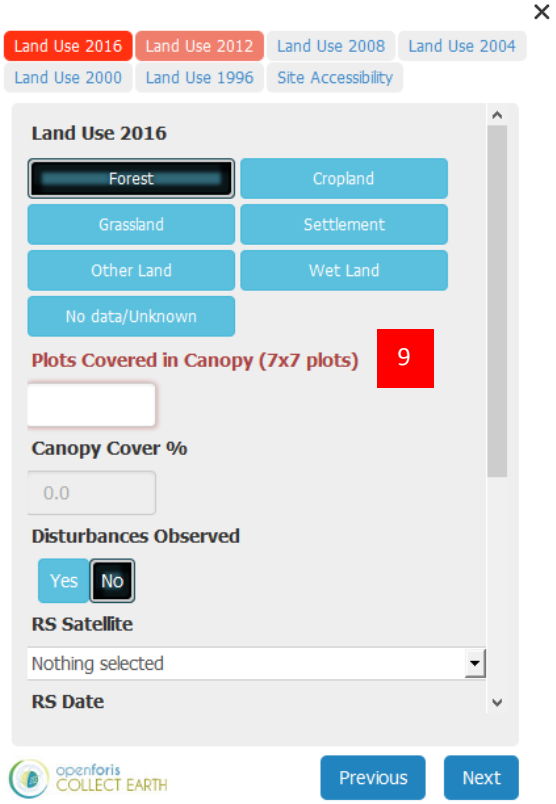


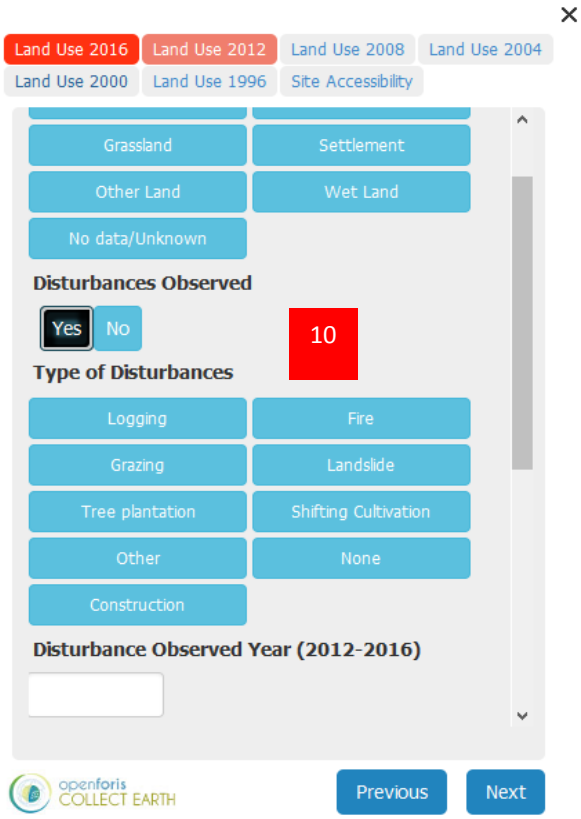


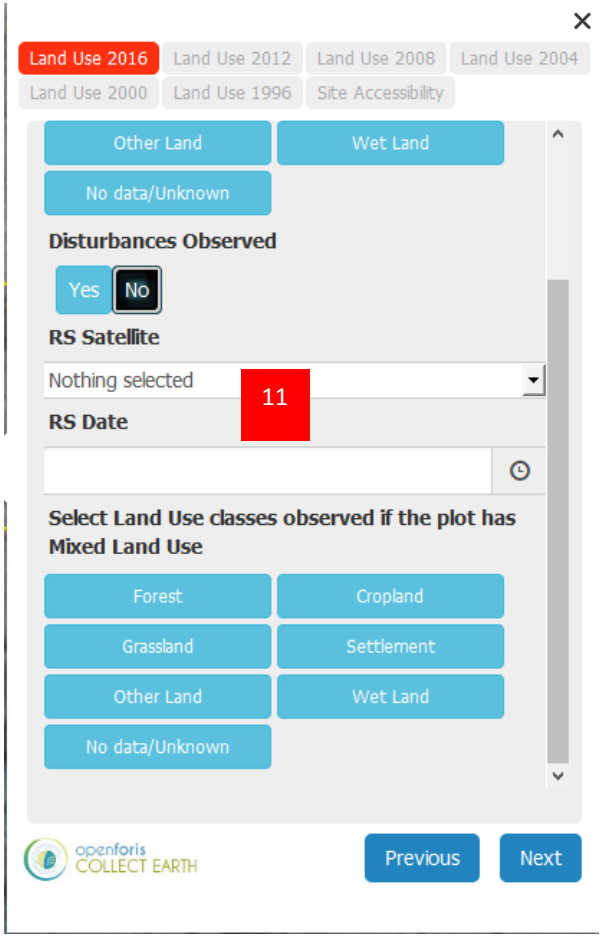
Note:

For this exercise, interpret 20-30 plots to understand the process and tool. For further analysis, all 194 plots have been interpreted and stored in the ***file …//ReferenceData/ajk_luse_interpreted_WWF.csv***

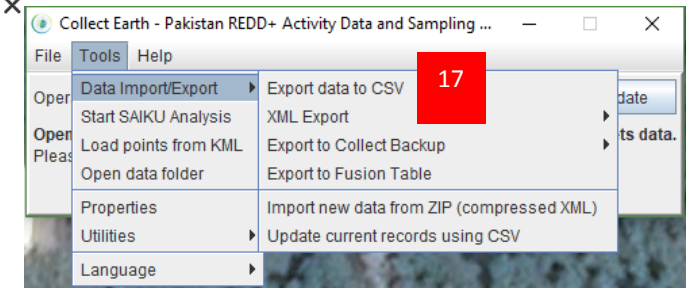

Land Use 2016
Land Use 2012
Land Use 2008
Land Use 2004
Land Use 2000
Land Use 2016 - This field is required.
Land Use 2016
Forest
Cropland
Grassland
Settlement
Other Land
Wet Land
No data/Unknown
Disturbances Observed
Yes
No
RS Satellite
Nothing selected
RS Date
Select Land Use classes observed if the plot has Mixed Land Use
Forest
Cropland
openforis COLLECT EARTH
Previous
Next


Land Use 2016
Land Use 2012
Land Use 2008
Land Use 2004
Land Use 2000
Land Use 1996
Site Accessibility
Land Use 2016
Forest
Cropland
Grassland
Settlement
Other Land
Wet Land
No data/Unknown
Plots Covered in Canopy (7x7 plots)
9
Canopy Cover %
0.0
Disturbances Observed
Yes
No
RS Satellite
Nothing selected
RS Date
openforis COLLECT EARTH
Previous
Next

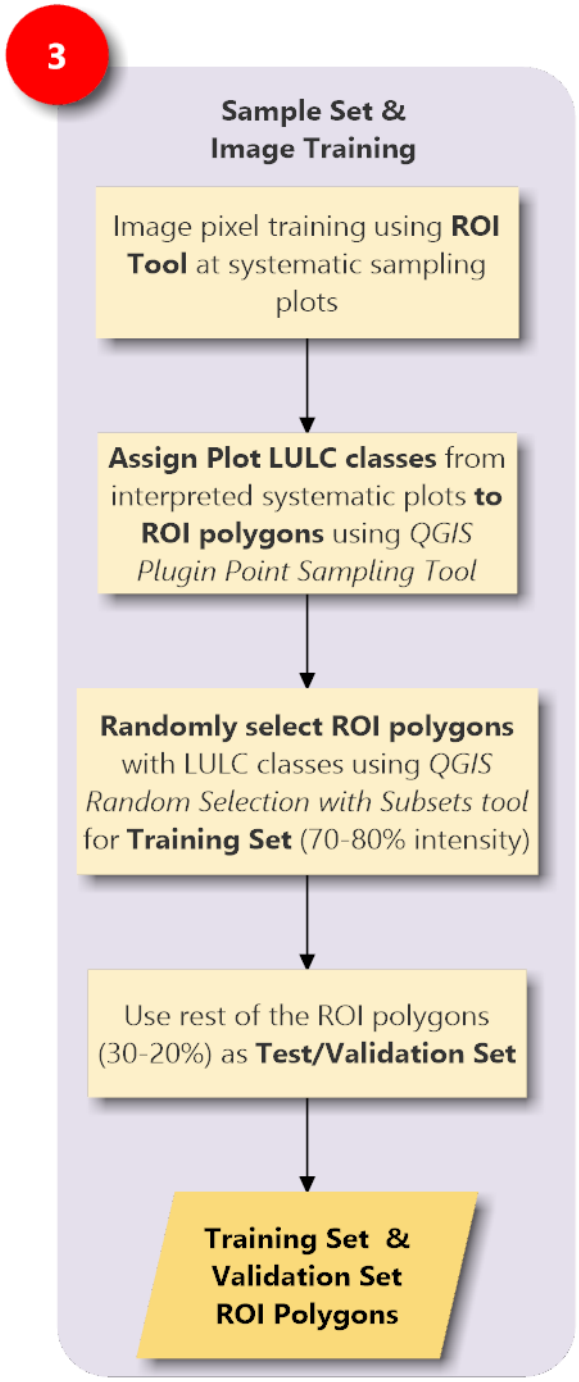

Land Use 2016
Land Use 2012
Land Use 2008
Land Use 2004
Land Use 2000
Land Use 1996
Site Accessibility
Grassland
Settlement
Other Land
Wet Land
No data/Unknown
Disturbances Observed
Yes
No
10
Type of Disturbances
Logging
Fire
Grazing
Landslide
Tree plantation
Shifting Cultivation
Other
None
Construction
Disturbance Observed Year (2012-2016)
openforis COLLECT EARTH
Previous
Next


Land Use 2016
Land Use 2012
Land Use 2008
Land Use 2004
Land Use 2000
Land Use 1996
Site Accessibility
Other Land
Wet Land
No data/Unknown
Disturbances Observed
Yes
No
RS Satellite
Nothing selected
11
RS Date
Select Land Use classes observed if the plot has Mixed Land Use
Forest
Cropland
Grassland
Settlement
Other Land
Wet Land
No data/Unknown
openforis COLLECT EARTH
Previous
Next

Land Use 2016
Land Use 2012
Land Use 2008
Land Use 2004
Land Use 2000
Land Use 1996
Site Accessibility
Accessibility
Accessible
Inaccessible due to Slope
Inaccessible due to Water body
Inaccessible due to Restrictions
Inaccessible due to Other Reasons
Unknown
15
openforis COLLECT EARTH
Previous
Send
16


Collect Earth - Pakistan REDD+ Activity Data and Sampling ...
File
Tools
Help
Data Import/Export
Start SAIKU Analysis
Load points from KML
Open data folder
Properties
Utilities
Language
Export data to CSV
XML Export
Export to Collect Backup
Export to Fusion Table
Import new data from ZIP (compressed XML)
Update current records using CSV
17

4. Sample Set for Image Training



Land Use classes (IPCC classes) of the plots interpreted in CollectEarth are used for training Landsat 5 TM/Landsat 8 OLI image pixels using QGIS Semi-Automatic Classification Plugin

Region of Interest (ROI) polygons are generated using the interpreted plots as 'seeds' using region growing algorithm. ROIs can also be generated manually.

Generated ROIs are randomly selected for training subset and validation subset.

4.1. Exercise 3 Creating Training and Validation Set

4.1.1. Background

Visually interpreted points with land use classes are used as 'seeds' for generating ROI sample sets to develop training and validation subsets. The training sets are used for training image pixels for Random Forests classification. The validation sets are used for independent check of the accuracy and subsequent area estimation.

Data Used

SN	Datasets	Sources	Descriptions
1	CollectEarth Interpreted data *(ajk_luse_interpreted_WWF.csv)*	CollectEarth interpreted data	CollectEarth interpreted data)
2	Landsat 8 OLI image scenes mosaics (150-36, 149-36, 149-37) *(L8_AJK_Mosaic.tif)*		Landsat 8 OLI atmospherically corrected image scene mosaic
3	Administrative Boundary Level 1 *(PAK_adm1.shp)* *(AJK_UTM43N.shp)*	Global Administrative Areas http://www.gadm.org/	administrative areas (administrative boundaries)

Software Used

- Quantum GIS (QGIS) [1] Ver 2.18 (or older) - Open source GIS software
- Semi-automatic Classification Plugin in QGIS (SCP tool)

[1] Latest version of QGIS can be downloaded from http://www.qgis.org/

4.1.2. Generating Region of Interest (ROI)

Step 1 Image Preparation – Clipping/Masking

1. Using AJK administrative boundary from boundary vector data '*PAK_adm1.shp',* clip the '*L8_AJK_Mosaic.tif*' Landsat 8 image data. *[Note: You will need to extract AJK boundary from the PAK_adm1.shp data and save as a separate polygon shape file for e.g. AJK_Boundary.shp. Make sure that the CRS of the image and the vector boundary data are the same. UTM 43N (EPSG:32643) in this case]*
2. Use *Raster>Extraction>Clipper* tool to clip the image mosaic raster with AJK boundary vector. Alternatively, you can create a masking raster layer with AJK boundary vector.

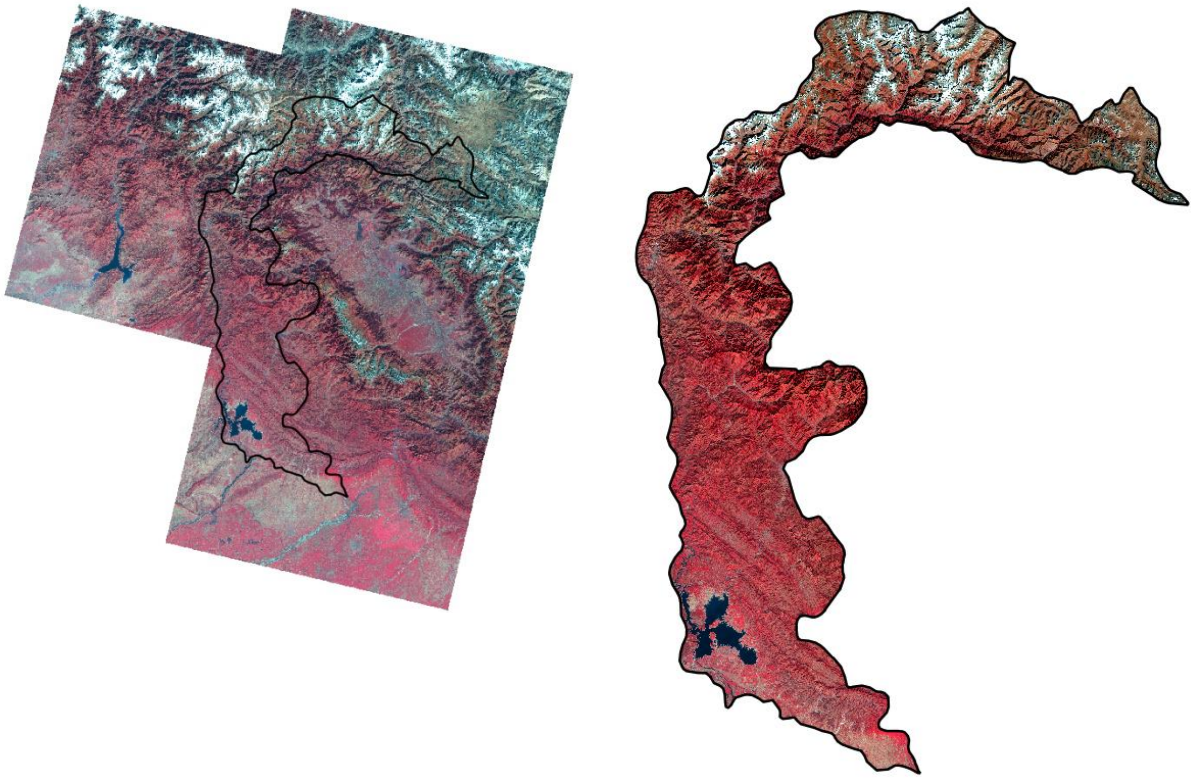
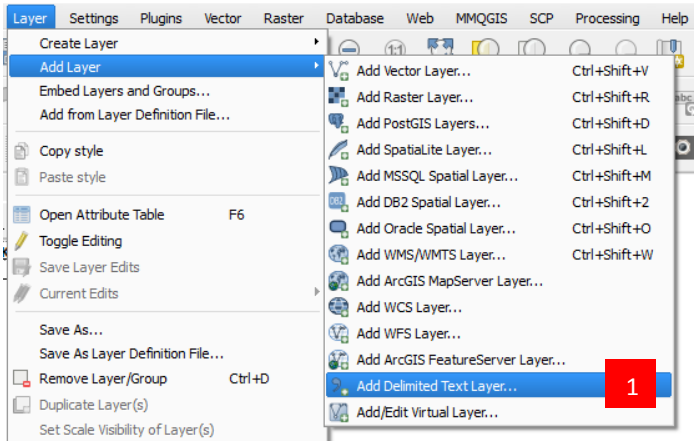


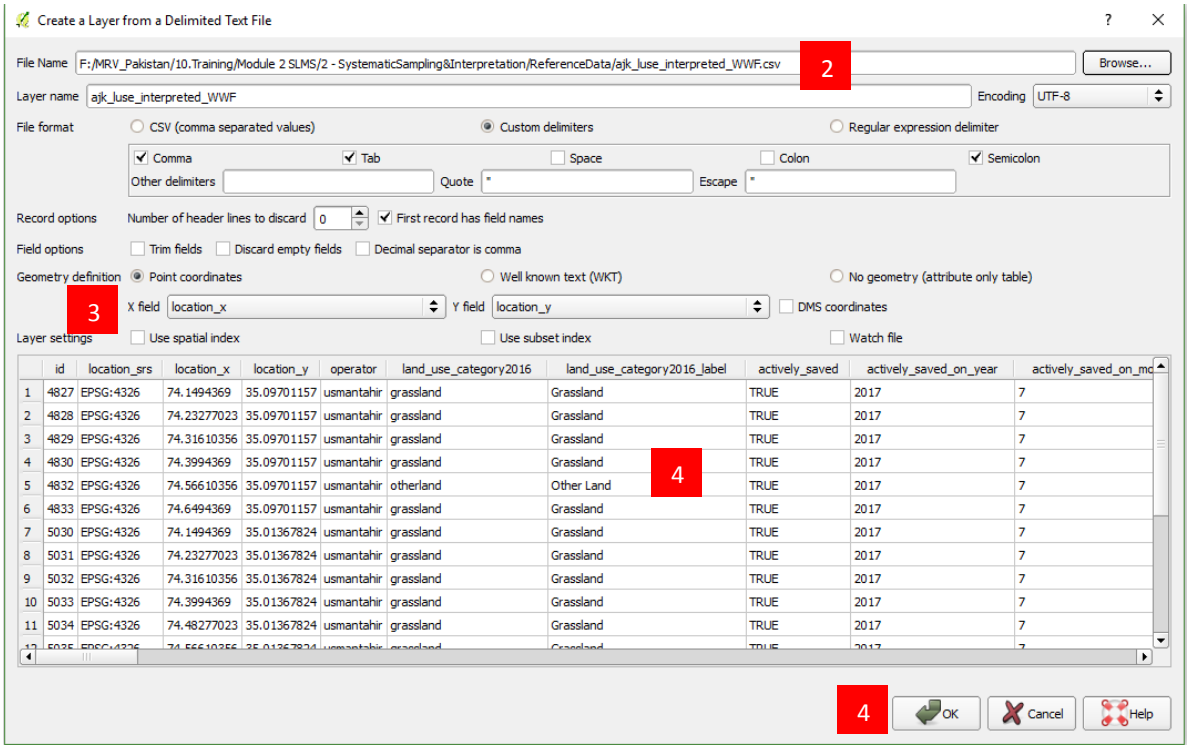
Figure 9 Mosaic L8 image with AJK clipping boundary vector data and clipped L8 image

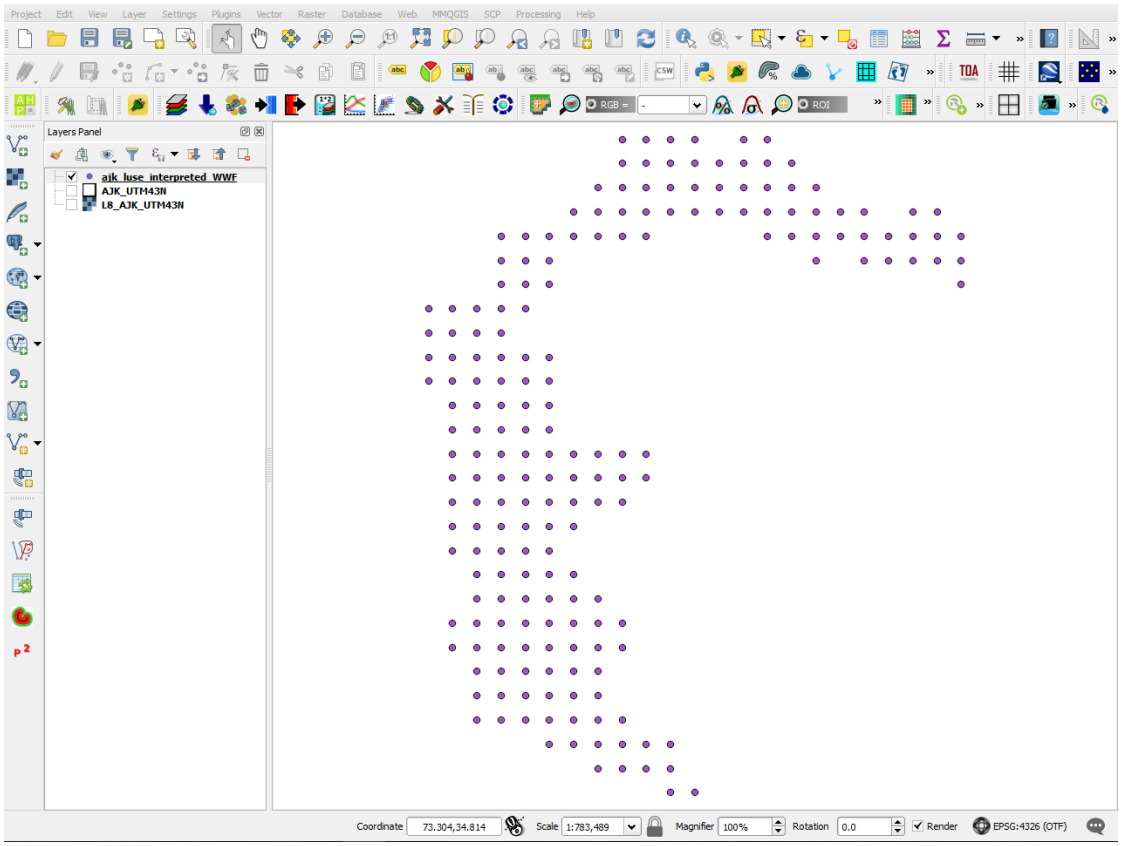
Step 2 Sample Preparation

1. Add the CollectEarth interpreted *ajk_luse_interpreted_WWF.csv* data to generate plot point vector data using *Layer>Add Layer>Add Delimited Text Layer* menu.
2. Browse to the '*ajk_luse_interpreted_WWF.csv'* file. Make sure to select *Custom delimiters.*
3. Select Point coordinates and browse to the X Field and Y Field (i.e. longitude and latitude). In this case, 'location_x' and 'location_y' fields respectively.

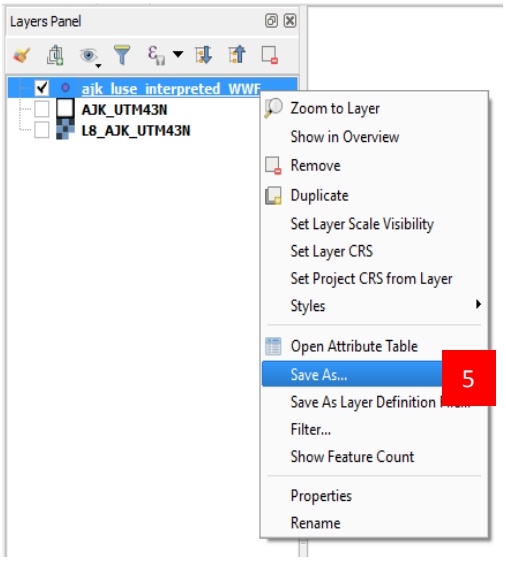


4. The list of points in the csv file is displayed as shown. Select *OK* to import the points to QGIS.

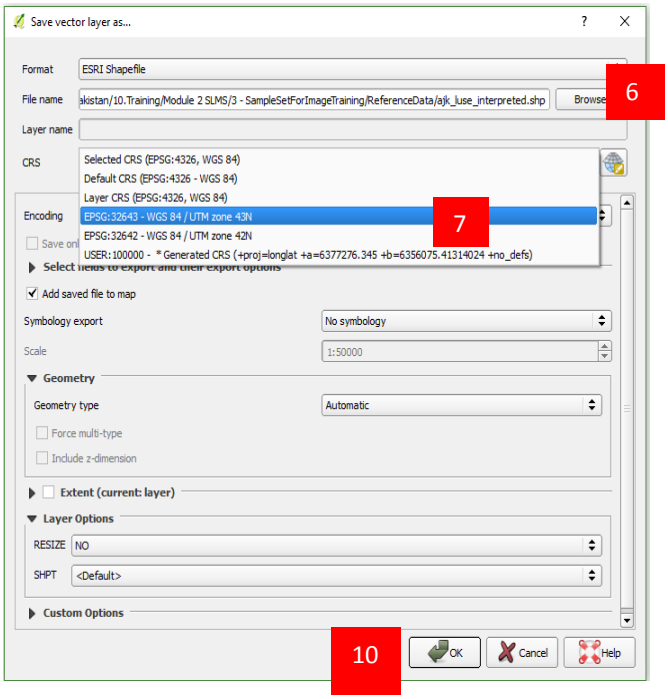


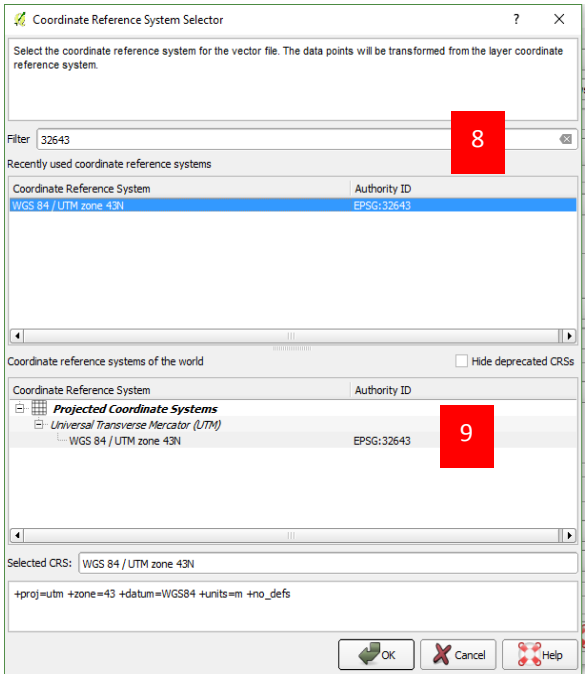


5. These points are imported as text csv file in default CRS (WGS84). Since the images has been processed in UTM projection (UTM43N), it is necessary to convert the CRS of the imported point/plots to the same CRS as that of the images. This can be done by **Saving** the added csv file as vector dataset in shape file format by *Right Click* on the layer *Save As…* option.

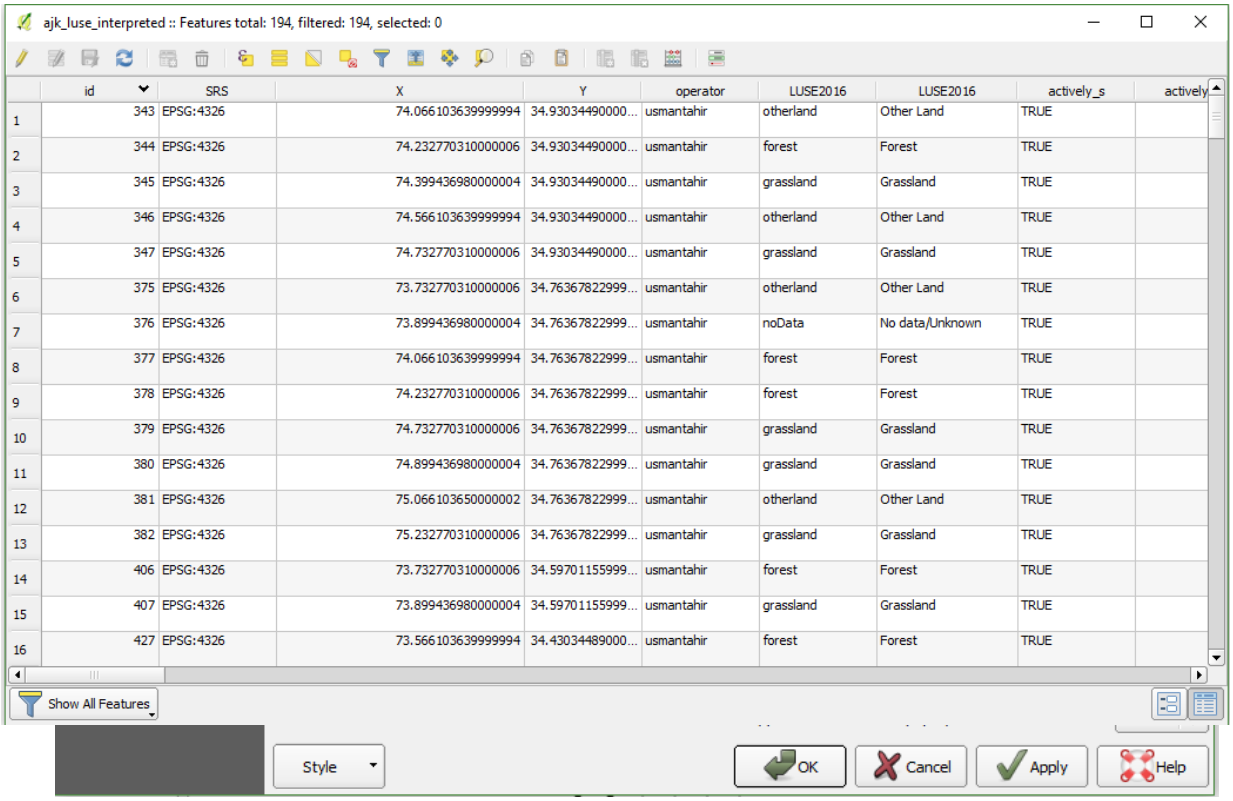


6. Browse to the appropriate location to save the file and provide the filename. In this case *'//Exercise3/ExerciseData/ajk_luse_interpreted.shp'*
7. In the CRS (Coordinate Reference System), select *EPSG:32643 – WGS84/UTM Zone 43N* for this dataset.
8. If EPSG:32643 does not appear in the list, use the icon on its side *Select CRS* to browse to the CRS lists in the system. Type 32643 in the Filter to search for EPSG:32643.
9. WGS 84/ UTM One 43N is displayed in the list, select the CRS and press OK to assign CRS to exported points.
10. Leave other options as default and finally, press OK to save the point plots with UTM 43N CRS. The imported csv file is now saved permanently as a vector point file in UTM 43N CRS.





11. Since the CSV file contained field names with more than 8 characters, the field names in the imported/saved '*ajk_luse_interpreted.shp' will be curtailed. To avoid this, you can rename the field names in the original csv file appropriately (for e.g.* LUSE2016, LUSE2012 etc.) or alternatively you can choose to provide "Alias" names to the fields from *Layer Properties>Fields* (right click on layer) as shown.

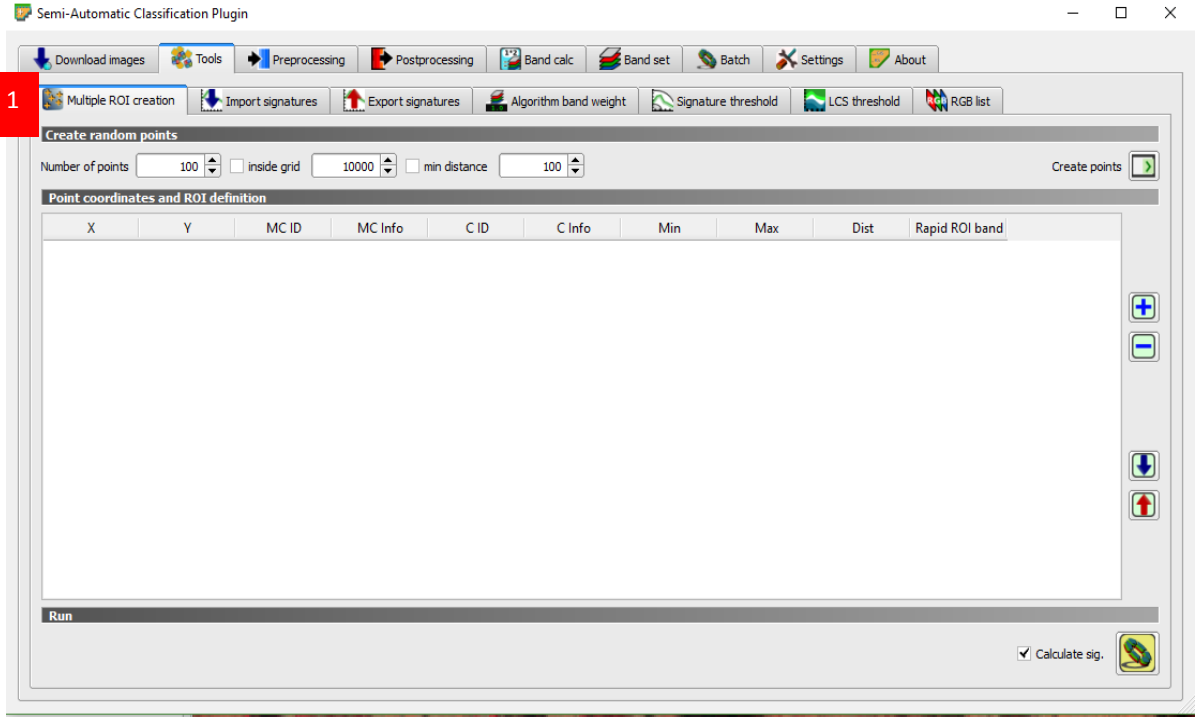


	id	SRS	X	Y	operator	LUSE2016	LUSE2016	actively_s	actively
1	343	EPSG:4326	74.066103639999994	34.93034490000...	usmantahir	otherland	Other Land	TRUE	
2	344	EPSG:4326	74.232770310000006	34.93034490000...	usmantahir	forest	Forest	TRUE	
3	345	EPSG:4326	74.399436980000004	34.93034490000...	usmantahir	grassland	Grassland	TRUE	
4	346	EPSG:4326	74.566103639999994	34.93034490000...	usmantahir	otherland	Other Land	TRUE	
5	347	EPSG:4326	74.732770310000006	34.93034490000...	usmantahir	grassland	Grassland	TRUE	
6	375	EPSG:4326	73.732770310000006	34.76367822999...	usmantahir	otherland	Other Land	TRUE	
7	376	EPSG:4326	73.899436980000004	34.76367822999...	usmantahir	noData	No data/Unknown	TRUE	
8	377	EPSG:4326	74.066103639999994	34.76367822999...	usmantahir	forest	Forest	TRUE	
9	378	EPSG:4326	74.232770310000006	34.76367822999...	usmantahir	forest	Forest	TRUE	
10	379	EPSG:4326	74.732770310000006	34.76367822999...	usmantahir	grassland	Grassland	TRUE	
11	380	EPSG:4326	74.899436980000004	34.76367822999...	usmantahir	grassland	Grassland	TRUE	
12	381	EPSG:4326	75.066103650000002	34.76367822999...	usmantahir	otherland	Other Land	TRUE	
13	382	EPSG:4326	75.232770310000006	34.76367822999...	usmantahir	grassland	Grassland	TRUE	
14	406	EPSG:4326	73.732770310000006	34.59701155999...	usmantahir	forest	Forest	TRUE	
15	407	EPSG:4326	73.899436980000004	34.59701155999...	usmantahir	grassland	Grassland	TRUE	
16	427	EPSG:4326	73.566103639999994	34.43034489000...	usmantahir	forest	Forest	TRUE	

[Note: Unnecessary fields can be deleted using Layer Properties and Delete OR from the Attribute Table view.]

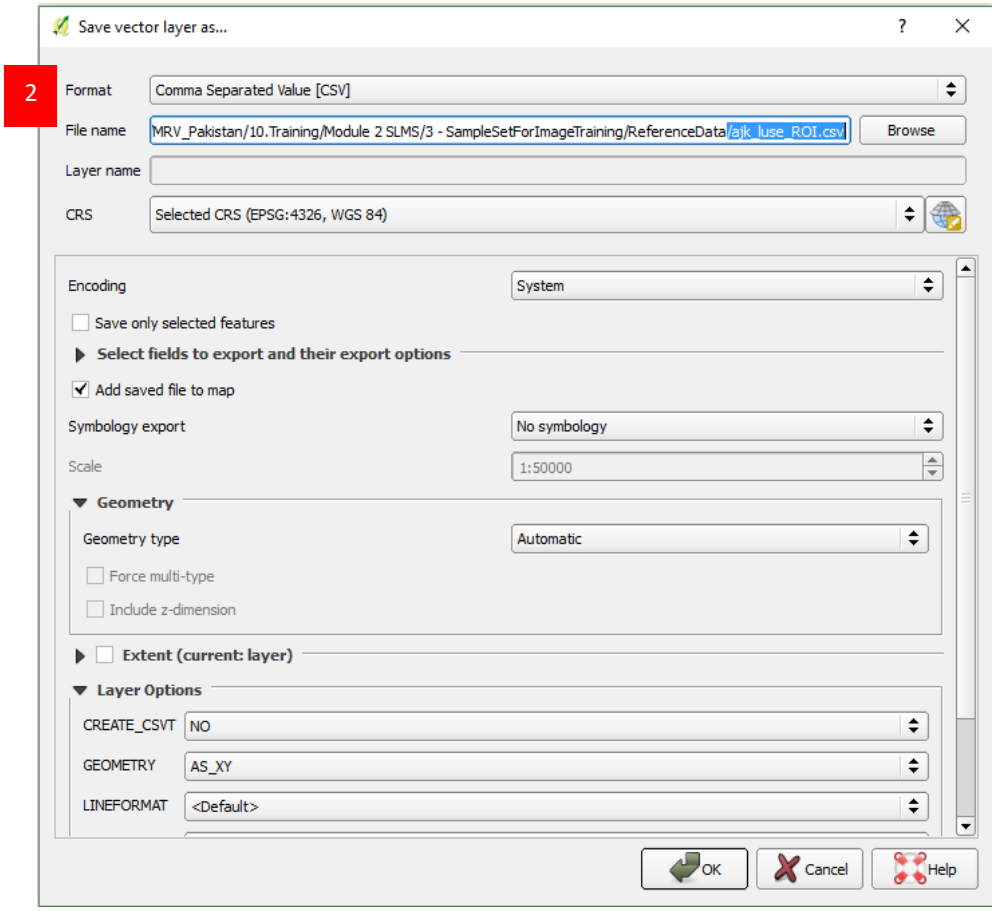
Step 3A Automatic Generation of Region of Interest (ROI) Polygons (Optional)

1. ROI polygons can be generated automatically from the sample points with designated land use categories in the attribute field (e.g. LUSE2016, LUSE2012) using *Multiple ROI Creation* tool in Semi-Automatic Classification plugin. ROIs can be generated manually for each of the point sample, however, since we already have point samples from higher resolution images, we can use these to train the image pixels for classification.



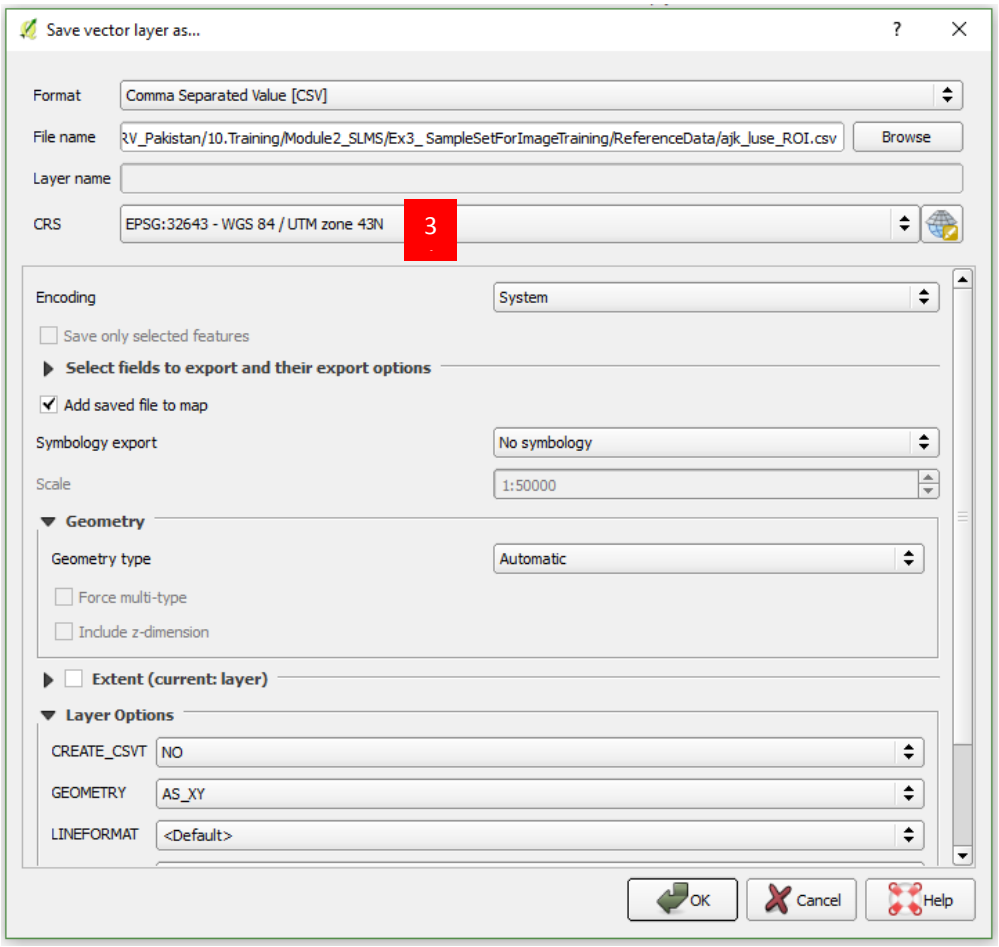
2. Prior to generating ROI's from the sample plot points, ROI definition file needs to be created from the sample plots with these fields:

X	X coordinate/longitude
Y	Y coordinate/latitude
MC ID	ROI Macro Class ID
MC Info	ROI Macro Class Label
C ID	ROI Class ID
C Info	ROI Class Label
Min size	Minimum area of ROI (in pixel unit)
Max width	Side of a square which inscribes the ROI, defining the maximum width thereof (in pixel unit)
Range radius	Similarity of pixels (distance in radiometry units)
Rapid ROI band	No. of image bands to select to create ROI (value 1 uses only one image band, hence ROI creating is faster. Value 0 uses all available bands, however the process might get slower using all the mands)



3. Using the saved '*ajk_luse_interpreted.shp', save the attribute table as 'csv' datafile with name '/ajk_luse_ROI.csv'*. This will extract the attribute table as a separate text csv data file to be edited for using as the source for generating ROIs. Select **UTM Zone 43N** to save the projected coordinate in UTM 43N.

[Note: 'ajk_luse_interpreted.shp' data is in WGS84 and needs to be converted to UTM projection for ROI tool to work, since the reference images are in UTM projection. Changing the CRS while exporting to csv, projected easting(X) and northing (Y) fields are added in the datafile, which is later used for generating ROIs at these plot locations.]



4. Open the exported *'ajk_luse_ROI.csv'* file in MS Excel. And start editing the field names. Delete the fields except **location_x, location_y, luse_2016.**
5. Rename location_x to X, location_y to Y and luse2016 to MC Info
6. Add additional fields **MC ID, C ID, C Info, Min size, Max width, Range radius, Rapid ROI band** in the data file and fill in with these data

X	X coordinate (easting)
Y	Y coordinate (northing)
MC ID	For corresponfing luse (MC Info) enter IDs 1-7 as shown 1 Forest 2 Cropland 3 Grassland 4 Wet Land 5 Settlement 6 Other Land 7 No data/Unclassified
MC Info	Forest Cropland Grassland Wet Land Settlement Other Land No data/Unclassified

C ID	For this exercise, we are not using any sub-clases, therefore copy the MC ID field and paste in C ID
C Info	For this exercise, we are not using any sub-clases, therefore copy the MC Iinfo field and paste in C Info
Min size	60 (i.e. 2 pixels)
Max width	100
Range radius	0.01
Rapid ROI band	<leave this blank to use all the image bands for ROI generation>

A	B	C	D	E	F	G	H	I	J
X	Y	MC Id	MC Info	C Id	C Info	Min size	Max width	Range radius	Rapid ROI band
422475.4523	3884132.231	3	Grassland	3	Grassland	60	100	0.01	
430071.0297	3884070.571	3	Grassland	3	Grassland	60	100	0.01	
437666.5565	3884015.265	3	Grassland	3	Grassland	60	100	0.01	
445262.0392	3883966.312	3	Grassland	3	Grassland	60	100	0.01	
460452.8896	3883887.463	6	Other Land	6	Other Land	60	100	0.01	
468048.27	3883857.567	3	Grassland	3	Grassland	60	100	0.01	
422396.6458	3874890.447	3	Grassland	3	Grassland	60	100	0.01	
429999.9457	3874828.853	3	Grassland	3	Grassland	60	100	0.01	
437603.1946	3874773.605	3	Grassland	3	Grassland	60	100	0.01	
445206.399	3874724.704	3	Grassland	3	Grassland	60	100	0.01	
452809.5624	3874682.148	3	Grassland	3	Grassland	60	100	0.01	
460412.6915	3874645.938	3	Grassland	3	Grassland	60	100	0.01	
468015.7926	3874616.074	3	Grassland	3	Grassland	60	100	0.01	
475618.8695	3874592.554	6	Other Land	6	Other Land	60	100	0.01	
422318.0039	3865648.788	6	Other Land	6	Other Land	60	100	0.01	
437539.965	3865532.07	6	Other Land	6	Other Land	60	100	0.01	
452761.7434	3865440.711	3	Grassland	3	Grassland	60	100	0.01	
467983.383	3865374.707	3	Grassland	3	Grassland	60	100	0.01	
483204.9276	3865334.057	3	Grassland	3	Grassland	60	100	0.01	
407001.9543	3856549.179	6	Other Land	6	Other Land	60	100	0.01	
414620.7716	3856475.049	6	Other Land	6	Other Land	60	100	0.01	
422239.5266	3856407.253	3	Grassland	3	Grassland	60	100	0.01	
429858.2232	3856345.791	3	Grassland	3	Grassland	60	100	0.01	
437476.8677	3856290.661	1	Forest	1	Forest	60	100	0.01	
445095.4669	3856241.864	6	Other Land	6	Other Land	60	100	0.01	
452714.0246	3856199.4	6	Other Land	6	Other Land	60	100	0.01	
460332.5472	3856163.267	3	Grassland	3	Grassland	60	100	0.01	
467951.0413	3856133.467	6	Other Land	6	Other Land	60	100	0.01	
475569.5108	3856109.998	6	Other Land	6	Other Land	60	100	0.01	
483187.9622	3856092.86	6	Other Land	6	Other Land	60	100	0.01	
490806.4012	3856082.054	6	Other Land	6	Other Land	60	100	0.01	
498424.8343	3856077.58	6	Other Land	6	Other Land	60	100	0.01	
513661.7007	3856087.624	6	Other Land	6	Other Land	60	100	0.01	
521280.1472	3856102.144	6	Other Land	6	Other Land	60	100	0.01	

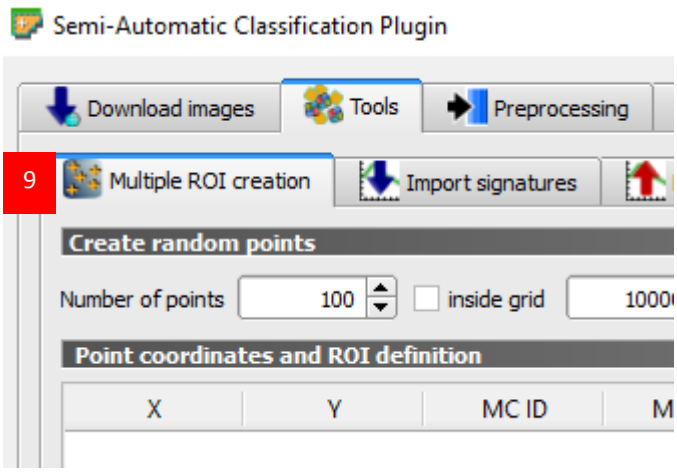
7. After editing, save the data file as *'ajk_luse_ROI.csv'.*
8. CSV file creates "comma , " delimited data, however, the SCP Multiple ROI creation tool uses "semi-colon :" delimited data (some bug in program), therefore, the "comma ," need to be converted into ";". This can be done by opening the 'ajk_luse_ROI.csv' file in Text Editor application (such as Notepad or WordPad and Replacing "," with "semi-colon ;"

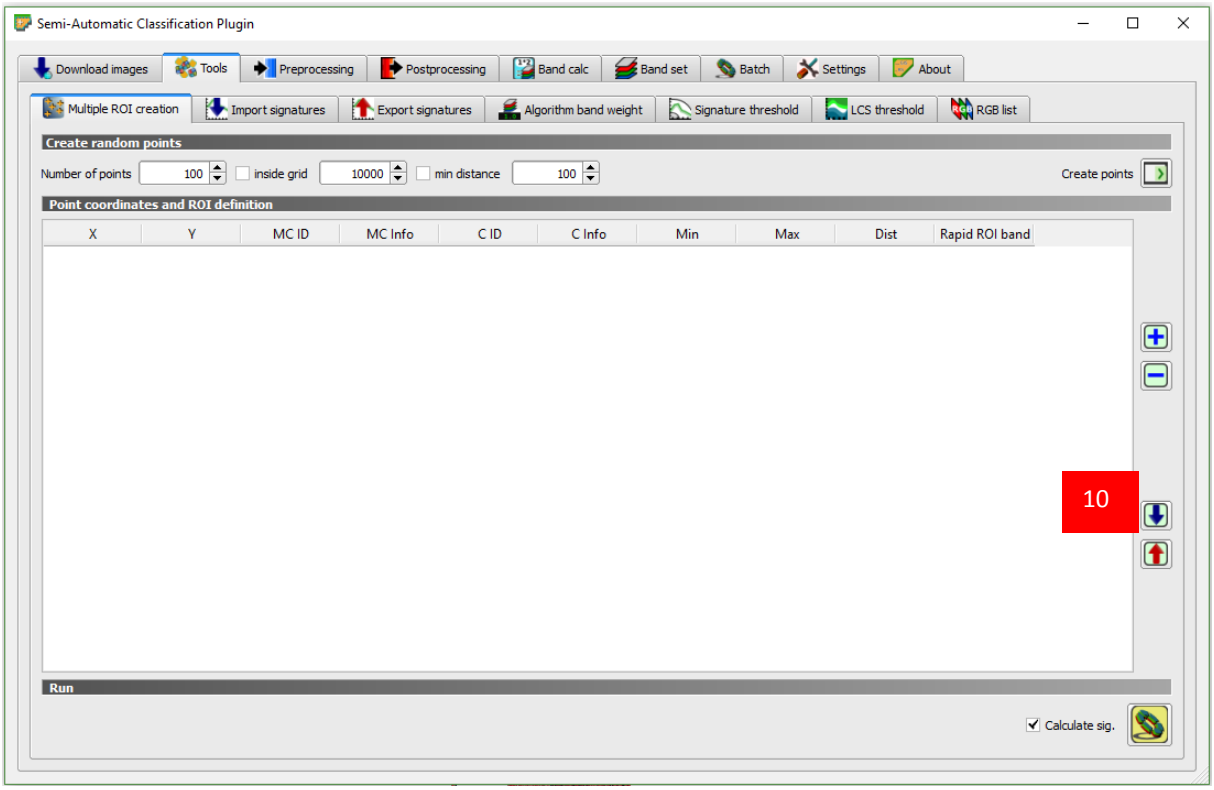
```
X;Y;MC Id;MC Info;C Id;C Info;Min size;Max width;Range radius;Rapid ROI band
422475.4523;3884132.231;3;Grassland;3;Grassland;60;100;0.01;
430071.0297;3884070.571;3;Grassland;3;Grassland;60;100;0.01;
437666.5565;3884015.265;3;Grassland;3;Grassland;60;100;0.01;
445262.0392;3883966.312;3;Grassland;3;Grassland;60;100;0.01;
460452.8896;3883887.463;6;Other Land;6;Other Land;60;100;0.01;
468048.27;3883857.567;3;Grassland;3;Grassland;60;100;0.01;
422396.6458;3874890.447;3;Grassland;3;Grassland;60;100;0.01;
429999.9457;3874828.853;3;Grassland;3;Grassland;60;100;0.01;
437603.1946;3874773.605;3;Grassland;3;Grassland;60;100;0.01;
445206.399;3874724.704;3;Grassland;3;Grassland;60;100;0.01;
452809.5624;3874682.148;3;Grassland;3;Grassland;60;100;0.01;
460412.6915;3874645.938;3;Grassland;3;Grassland;60;100;0.01;
468015.7926;3874616.074;3;Grassland;3;Grassland;60;100;0.01;
475618.8695;3874592.554;6;Other Land;6;Other Land;60;100;0.01;
422318.0039;3865648.788;6;Other Land;6;Other Land;60;100;0.01;
437539.965;3865532.07;6;Other Land;6;Other Land;60;100;0.01;
452761.7434;3865440.711;3;Grassland;3;Grassland;60;100;0.01;
467983.383;3865374.707;3;Grassland;3;Grassland;60;100;0.01;
483204.9276;3865334.057;3;Grassland;3;Grassland;60;100;0.01;
407001.9543;3856549.179;6;Other Land;6;Other Land;60;100;0.01;
414620.7716;3856475.049;6;Other Land;6;Other Land;60;100;0.01;
422239.5266;3856407.253;3;Grassland;3;Grassland;60;100;0.01;
429858.2232;3856345.791;3;Grassland;3;Grassland;60;100;0.01;
437476.8677;3856290.661;1;Forest;1;Forest;60;100;0.01;
445095.4669;3856241.864;6;Other Land;6;Other Land;60;100;0.01;
```

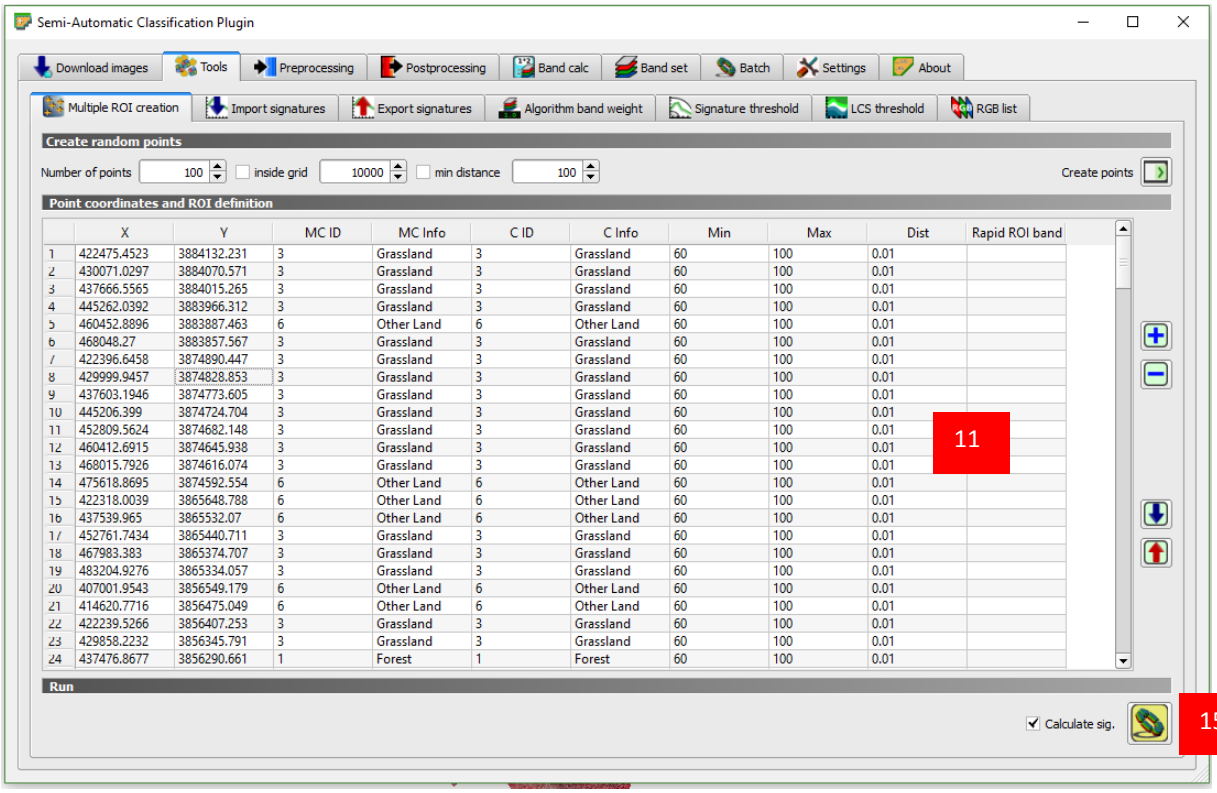
8

9. In the SCP plugin window, go to *Tools and Multiple ROI creation tab*
10. *Import point list from text file* button to import the csv data from *'ajk_luse_ROI.csv'*.

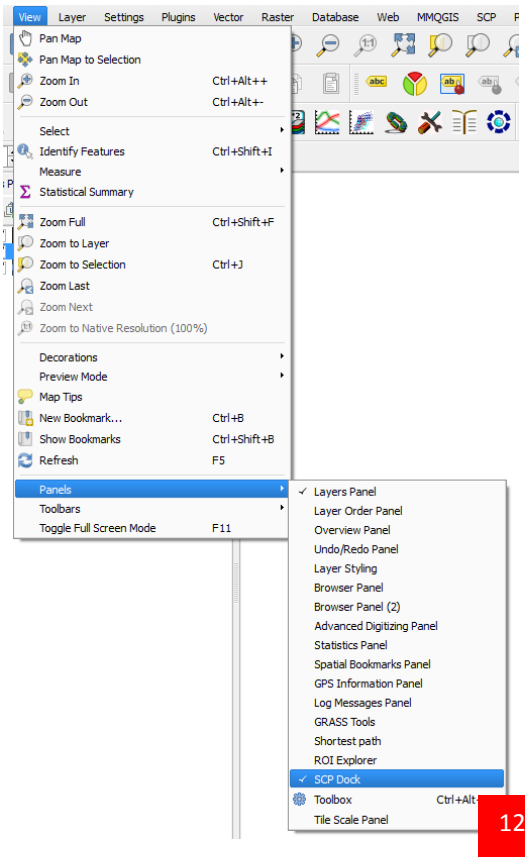
11. The imported csv data is displayed on the window.




Semi-Automatic Classification Plugin
Download images
Tools
Preprocessing
Postprocessing
Band calc
Band set
Batch
Settings
About
Multiple ROI creation
Import signatures
Export signatures
Algorithm band weight
Signature threshold
LCS threshold
RGB list
Create random points
Number of points
100
inside grid
10000
min distance
100
Create points
Point coordinates and ROI definition
X
Y
MC ID
MC Info
C ID
C Info
Min
Max
Dist
Rapid ROI band
10
Run
Calculate sig.


Semi-Automatic Classification Plugin
Download images
Tools
Preprocessing
Postprocessing
Band calc
Band set
Batch
Settings
About
Multiple ROI creation
Import signatures
Export signatures
Algorithm band weight
Signature threshold
LCS threshold
RGB list
Create random points
Number of points
100
inside grid
10000
min distance
100
Create points
Point coordinates and ROI definition
X Y MC ID MC Info C ID C Info Min Max Dist Rapid ROI band
1 422475.4523 3884132.231 3 Grassland 3 Grassland 60 100 0.01
2 430071.0297 3884070.571 3 Grassland 3 Grassland 60 100 0.01
3 437666.5565 3884015.265 3 Grassland 3 Grassland 60 100 0.01
4 445262.0392 3883966.312 3 Grassland 3 Grassland 60 100 0.01
5 460452.8896 3883887.463 6 Other Land 6 Other Land 60 100 0.01
6 468048.27 3883857.567 3 Grassland 3 Grassland 60 100 0.01
7 422396.6458 3874890.447 3 Grassland 3 Grassland 60 100 0.01
8 429999.9457 3874828.853 3 Grassland 3 Grassland 60 100 0.01
9 437603.1946 3874773.605 3 Grassland 3 Grassland 60 100 0.01
10 445206.399 3874724.704 3 Grassland 3 Grassland 60 100 0.01
11 452809.5624 3874682.148 3 Grassland 3 Grassland 60 100 0.01
12 460412.6915 3874645.938 3 Grassland 3 Grassland 60 100 0.01
13 468015.7926 3874616.074 3 Grassland 3 Grassland 60 100 0.01
14 475618.8695 3874592.554 6 Other Land 6 Other Land 60 100 0.01
15 422318.0039 3865648.788 6 Other Land 6 Other Land 60 100 0.01
16 437539.965 3865532.07 6 Other Land 6 Other Land 60 100 0.01
17 452761.7434 3865440.711 3 Grassland 3 Grassland 60 100 0.01
18 467983.383 3865374.707 3 Grassland 3 Grassland 60 100 0.01
19 483204.9276 3865334.057 3 Grassland 3 Grassland 60 100 0.01
20 407001.9543 3856549.179 6 Other Land 6 Other Land 60 100 0.01
21 414620.7716 3856475.049 6 Other Land 6 Other Land 60 100 0.01
22 422239.5266 3856407.253 3 Grassland 3 Grassland 60 100 0.01
23 429858.2232 3856345.791 3 Grassland 3 Grassland 60 100 0.01
24 437476.8677 3856290.661 1 Forest 1 Forest 60 100 0.01
11
Run
Calculate sig.
15

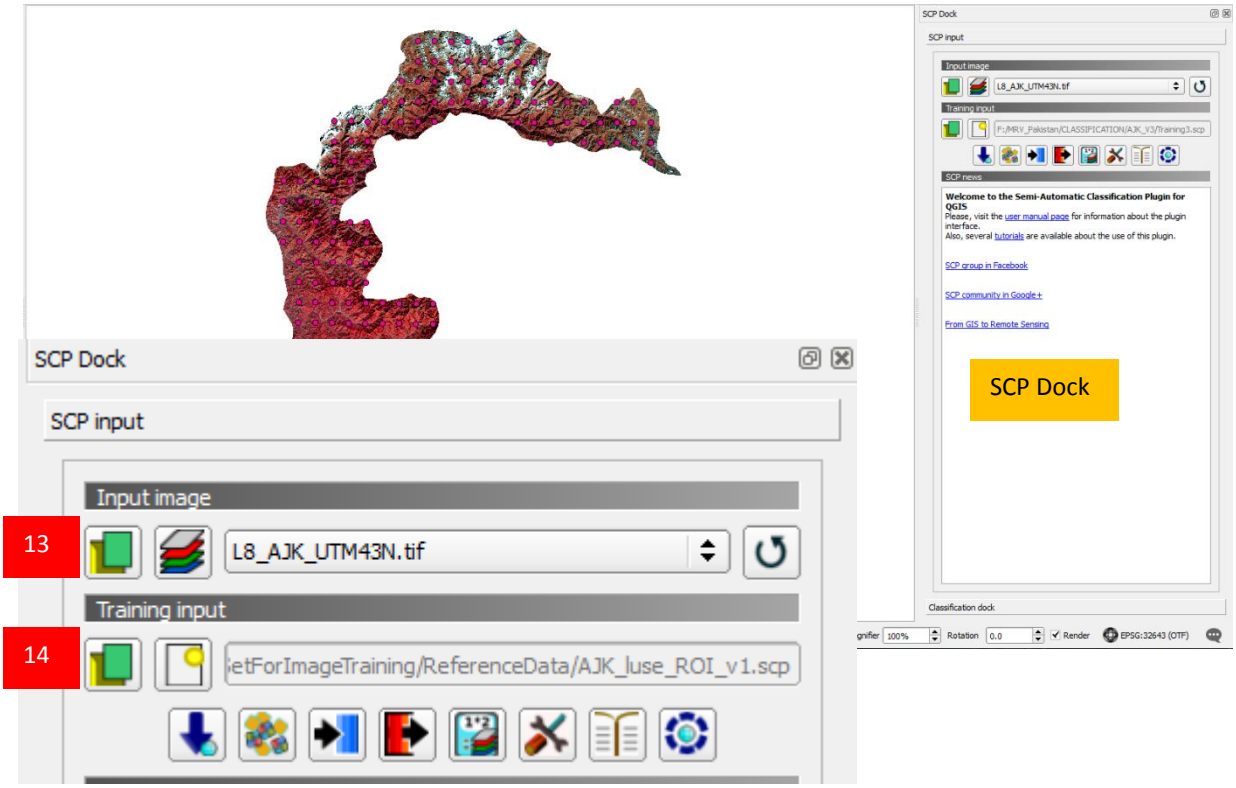
12. Enable ***SCP Dock,*** from the *View>Panels>SCP Dock* menu to enable SCP docking in the right side of QGIS window.
13. *Open/Select* the image file
14. *Create a new training input* and save the file as *"ajk_luse_ROI.scp".*
15. Go back to the Multiple ROI Creation tab in the SCP tool and enable *Calculate sig and Run the process.*
16. *Save* the generated ROI data *'ajk_luse_ROI.scp' as a* shape file *'ajk_luse_ROI.shp'* to be used for generating training and sample ROI subsets.



Caution!

This process may take several minutes to complete depending on the number of plots, area of interest for analysis and computer resources. Do not use any other QGIS tools during the process. It is recommended to leave QGIS without any disturbances to the process.

You may see some Python error. This is likely to occur if QGIS and SCP plugins are not installed properly.

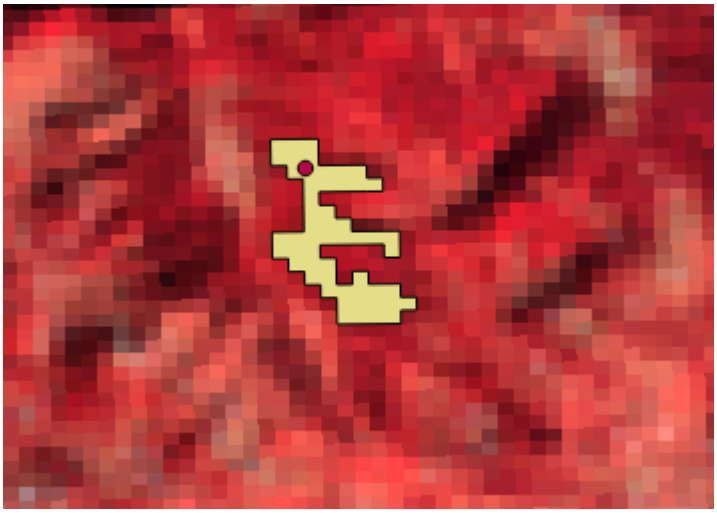


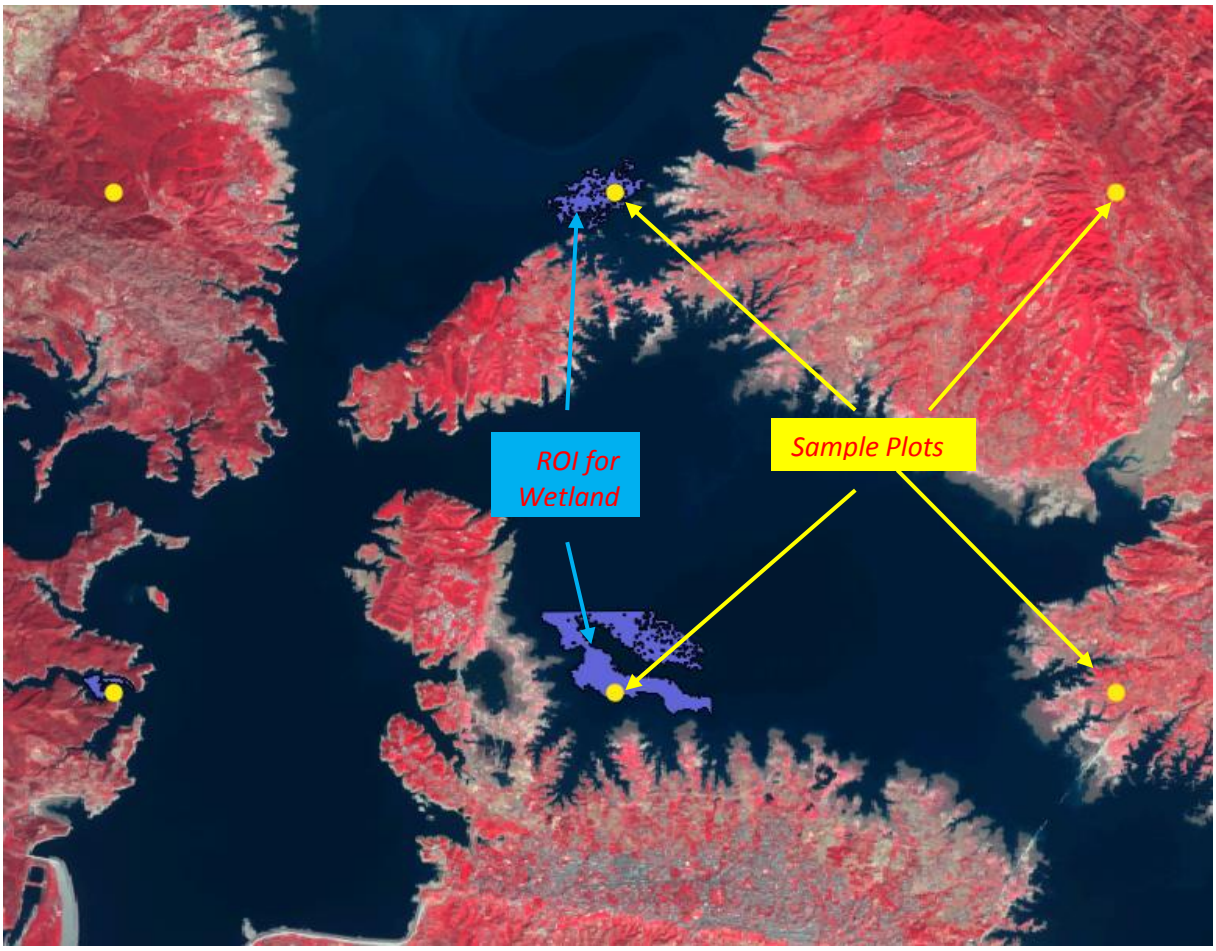
AJK_luse_ROI_v1 :: Features total: 193, filtered: 193, selected: 0

	MC_ID	MC_info	C_ID	C_info	SCP_UID
1	3	Grassland	3	Grassland	20170803_1028...
2	3	Grassland	3	Grassland	20170803_1028...
3	3	Grassland	3	Grassland	20170803_1028...
4	3	Grassland	3	Grassland	20170803_1028...
5	6	Other Land	6	Other Land	20170803_1028...
6	3	Grassland	3	Grassland	20170803_1028...
7	3	Grassland	3	Grassland	20170803_1028...
8	3	Grassland	3	Grassland	20170803_1028...
9	3	Grassland	3	Grassland	20170803_1028...
10	3	Grassland	3	Grassland	20170803_1028...
11	3	Grassland	3	Grassland	20170803_1028...
12	3	Grassland	3	Grassland	20170803_1028...
13	3	Grassland	3	Grassland	20170803_1028...
14	6	Other Land	6	Other Land	20170803_1028...
15	6	Other Land	6	Other Land	20170803_1028...
16	6	Other Land	6	Other Land	20170803_1028...
17	3	Grassland	3	Grassland	20170803_1028...

Show All Features

17. Explore and navigate the generated ROI polygons. These ROI polygons are observed as shown here

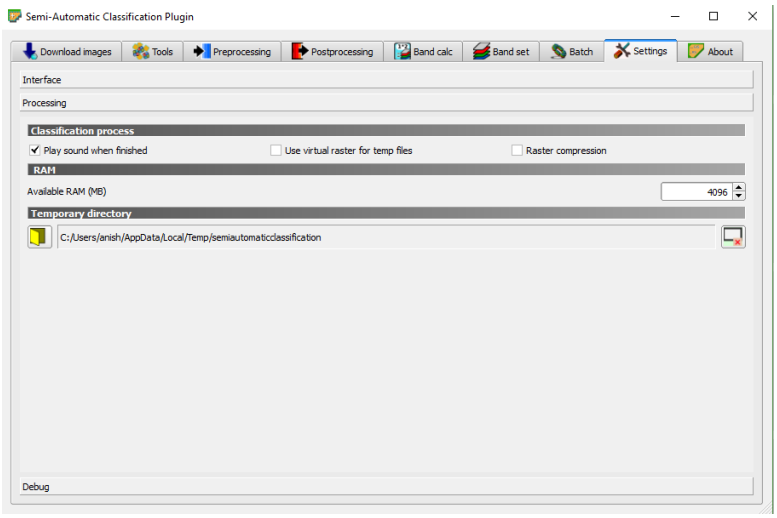




Caution!

If the process/application crashes and shows error message, increase the processing RAM of SCP tool using Settings. The recommended value for the available RAM in SCP settings is half of the system RAM. For instance, if your system has 8 GB of RAM, set the value to 4096 MB

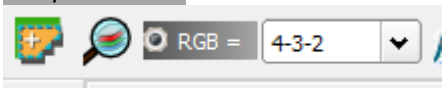
Also clear the temporary directory to clear the cache memory. It is advisable to use computer with minimum 8 GB RAM resources for this process.



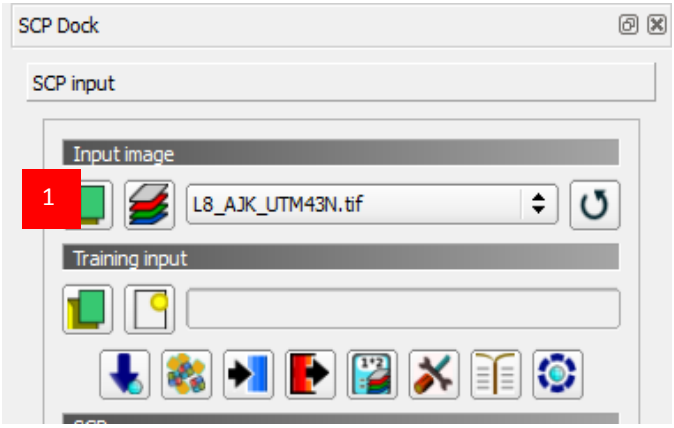
Step 3B Manual Generation of Region of Interest (ROI) Polygons [1]

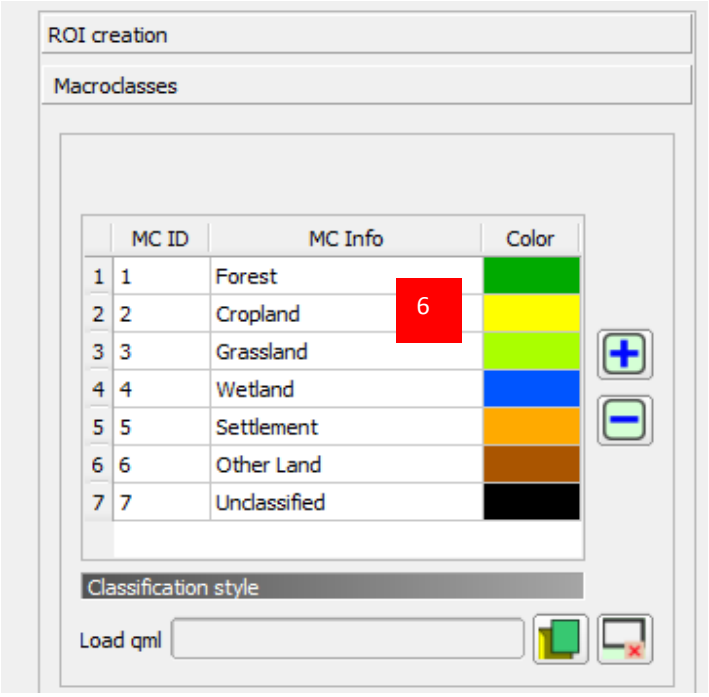
Region of Interest (ROI) for image signature can be generated manually as well, by selecting the 'seed' pixels manually in Semi-Automatic Classification plugin. ROIs can be generated manually for each of the systematic point sample, or can also be generated randomly, selecting seed pixels at random locations. However, these seed pixels need to be "pure pixels" i.e. should not contain mixed land cover/land use categories.

1. Enable ***SCP Dock,*** from the *View>Panels>SCP Dock* menu to enable SCP docking in the right side of QGIS window.
2. *Open/Select* the image file.
3. Image band composite cane be changed to 4-3-2 or 3-2-1 band sequence using *band composite tool* in SCP toolbar.



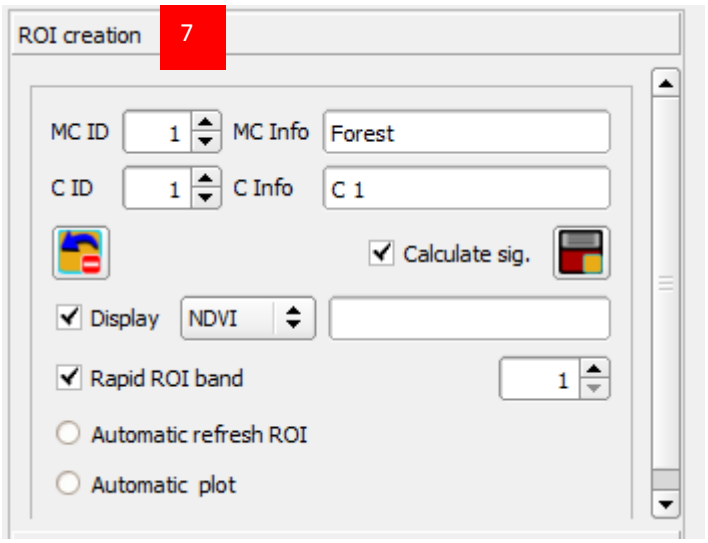
4. For reference, you can use *'ajk_luse_interpreted.shp'* plots to guide to the locations where ROIs are to be created.
5. *Create a new training input* and save the file as *"ajk_maual_ROI.scp"*.
6. In the SCP Dock, go to *Microclasses* and *define Macro Class ID, Macro Class Info and desired color* for Macro Class. These Macro Classes are our six IPCC Land Use Classes plus an Unclassified (if necessary) as shown.





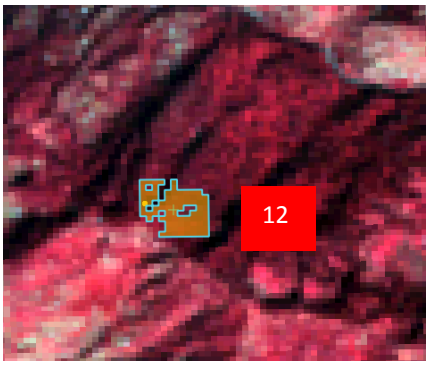
[1] Automatic ROI generation in Step 3A may encounter with software crash due to available RAM in the computer and potential bug in the Semi-automatic Classification Plugin. In this case, manual generation of ROIs described in Step 3B should be followed.

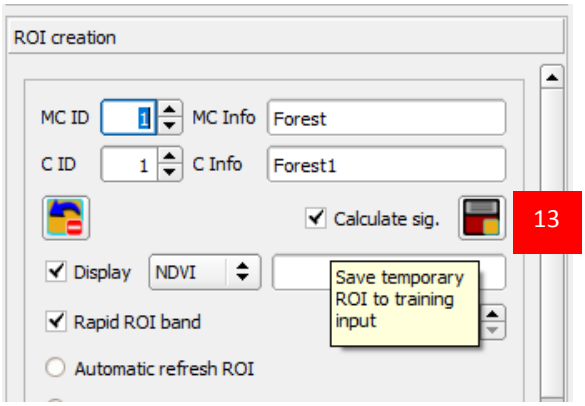
7. Go back to **ROI Creation** tab in the SCP Dock.
8. In the SCP toolbar, enable *Activate ROI Pointer* to use the cursor pointer to generate ROIs. In the **Dist.** enter **0.001**, which is radiometric distance between pixels (i.e. pixel values in reflectance of the TOA corrected image we are using).
9. Similarly, in **Min** enter **2** (Minimum area of ROI in pixel units i.e. 2 pixels in this case).
10. In **Max**, enter **10** to define size of the square that inscribes the ROI, i.e. maximum width of the ROI generated in pixel units.

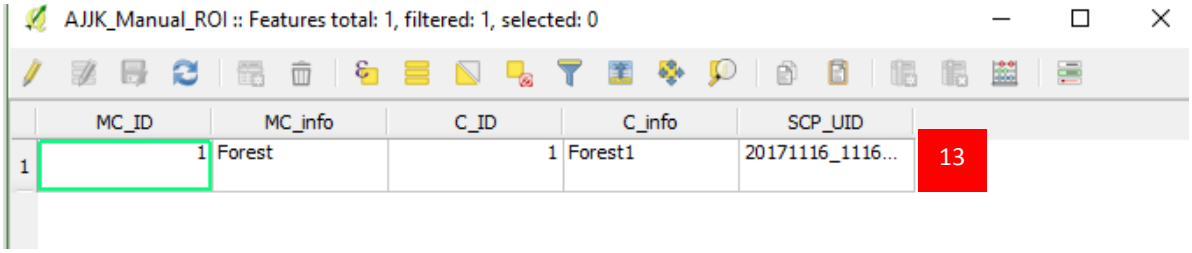


[Note: Smaller the ROI for a 'seed' pixel, it is likely to contain pure pixels required for classification]

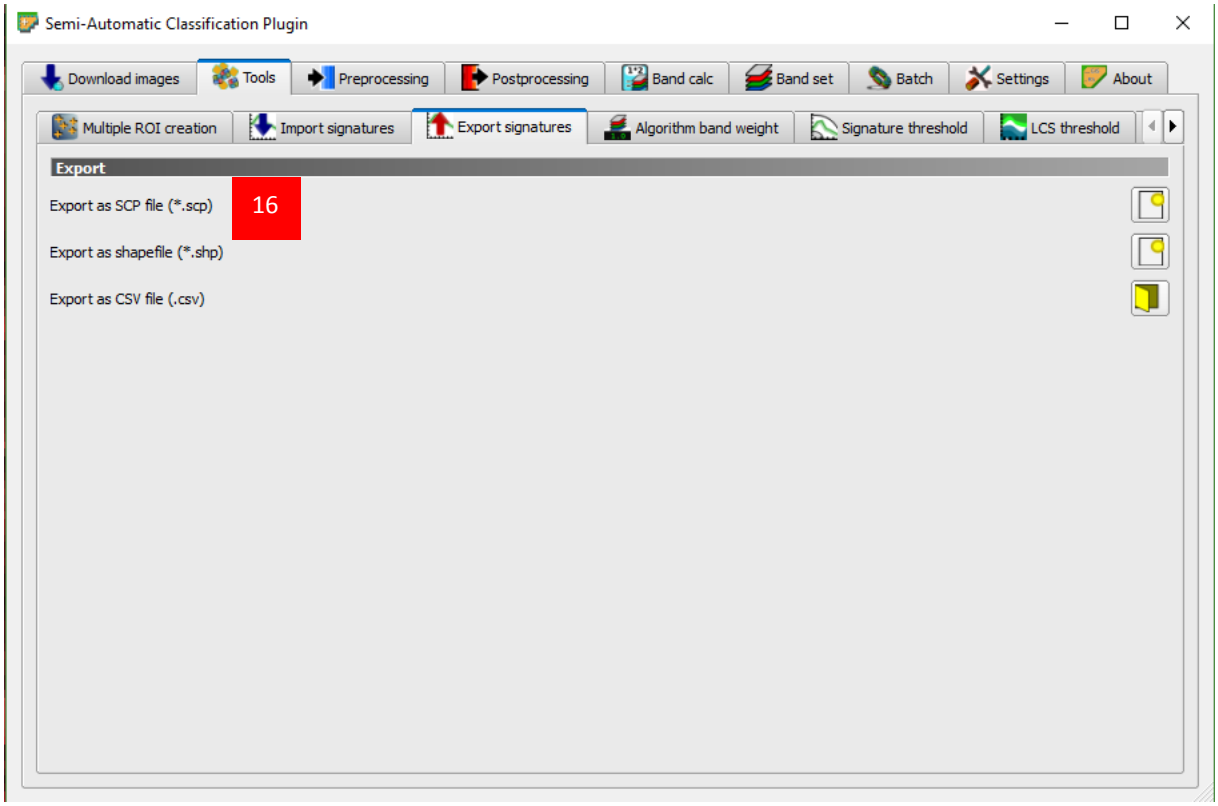
11. Go to the location of one of the plots, and using the **Activate ROI pointer**, click on the nearest pure pixel. ROI is created, clustering the nearest pixel as shown.
12. In the ROI creation, use the **MC ID** to browse to the correct ID and associated MC Info class defined in Step 6. For e.g. MC ID 1 for MC Info Forest. For each individual ROIs, you can assign individual **C ID** as 1…and **C Info** as unique values Forest1, Forest2, Forest3…..Cropland1, Corpland2 and so on.
13. Once ROI is created, use *Save Temporary ROI to training input* to save individually created ROIs to the earlier created *'ajk_manual_ROI'* ROI file. The created ROI attributes is stored in the attribute table of the data file as shown.
14. Create more ROIs following the above steps







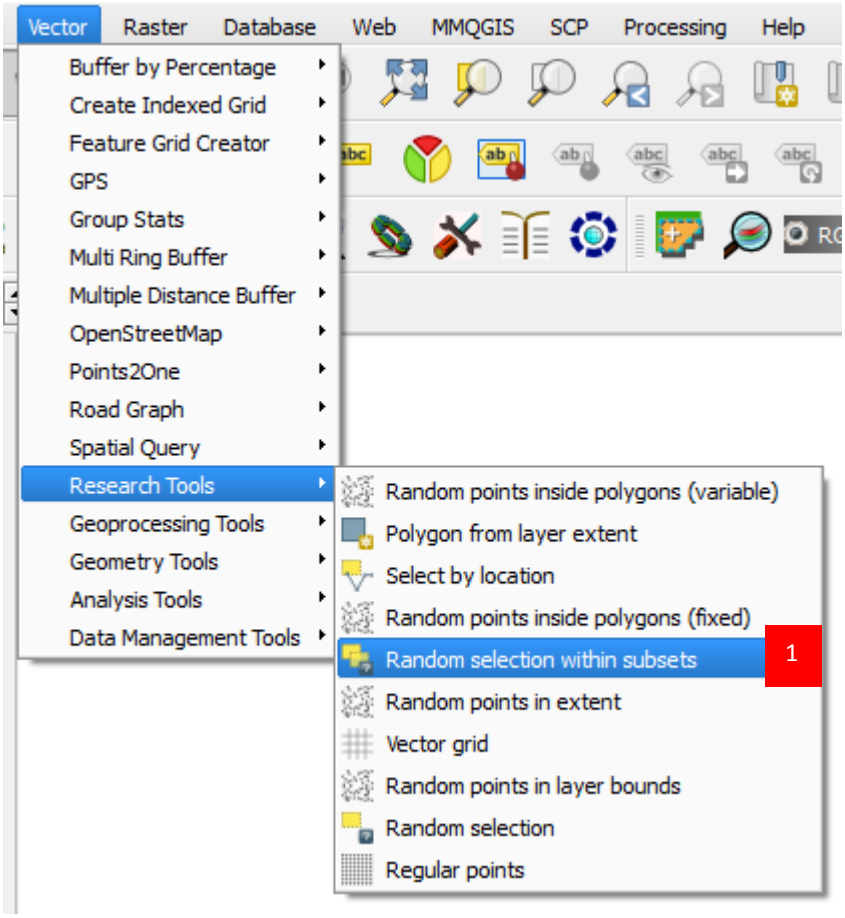
15. Once sufficient ROIs have been created, export this ROI file '*ajk_manual_ROI*' to a new shape file as *ájk_luse_ROI.shp'* to be used for further random sub-set selection and/or classification.
16. Go to *Semi-automatic Classification Plugin>Tools>Export Signatures* and use Export as shapefile (*.shp) to export populated ROI signature file to a permanent shape file to be used for classification in next exercise.



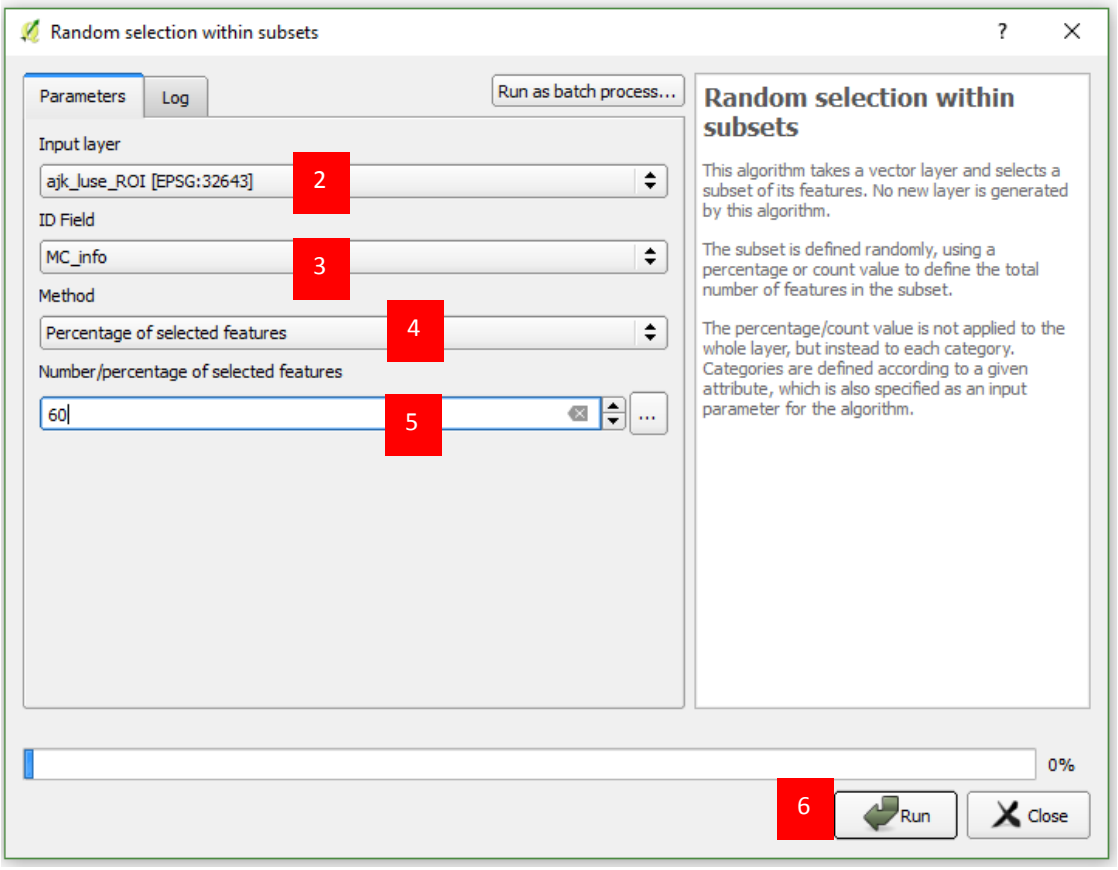
4.1.3. Training and Validation Subsets

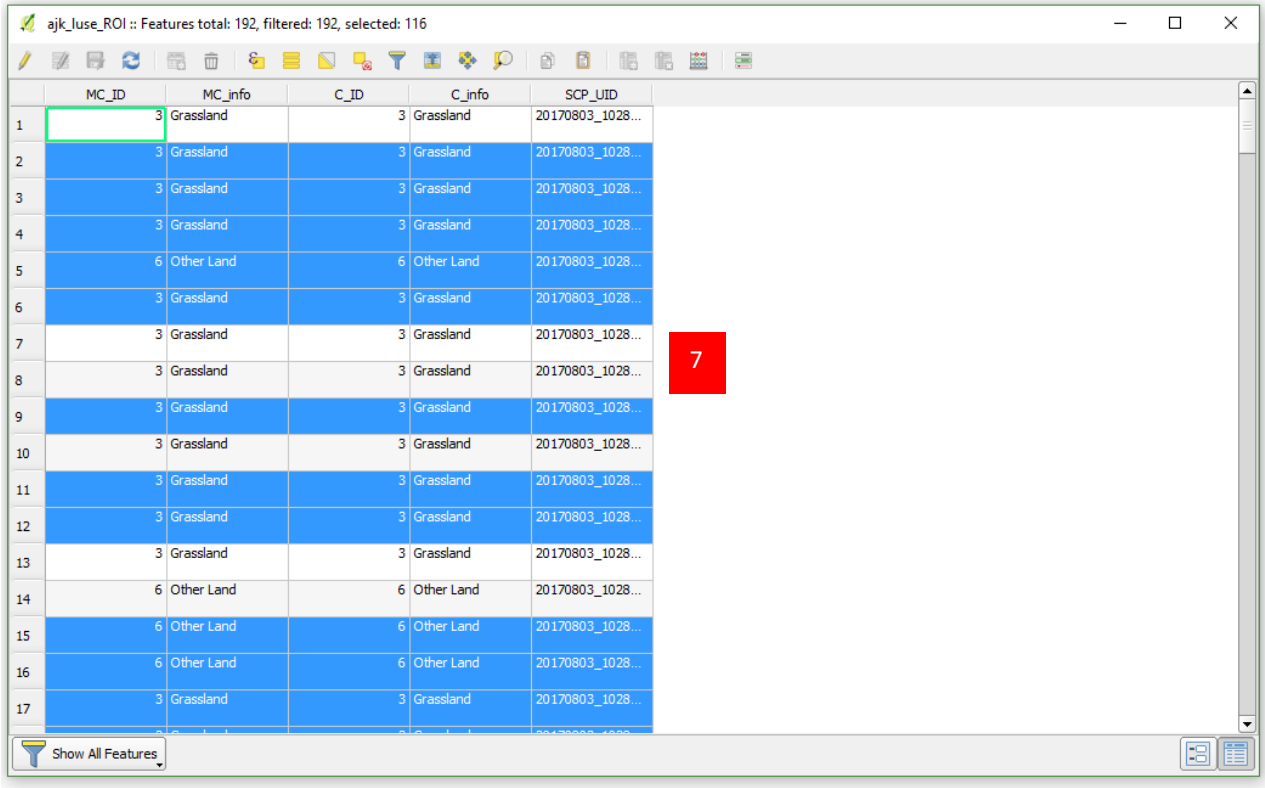
Step 4 Generating Training and Validation Subsets

1. Once the ROI polygons are created, image training and validation sets are created using Random Sampling in *Vector>Research Tools>Random selection within subset* menu.

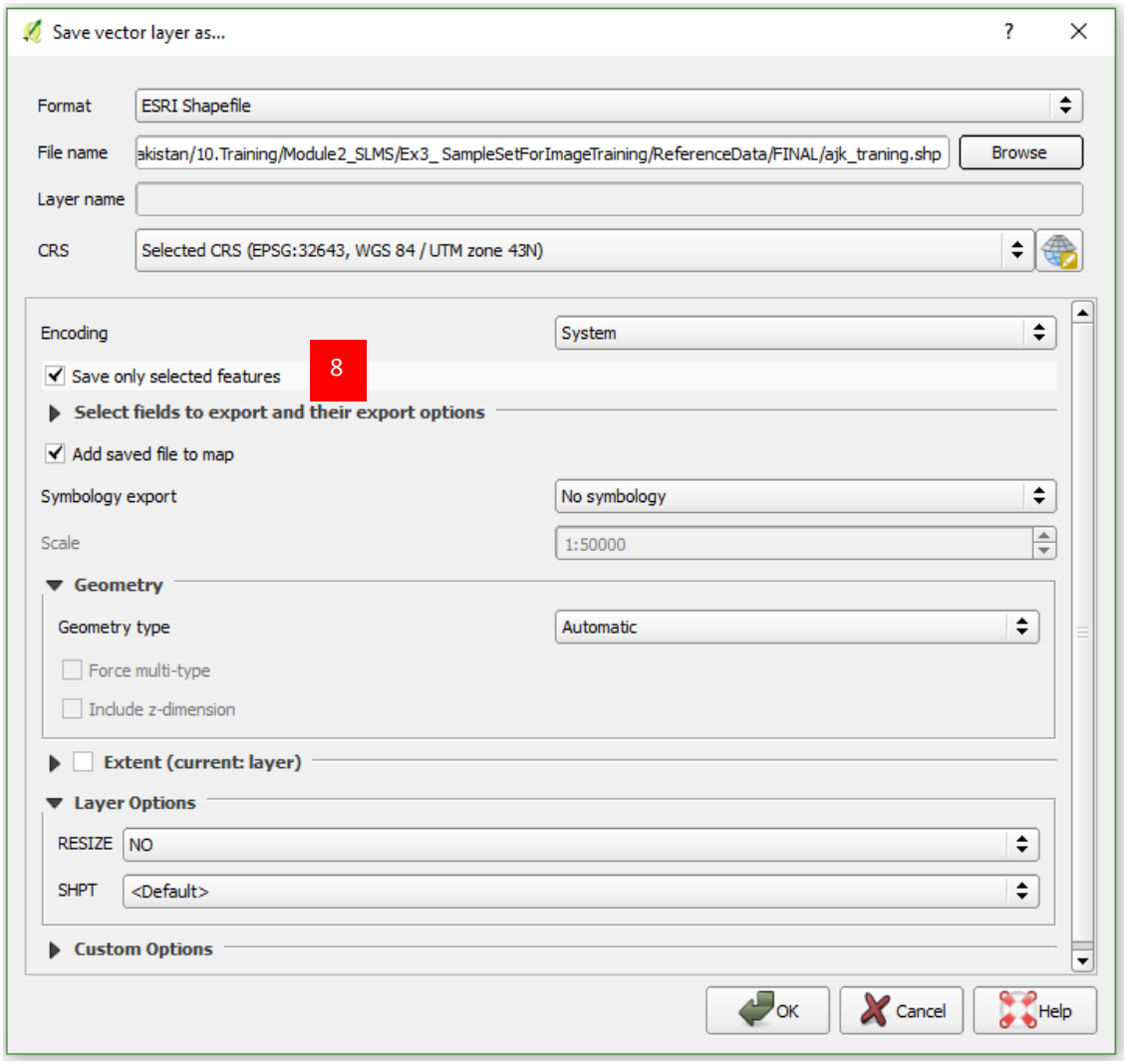


2. User *ájk_luse_ROI.shp'* as the Input Layer.
3. Select MC_info as Id Field for subset/strata.
4. Use *Percentage of selected features* as Method.
5. Enter 60 for *Number/percentage of selected features* to indicate 60% of the ROI polygons will be selected as training subset and remaining 40% as validation subset. From all the strata defined by MC_info field, 60% of each is selected.
6. *Run* the process.

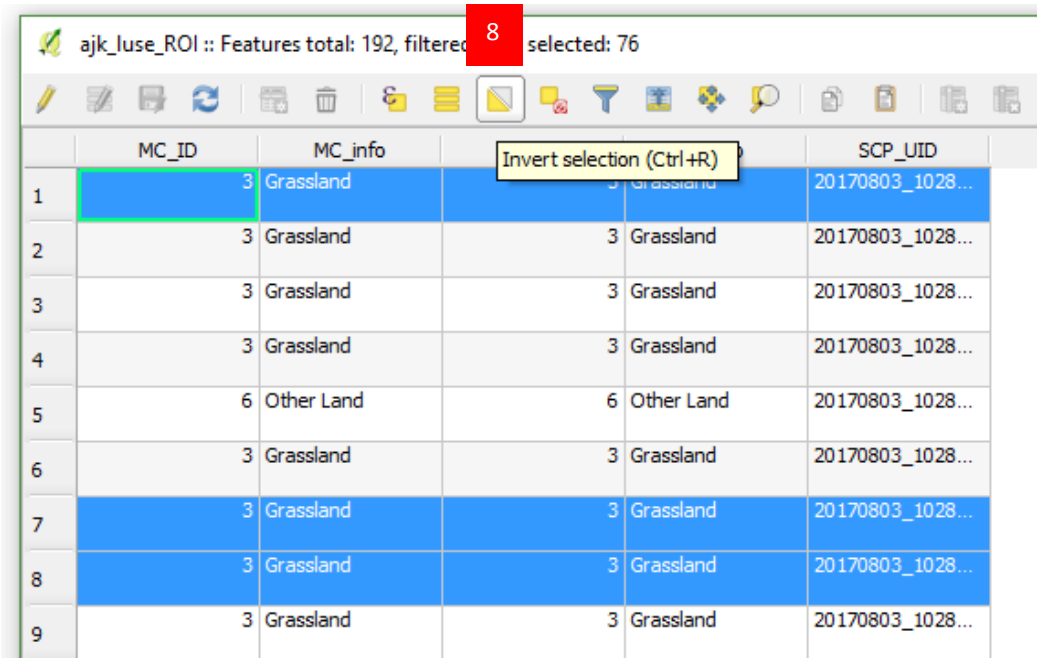

Random selection within subsets
Parameters
Log
Run as batch process...
Input layer
ajk_luse_ROI [EPSG:32643]
2
ID Field
MC_info
3
Method
Percentage of selected features
4
Number/percentage of selected features
60
5
Random selection within subsets
This algorithm takes a vector layer and selects a subset of its features. No new layer is generated by this algorithm.
The subset is defined randomly, using a percentage or count value to define the total number of features in the subset.
The percentage/count value is not applied to the whole layer, but instead to each category. Categories are defined according to a given attribute, which is also specified as an input parameter for the algorithm.
0%
6
Run
Close


ajk_luse_ROI :: Features total: 192, filtered: 192, selected: 116
MC_ID MC_info C_ID C_info SCP_UID
1 3 Grassland 3 Grassland 20170803_1028...
2 3 Grassland 3 Grassland 20170803_1028...
3 3 Grassland 3 Grassland 20170803_1028...
4 3 Grassland 3 Grassland 20170803_1028...
5 6 Other Land 6 Other Land 20170803_1028...
6 3 Grassland 3 Grassland 20170803_1028...
7 3 Grassland 3 Grassland 20170803_1028...
8 3 Grassland 3 Grassland 20170803_1028...
9 3 Grassland 3 Grassland 20170803_1028...
10 3 Grassland 3 Grassland 20170803_1028...
11 3 Grassland 3 Grassland 20170803_1028...
12 3 Grassland 3 Grassland 20170803_1028...
13 3 Grassland 3 Grassland 20170803_1028...
14 6 Other Land 6 Other Land 20170803_1028...
15 6 Other Land 6 Other Land 20170803_1028...
16 6 Other Land 6 Other Land 20170803_1028...
17 3 Grassland 3 Grassland 20170803_1028...
7
Show All Features

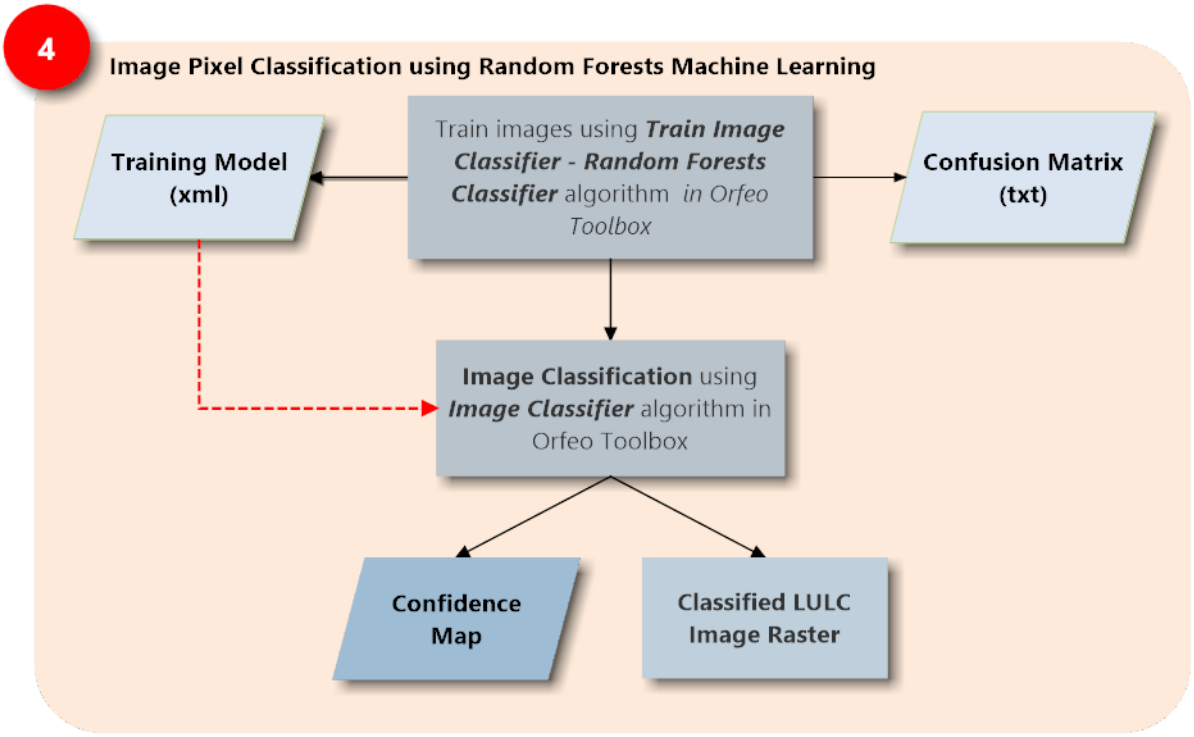
7. *Open Attribute table* of *ájk_luse_ROI.shp'* and review the selected ROI polygon.
8. *Save* the selected data as a separate file *'ajk_traning.shp'* making sure that the option ***Save only selected features*** is enabled.



9. In the attribute table, use *Invert selection* tool to toggle the selected records to remaining 40%.
10. *Save* the remaining selected data as a separate file *'ajk_validation.shp'* making sure that the option ***Save only selected features*** is enabled.


8
ajk_luse_ROI :: Features total: 192, filtered
selected: 76
Invert selection (Ctrl+R)
MC_ID
MC_info
SCP_UID
1 3 Grassland 20170803_1028...
2 3 Grassland 3 Grassland 20170803_1028...
3 3 Grassland 3 Grassland 20170803_1028...
4 3 Grassland 3 Grassland 20170803_1028...
5 6 Other Land 6 Other Land 20170803_1028...
6 3 Grassland 3 Grassland 20170803_1028...
7 3 Grassland 3 Grassland 20170803_1028...
8 3 Grassland 3 Grassland 20170803_1028...
9 3 Grassland 3 Grassland 20170803_1028...

5. Image Classification using Random Forests



Landsat 8 OLI image is trained using the generated ROI Training subsets for random Forests Classifier in OrfeoToolBox Monteverdi Interface. This can be done within QGIS OrfeoToolBox geoalgorithm as well.

Random Forests algorithm generates Confusion Matrix using the validation subset during the process of training the image.

Image is then classified using the training model generated producing classification raster and confidence raster.

5.1. Exercise 4 Image Classification Using Random Forests

5.1.1. Background

Random Forests (RF) is a classification and regression algorithm used in machine learning and is increasingly being used to satellite image classification for creation of continuous field data such as percent tree cover, biomass, land cover etc. Random Forests is a non-parametric regression model capable of using continuous and categorical data and is easy to parametrize, is non-sensitive to over fitting, can handle outliers in training data and also calculates ancillary information such as classification error and variable importance.

RF algorithm is based on ensemble model, that is it uses the results from many different models to calculate a response. In RF, several decision trees are created and the response is calculated based on the outcome of all the decisions trees. Some common applications of random forests classification include land cover and land cover change mapping. Regression applications include continuous filed mapping e.g. percent tree cover, percent shrub cover, impervious surfaces and biomass mapping.

Data Used

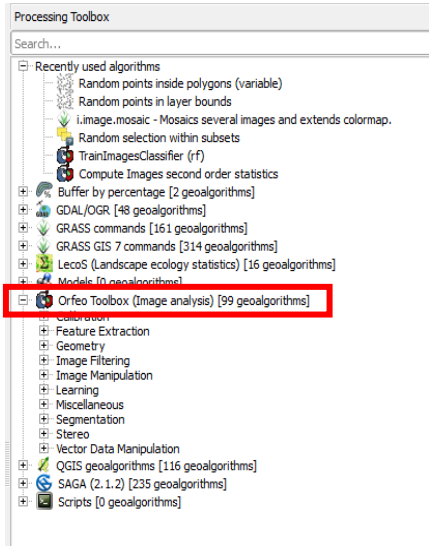
SN	Datasets	Sources	Descriptions
1	ROI training data *(ajk_training.shp)*	ROI generated training set in Exercise 3	ROI generated in Exercise 3
2	ROI validation data *(ajk_vlidation.shp)*	ROI generated validation set in Exercise 3	ROI generated in Exercise 3
3	Landsat 8 OLI image scenes mosaick (150-36, 149-36, 149-37) *(L8_AJK_Mosaic.tif)*	Mosaicked L8 image from Exercise 1	Landsat 8 OLI atmospherically corrected image scene mosaic

Software Used

- Orfeo ToolBox[1] Ver 6.0.0 – Open source remote sensing analysis software
- Quantum GIS (QGIS) [2] Ver 2.18 (or older) - Open source GIS software

[1] Latest version of Orfeo Toolbox can be downloaded from https://www.orfeo-toolbox.org/
[2] Latest version of QGIS can be downloaded from http://www.qgis.org/

[Note: Orfeo ToolBox can be used from within QGIS Interface from Processing Toolbox geoalgorithm. However, for greater control on the options, it is recommended to use standalone Orfeo ToolBox with Monteverdi Graphical User Interface (GUI). In this exercise, standalone Orfeo ToolBox with Monteverdi GUI is used]

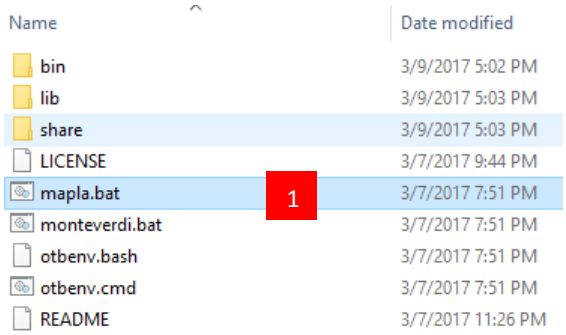


5.1.2. Image Classification

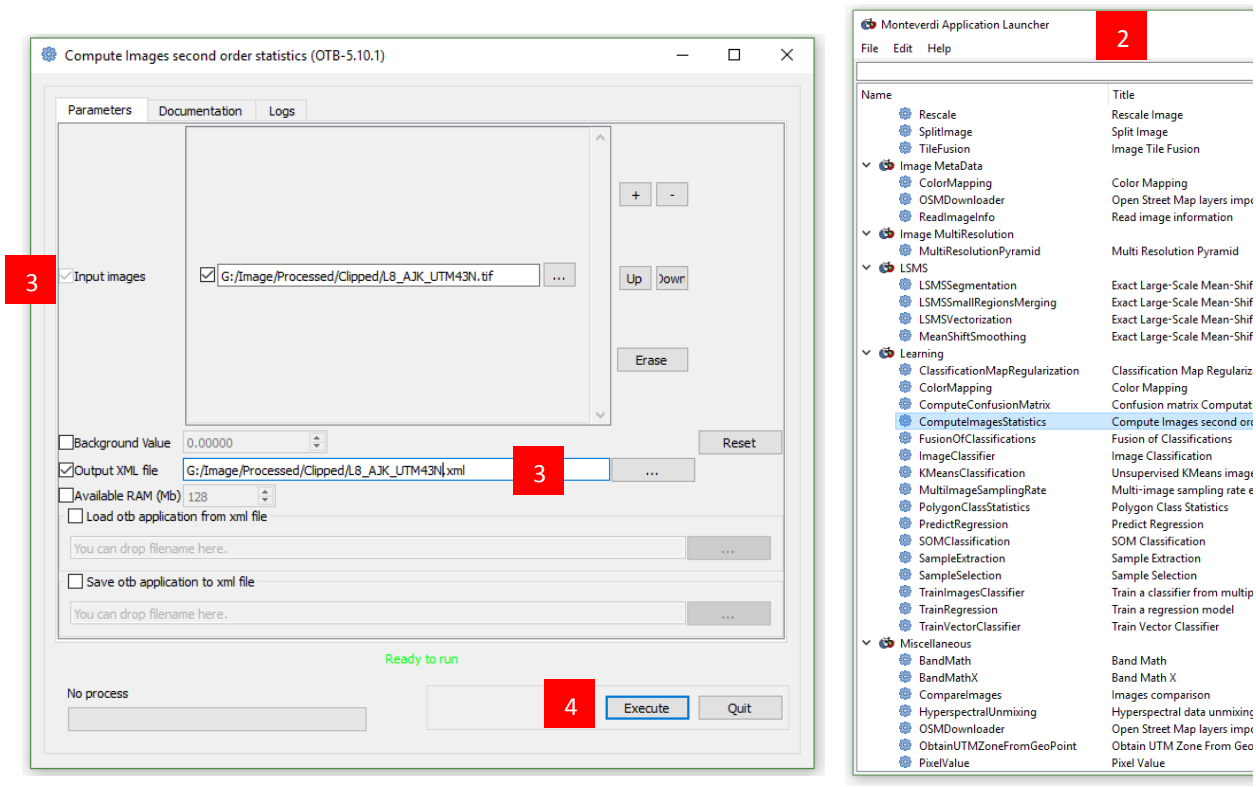
Step 1 Compute Image Second Order Statics

Image Second Order Statistics is computed to generate global mean and standard deviation for each band of a set of images and optionally saves the results in an XML file. The output XML is intended to be used an input for the **TrainImagesClassifier** application to normalize samples before learning.

1. From the Orfeo ToolBox installer folder, use **mapla.bat** (i.e. double click on maple.bat) batch file to use Orfeo ToolBox Monteverdi Application Launcher.
2. Browse through the list of algorithms and select *Learning>ComputeImageStatistics*.
3. In *Compute Image second order statics* window, browse image *'L8_AJK_UTM43n.tif'* as input and an output filename *'L8_AJK_UTM43N_stat.xml'* in output xml file.



4. *Execute* the process

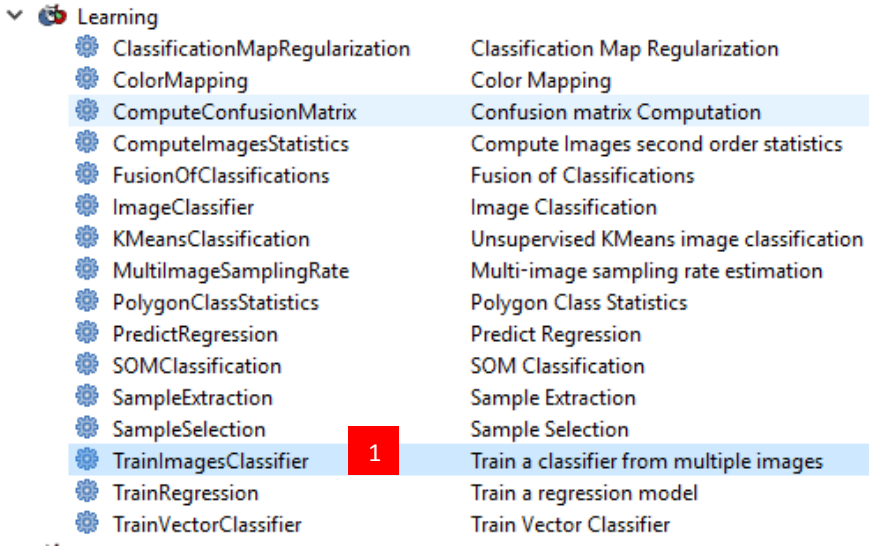


5. The second order statistics of the image is calculated in xml *'L8_AJK_UTM43N_stat.xml'* file as shown. The XML file can be open in any text editor. This will be used as an input for image training.

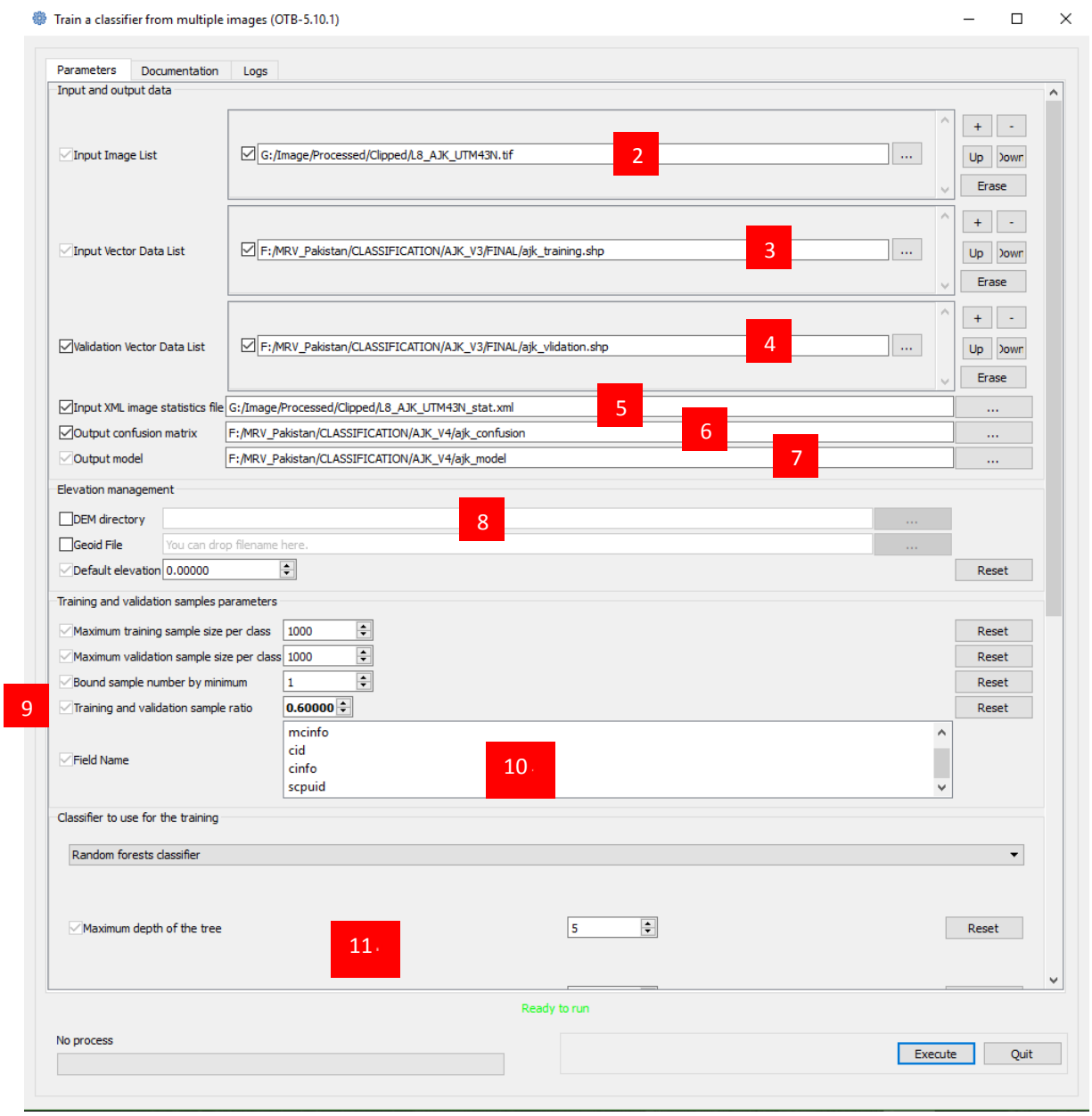
```
<?xml version="1.0" ?>
<FeatureStatistics>
    <Statistic name="mean">
        <StatisticVector value="7.8386" />
        <StatisticVector value="7.07163" />
        <StatisticVector value="6.89425" />
        <StatisticVector value="14.1226" />
        <StatisticVector value="12.0016" />
        <StatisticVector value="8.64314" />
    </Statistic>
    <Statistic name="stddev">
        <StatisticVector value="14.611" />
        <StatisticVector value="13.9003" />
        <StatisticVector value="14.3612" />
        <StatisticVector value="25.2968" />
        <StatisticVector value="22.6405" />
        <StatisticVector value="17.1568" />
    </Statistic>
</FeatureStatistics>
```

Step 2 Train Random Forests Classifier

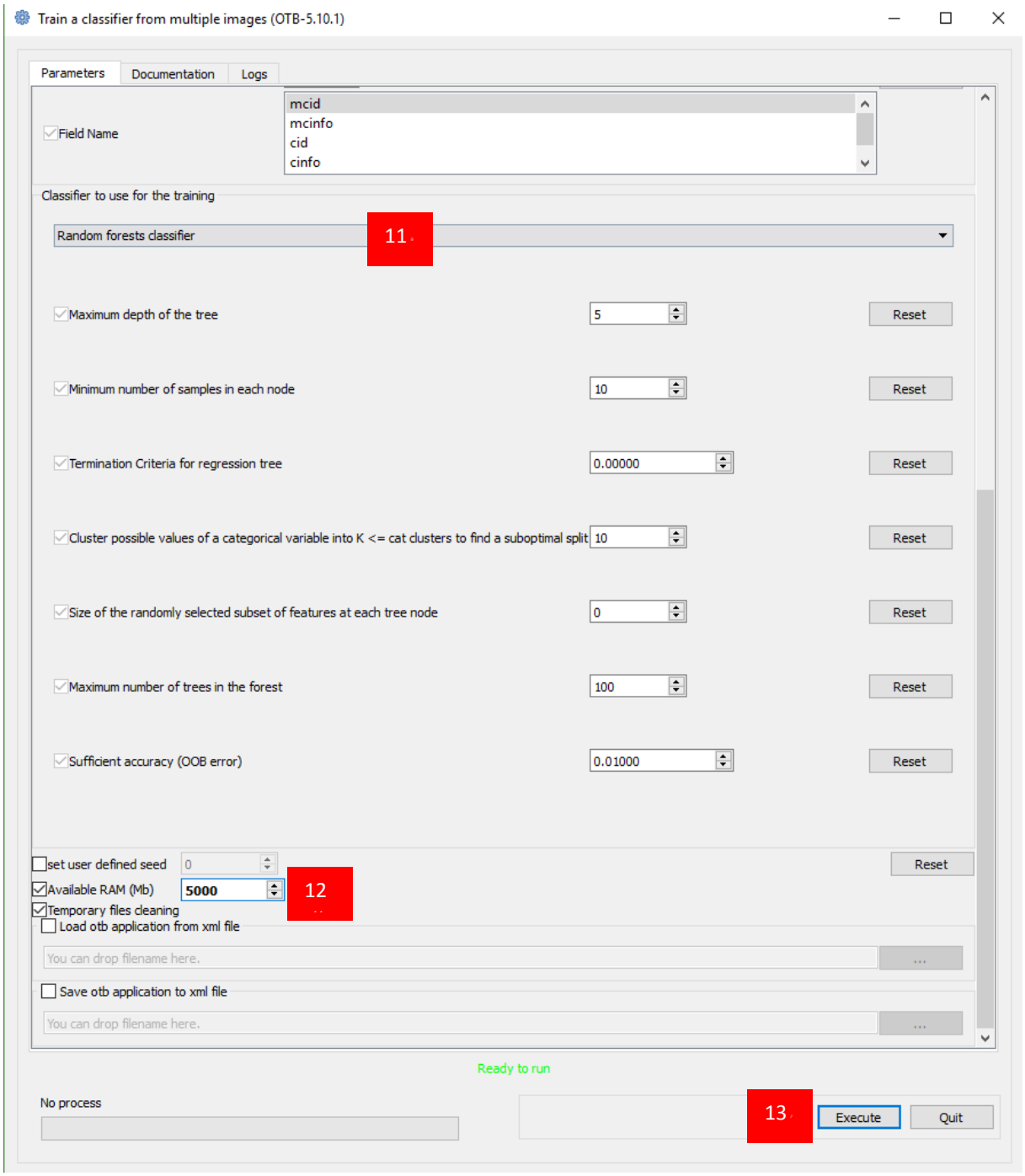
1. In Monteverdi Application launcher go to *Learning>TrainImagesClassifier*



2. Browse to the image file *'L8_AJK_UTM43n.tif'*
3. Browse to training dataset file '*ajk_training.shp'* for **Input Vector Data List.**
4. Browse to validation dataset file '*ajk_validation.shp'* for **Validation Vector Data List.**
5. Browse previously created image statistics file *'L8_AJK_UTM43N_stat.xml'* for **Input XML image statistics file.**
6. Browse to the output folder and give name *'ajk_confusion'* for generating confusion matrix from the RF training. This will later be used for accuracy assessment and area estimation in Exercise 5.
7. Browse to output folder and save ***'ajk_model'*** file which is and output of generate RF training and is sued later for classification. *[Note: This is the RF model file and is very important to save appropriately. In Steps 6 and 7, **DO NOT** give any extension to the confusion matrix and training model outputs]*
8. You can choose to use DEM file as input parameter for RF regression.
9. Scroll down the window enter **0.6 (60%)** for **Training and validation sample ratio.**
10. Select field name **mcid** (this is the field name MCID of training set ROI file *'ajk_traning.shp'*. That is, we are using **mcid** field data (which contains 1,2,3,4,5,6,7lookup vales corresponding to Forest, Cropland, Grazing, Wetland, Settlement, Other Land and NoData/Unclassified. These will be used as training classes (strata).



11. Scroll down the window and in **Classifier to use for the training,** select ***Random forests classifier.*** Leave other parameters as default for this exercise. You can increase the CPU Ram for processing. *[Note: CPU RAM depends on your computer configuration]*
12. After defining all the above parameters, *Execute* the algorithm


Train a classifier from multiple images (OTB-5.10.1)
Parameters
Documentation
Logs
Field Name
mcid
mcinfo
cid
cinfo
Classifier to use for the training
Random forests classifier
11
Maximum depth of the tree
5
Reset
Minimum number of samples in each node
10
Reset
Termination Criteria for regression tree
0.00000
Reset
Cluster possible values of a categorical variable into K <= cat clusters to find a suboptimal split
10
Reset
Size of the randomly selected subset of features at each tree node
0
Reset
Maximum number of trees in the forest
100
Reset
Sufficient accuracy (OOB error)
0.01000
Reset
set user defined seed
0
Reset
Available RAM (Mb)
5000
12
Temporary files cleaning
Load otb application from xml file
You can drop filename here.
Save otb application to xml file
You can drop filename here.
Ready to run
No process
13
Execute
Quit

13. Two output files *ájk_confusion'* and *ájk_model'* are created. These are both txt files and can be open in any text editor.

```
#Reference labels (rows):1,2,3,4,5,6,7
#Produced labels (columns):1,2,3,4,5,6,7
0,6,0,0,2,3,4
0,9,0,0,6,0,0
0,7,1,0,2,4,1
0,0,0,15,0,0,0
0,4,0,0,3,8,0
0,0,2,0,3,10,0
0,4,0,0,4,4,3
```

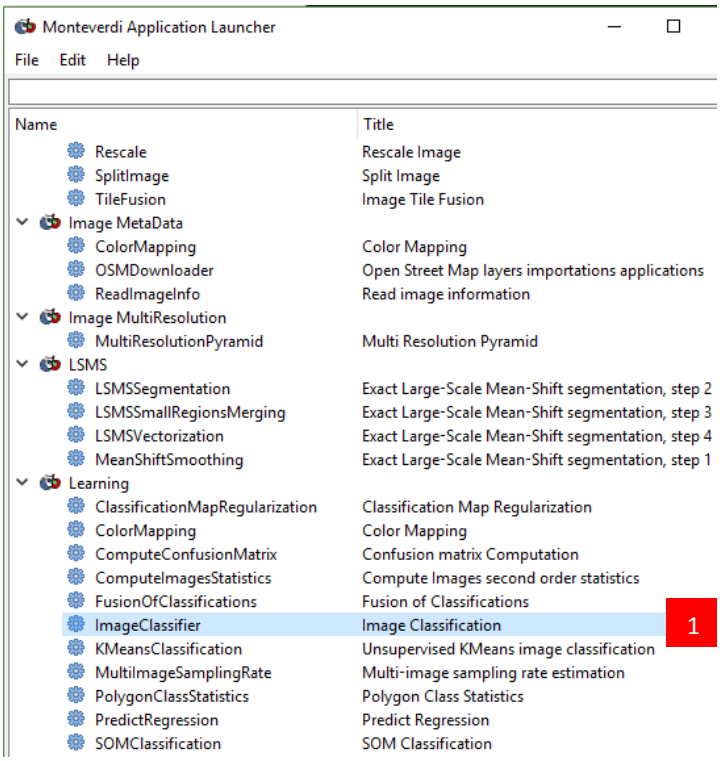
Figure 10 Confusion matrix text output

```
%YAML:1.0
my_random_trees: !!opencv-ml-random-trees
   nclasses: 7
   nsamples: 385
   nactive_vars: 2
   oob_error: 3.0821917808219179e-01
   ntrees: 100
   is_classifier: 1
   var_all: 6
   var_count: 6
   ord_var_count: 6
   cat_var_count: 0
   training_params:
      use_surrogates: 0
      max_categories: 10
      max_depth: 5
      min_sample_count: 10
      cross_validation_folds: 0
      priors: !!opencv-matrix
         rows: 1
         cols: 7
         dt: d
         data: [ 1., 1., 1., 1., 1., 1., 1. ]
   var_type: [ 0, 0, 0, 0, 0, 0 ]
   cat_count: !!opencv-matrix
      rows: 1
      cols: 1
      dt: i
      data: [ 7 ]
   cat_map: !!opencv-matrix
      rows: 1
      cols: 7
      dt: i
      data: [ 1, 2, 3, 4, 5, 6, 7 ]
   trees:
      -
```

Figure 11 Random Forest training model output

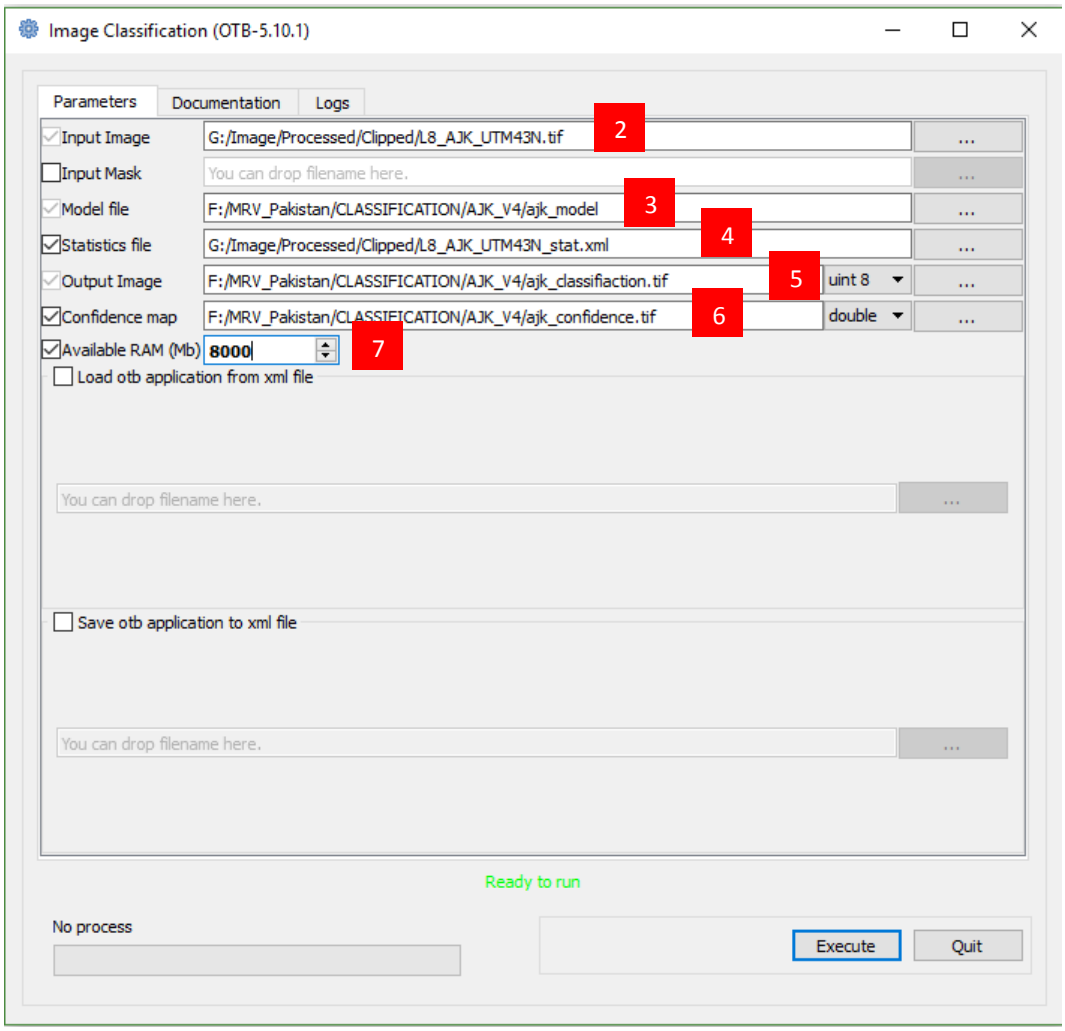
Step 3 Execute Random Forests Image Classifier

14. In Monteverdi Application launcher go to *Learning>ImageClassifier*



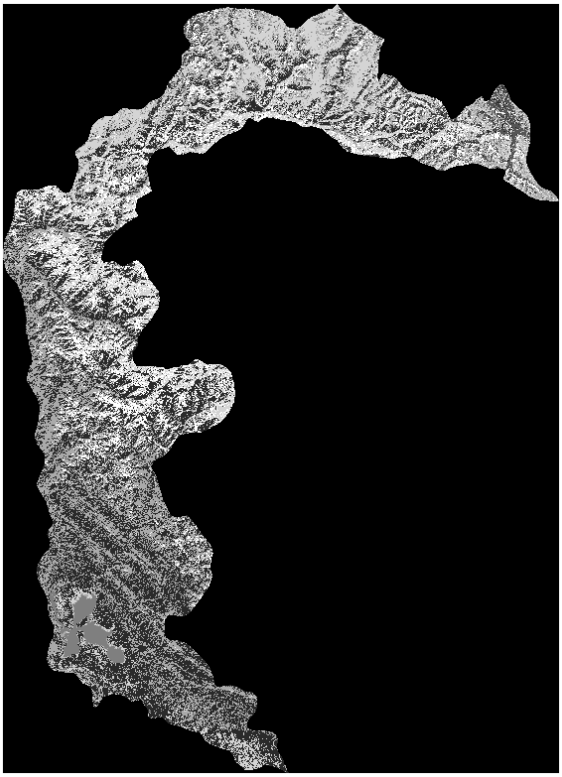
15. Browse to the image file *'L8_AJK_UTM43n.tif'*
16. Browse to the training mode file '*ajk_model'* for **Model file.**
17. Browse previously created image statistics file *'L8_AJK_UTM43N_stat.xml'* for **Statistics file** (though this is optional).
18. In the **Output Image,** browse to the output folder and save classification raster file as ***'ajk_classification.tif'.***
19. Similarly, classification confidence level raster can be generated and saved as 'a*jk_confidence.tif*'.
20. Processing **Available RAM** can be increased to allocate higher processing resources depending on our computer's configuration.

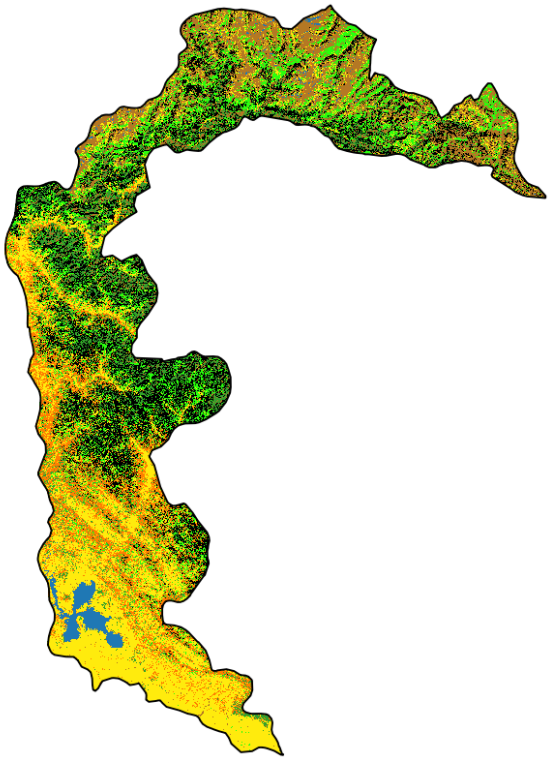
[Note: Processing may take a while 5-15 minutes depending on your computer configuration]



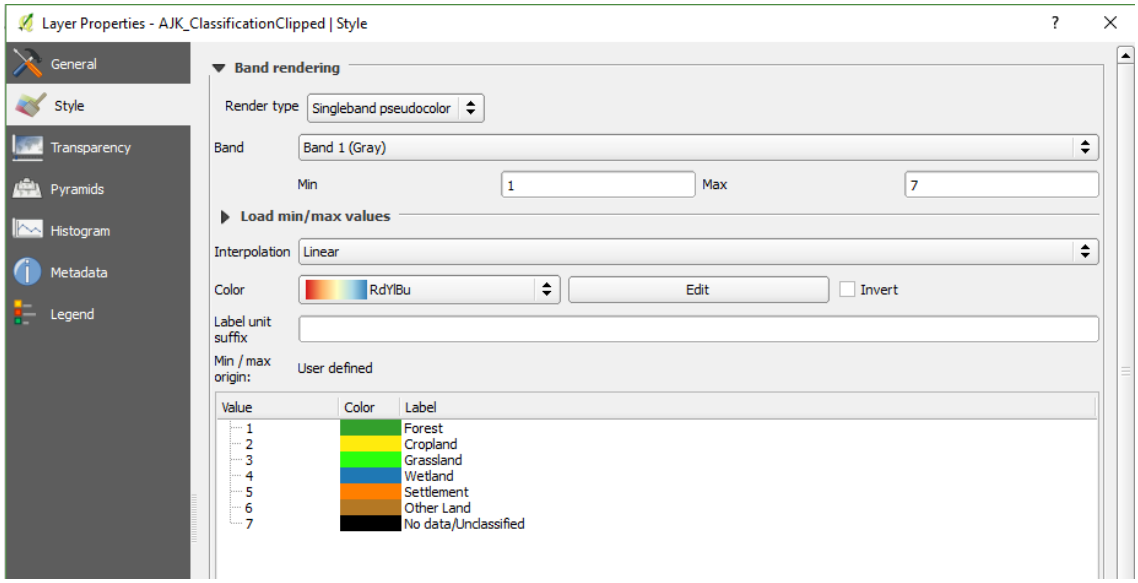
Step 4 Land Use Map Composition in QGIS

1. In QGIS, add recently created *'ajk_classification.tif'* raster.
2. Using AJK Province boundary polygon file, clip the raster to exclude the NoData values beyond the AJK boundary and save new clipped raster dataset as ***'ajk_luse.tif'***

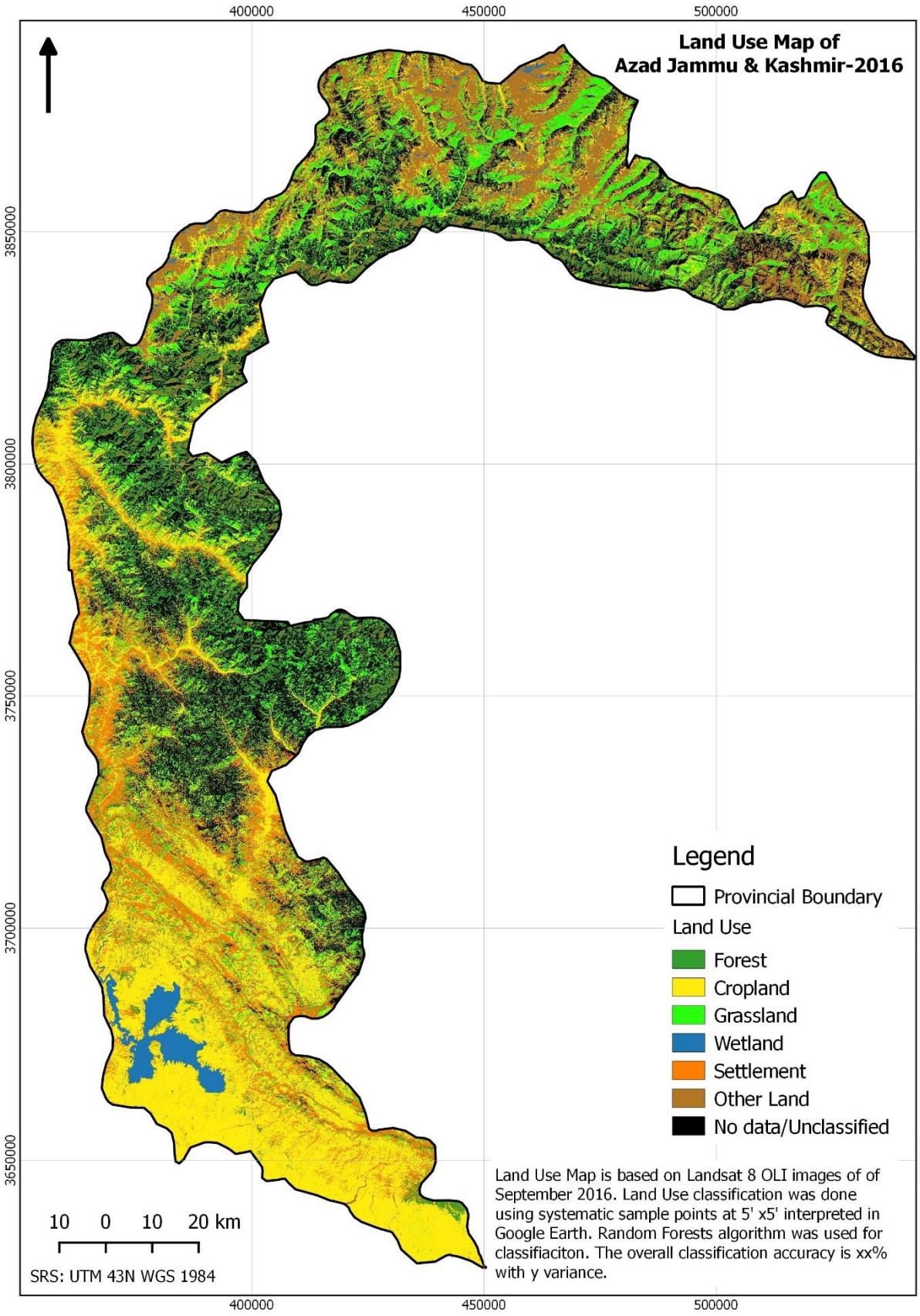




3. From the layer properties of classified land use raster layer, render the pixels with appropriate colour as shown.

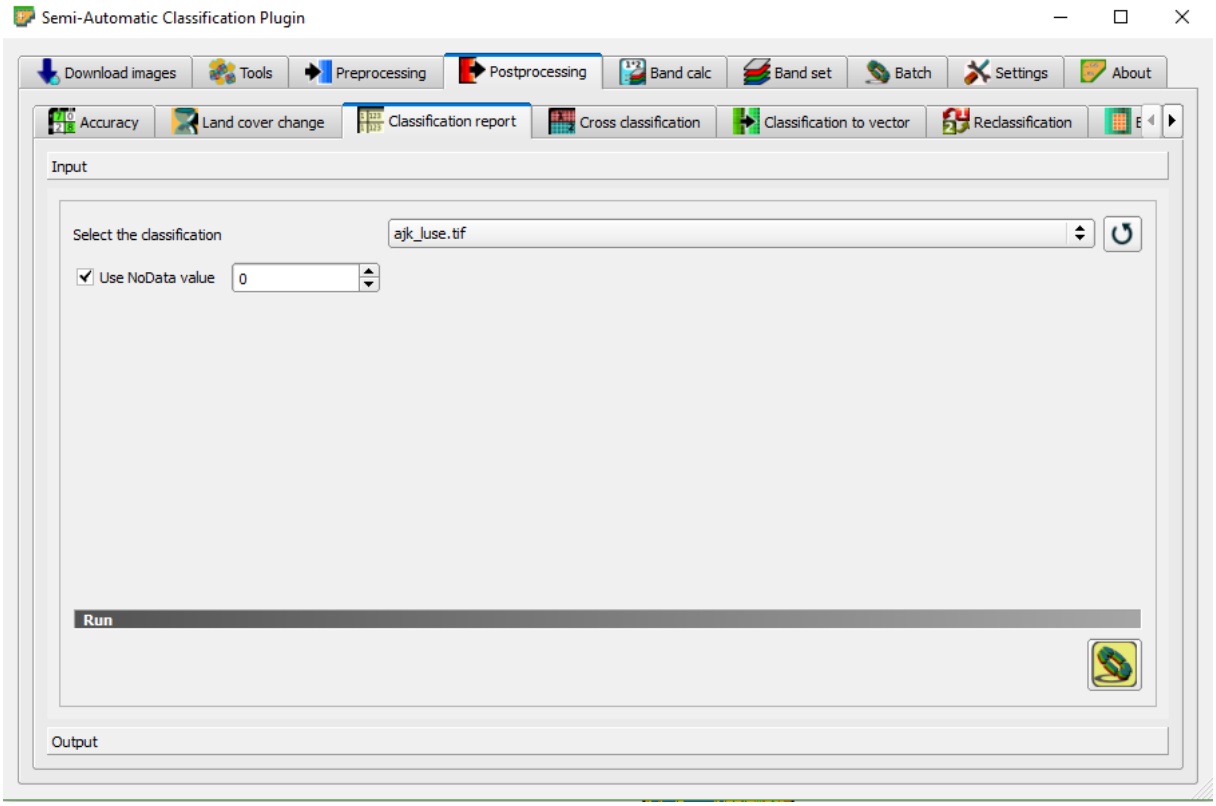


4. Compose the land use map appropriately using QGIS map composer tool.


Land Use Map of
Azad Jammu & Kashmir-2016
400000
450000
500000
3850000
3800000
3750000
3700000
3650000
Legend
Provincial Boundary
Land Use
Forest
Cropland
Grassland
Wetland
Settlement
Other Land
No data/Unclassified
Land Use Map is based on Landsat 8 OLI images of of September 2016. Land Use classification was done using systematic sample points at 5' x5' interpreted in Google Earth. Random Forests algorithm was used for classifiaciton. The overall classification accuracy is xx% with y variance.
10 0 10 20 km
SRS: UTM 43N WGS 1984

Step 5 Calculating Land Use Raster Statistics

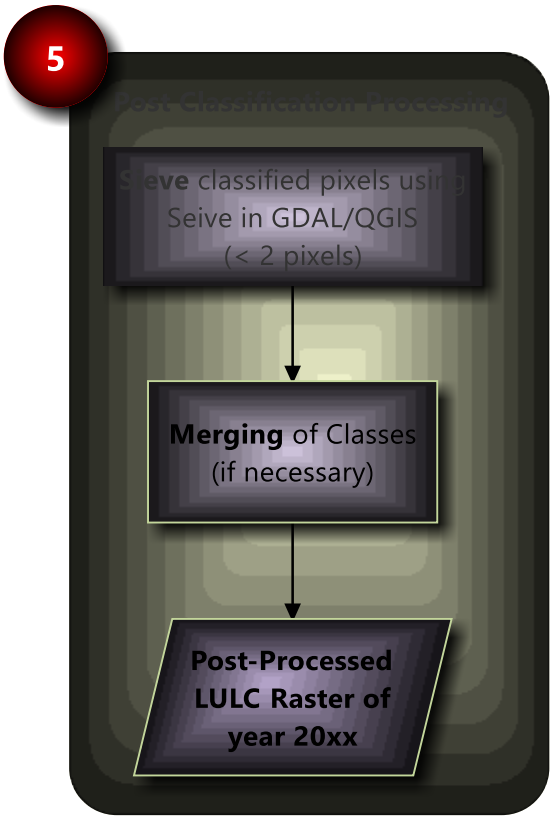
1. In QGIS **Semi-Automatic Classification Plugin** tool and go to *Postprocessing>Classification report tab*



2. Select the classified and clipped raster *'ajk_luse.tif' in* **Select the classification**
3. *Run* the process and save the file as *'ajk_luse_area.csv'* text file.
4. The statistics of the classified land use raster is extracted for Class (0-7), PixelSum (total pixels), Percentage%, Area in m^2 (no. of pixels x 30 x 30) as shown.

Class	PixelSum	Percentage %	Area [metre^2]
1.0	1535127	9.94740558134	1381647878.86
2.0	3940740	25.5354371792	3546752198.45
3.0	2279402	14.7702021897	2051511658.89
4.0	280542	1.8178724344	252493936.483
5.0	1971933	12.7778466083	1774782833.41
6.0	2205148	14.2890467843	1984681434.68
7.0	3219544	20.8621892228	2897660023.24

6. Post Classification Processing



Post classification processing is done to remove noises such as salt and pepper" effects of individual classified pixels. This is done by "sieving" isolated pixels and replacing them with the classification of surrounding majority class pixels.

For the case of LULC of Pakistan, the minimum mapping unit is considered as 2 pixels (i.e. 60 x 60 m). Threshold for sieving needs to be defined accordingly.

6.1. Exercise 5 Post Classification Processing

6.1.1. Background

Post classification processing is done to remove noises such as salt and pepper" effects of individual classified pixels. These classified pixels need to be removed if you are confident that these happen to be due to classification errors. This is often done by "sieving" isolated pixels and replacing them with the classification of surrounding majority class pixels.

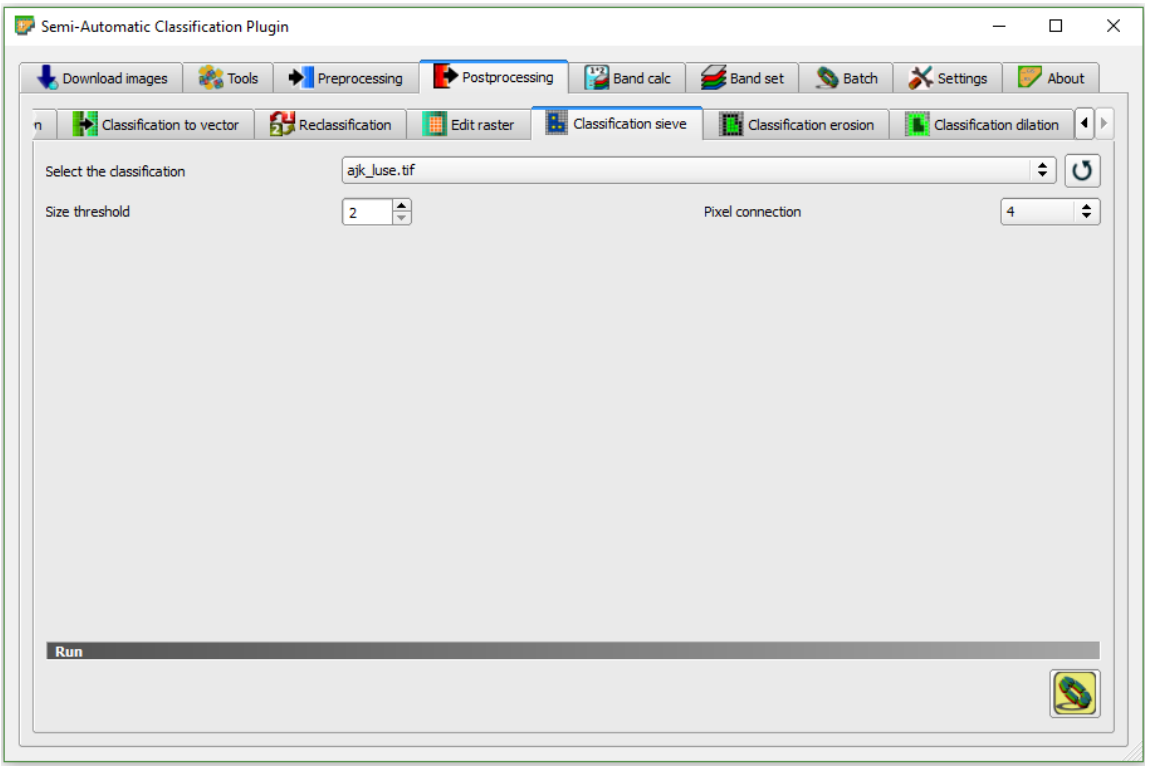
Data Used

SN	Datasets	Sources	Descriptions
1	Classified AJK Land Use/Land Cover raster *(ajk_classification.tif)*	Classified LULC raster data from Exercise 4	

Software Used

- Quantum GIS (QGIS) Ver 2.18 (or older) - Open source GIS software

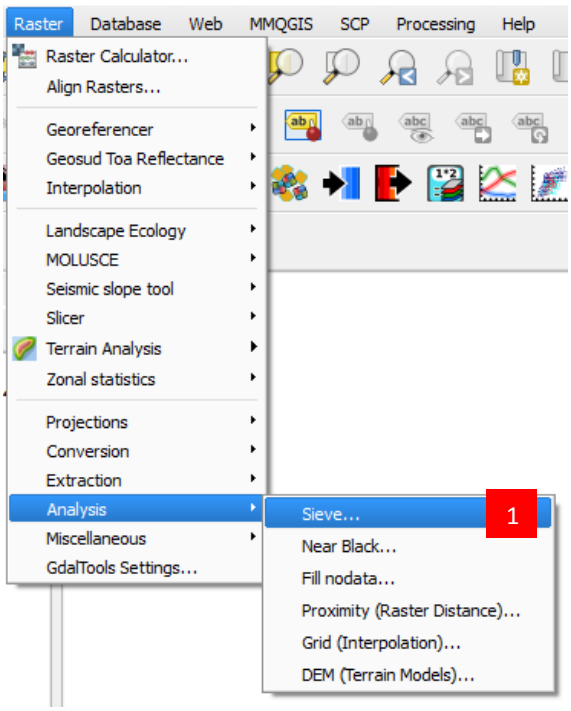
[Note: This step can be done using Semi-Automatic Classification (SCP) Postprocessing>Classification Sieve in conjunction with other tools Classification Erosion and Classification Dilation as well. In this exercise, simple raster function Sieve is used]



6.1.2. Classification Post Processing

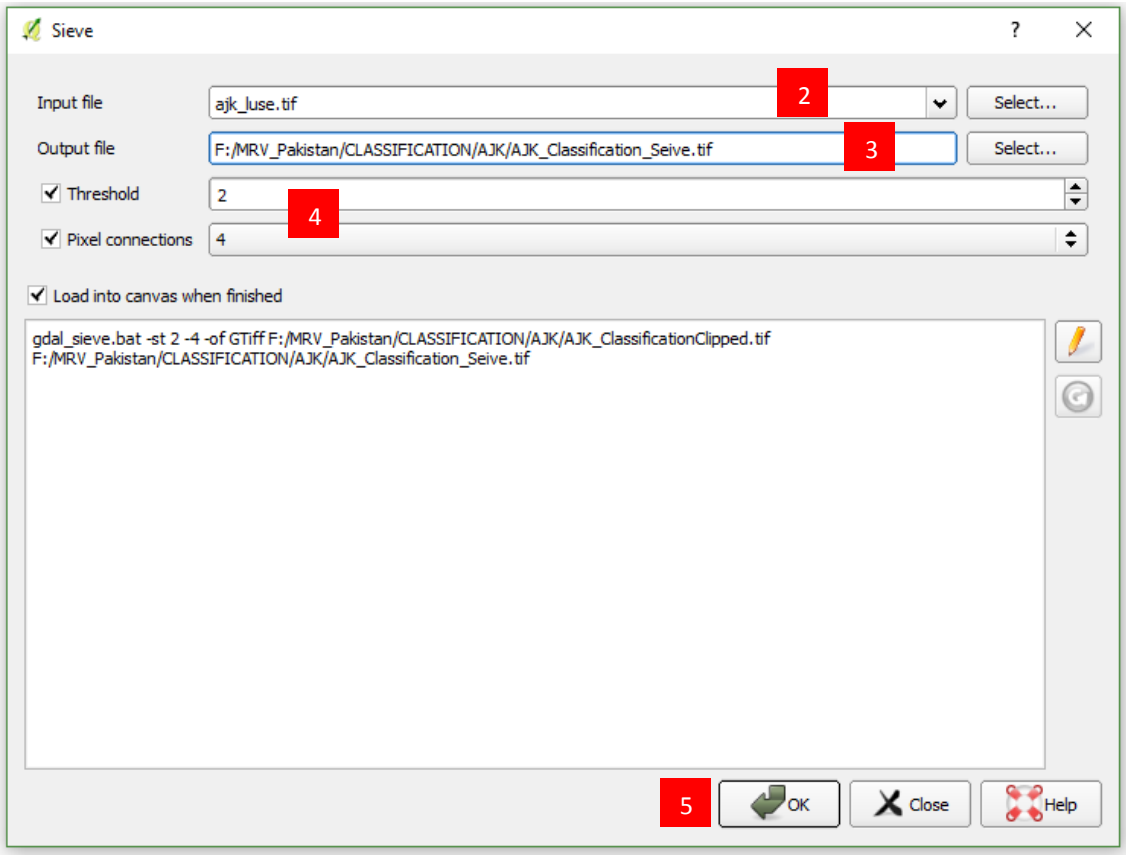
Step 1 Sieve

5. From QGIS *Raster* menu select *Analysis>Sieve..*
6. In the *Sieve* window, select the classified raster '*ajk_classification.tif*'.
7. Browse to the output location and name the output as '*ajk_classifiaction_seive.tif*'.
8. Define the ***Threshold*** as ***2*** and ***Pixel Connection as 4***
9. Run the algorithm



Caution!

If the pixels are 'unclassified', as in the case in above example, be cautious on using larger thresholds.



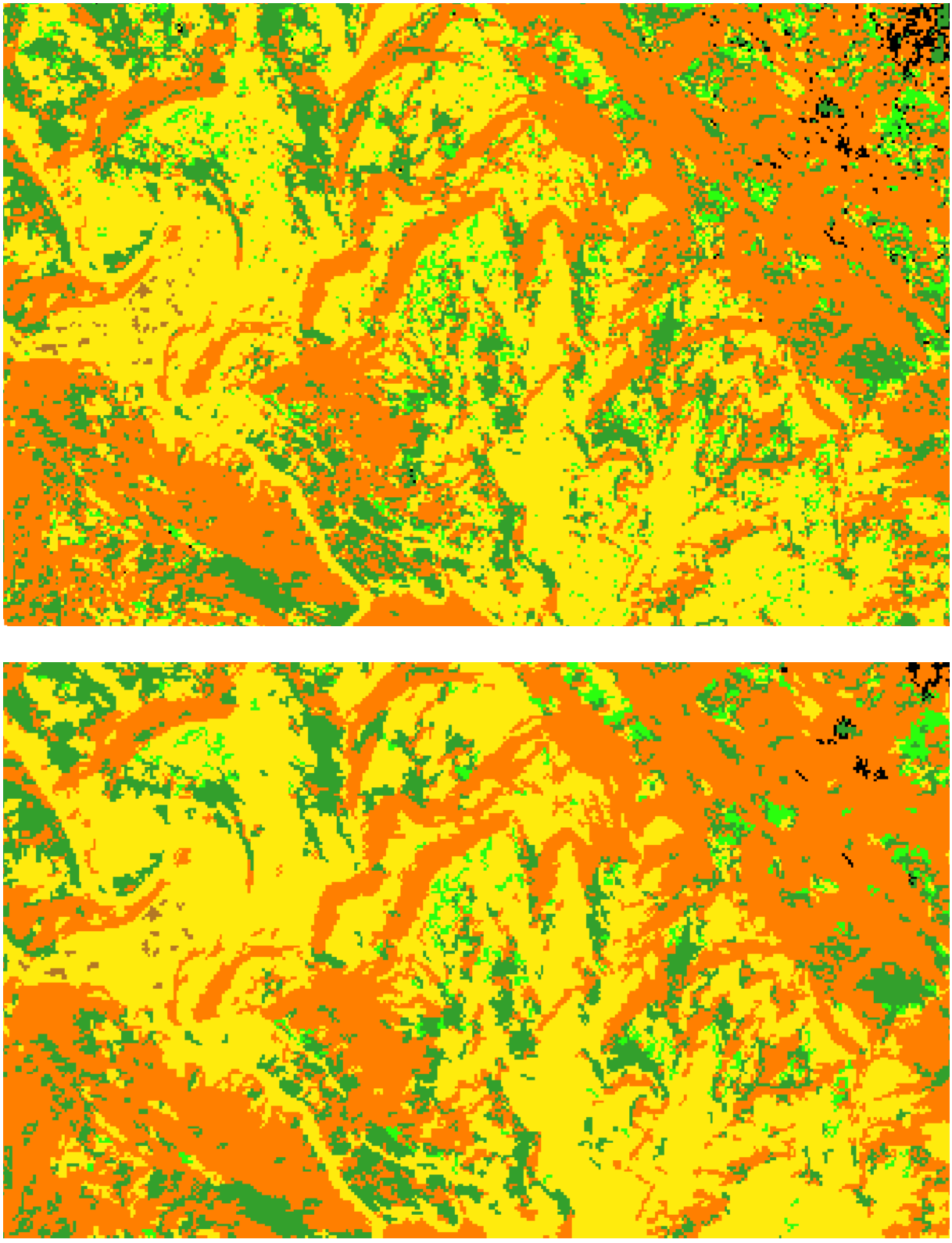
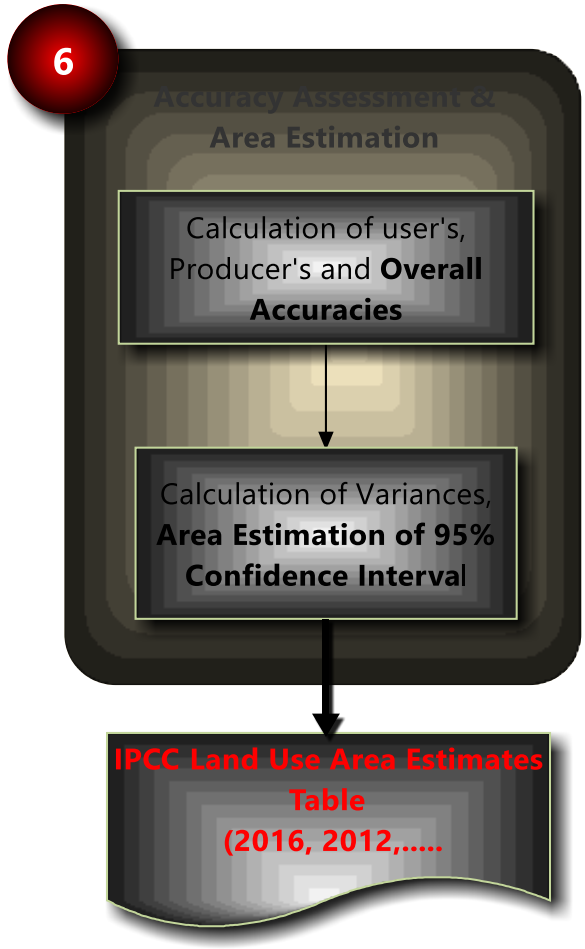


Figure 12 Classified raster before and after running Sieve algorithm

6.1.3. Assignment

1. *Re-run the Sieve process using different threshold of 4 and Pixel Connections of 8.*
2. *Compare the result with the previous attempt*

7. Accuracy Assessment and Area Estimation



Error Matrix is most commonly used to assess the overall and class wise accuracy of classified map products.

Area estimations along with standard errors is calculated at 95% confidence interval.

(Olofsson et al., 2014) method is adopted for accuracy assessment and area estimation.

In this exercise, a simplistic approach is used to demonstrate the method by using the reference data which were visually interpreted and selected.

In the actual case, independent samples sets are designed (stratified random) and used to assess the accuracy and estimate

Detailed methodology is given in (FAO, 2016):

FAO. 2016. "Map Accuracy Assessment and Area Estimation: A Practical Guide." In *National forest monitoring assessment working paper*. Rome: Food and Agriculture Organization of the United Nations.

7.1. Exercise 6 Accuracy Assessment and Area Estimation

7.1.1. Background

Data Used

SN	Datasets	Sources	Descriptions
1	AJK validation sample set *(AJK_ValidationRandomplots.shp)*	Independent validation sample subset	Independent sample subset generated for validation and area estimation
2	Classified land use raster data *(ajk_luse.tif)*	Classified land use raster from Exercise 4	
3	MS Excel/Libre Office worksheet *(AJK_AreaEstimation.xlsx/AJK_AreaEstimation.ods)*		Excel worksheet with estimation, error matrix and area estimation formula

Software Used

- Quantum GIS (QGIS) Ver 2.18 (or older) - Open source GIS software
- MS Excel or LibreOffice Calc (AJK_AreaEstimation.xlsx/AJK_AreaEstimation.ods)
- Python Script 'sample_map.py'

Caution!

To complete Chapter 7.1.2, make sure you have Python environment installed in your system (https://www.python.org/).

7.1.2. Exercise 6A Designing Reference Sample (Optional)[1]

A stratified random sampling is used for designing the reference sample. In this this exercise, independent stratified random samples subset is used to assess the accuracy of the land use classification.

Step 1 Determining Sample Size and Allocation

Sample size is determined using Cochran equation $n \approx \left(\frac{\sum W_i S_i}{S(\hat{P})}\right)^2$, where

W_i is the stratum weight

S_i is the standard error for stratum i given by $S_i = \sqrt{p_i(1-p_i)}$, p_i is the expected User's Accuracy (i.e. 0.8 for all the classes in this exercise)

$S(\hat{P})$ is the target standard error of overall accuracy (i.e. 0.010 in this exercise)

The sample size can be computed using the above equations in Excel as shown:

Table 1 Map are proportion and assumed User's Accuracy

LU Strata	Area (ha^2)	W_i	p_i	S_i	W_i*S_i
Forest	138,164.79	0.126	0.80	0.400	0.0503
Cropland	354,675.22	0.323	0.80	0.400	0.1291
Grassland	205,151.17	0.187	0.80	0.400	0.0747
Wetland	25,249.39	0.023	0.80	0.400	0.0092
Settlement	177,478.28	0.161	0.80	0.400	0.0646
Other Land	198,468.14	0.181	0.80	0.400	0.0722
No data*	289,766.00				
	1,099,186.99*				

** No data/Unclassified value is excluded from the calculation. No data value was due to classification error in terrain shadow regions.*

For $S(\hat{P}) = 0.010$

n = 1600 (i.e. 1600 random samples for 80% per class accuracy and 0.010 standard error of overall accuracy of 80%)

To allocate 1600 random samples amongst the land use strata, equal number of samples or proportionate samples can be used. Good practices stipulate that 50, 75 or 100 units are allocated to the smaller classes depending on the total sample size and that the rest is proportionally allocated to the larger strata (Congalton & Green, 2009). Example of sample allocation is shown in the following table.

[1] Trainees could choose to attempt this exercise (7.1.2 Exercise 6A Designing Reference Sample) to understand the principle of stratified sampling for validation, if the time permits. However, a simplified Exercise 6B is given to attempt.

[2] Pixel count could be used for area as well

Assignment

3. Review the accompanying Excel/Libre Office Calc worksheet 'SampleDesign' in *'AJK_AreaEstimation.xlsx or AJK_AreaEstimation.ods'* file

Table 2 Allocation of reference samples per strata

LU Strata	Equal	Proportional	Scheme1	Scheme2	Scheme3
Forest	267	201	100	75	50
Cropland	267	516	100	75	50
Grassland	267	299	100	75	50
Wetland	267	37	82	87	91
Settlement	267	258	575	608	641
Other Land	267	289	643	680	717
	1,600	1,600	1,600	1,600	1,600

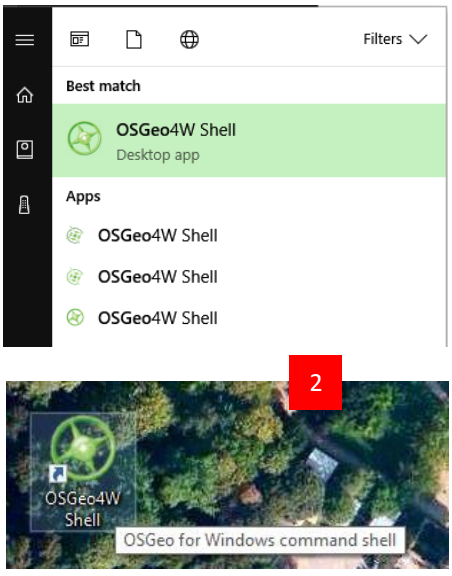
Step 2 Generating Validation Stratified Random Samples

There is no inbuilt tool in QGIS to generate stratified random samples based on a defined sample size and its allocations, therefore a customized Python Script (sample_map.py1) is used. Alternatively, an interactive web based tool developed using R, RStudio and the Shiny package, following (Olofsson et al., 2014) can be used from https://github.com/openforis/accuracy-assessment. In this exercise, a Python Script is used to generate stratified random samples based on the above sample size (n=1600) and allocation in

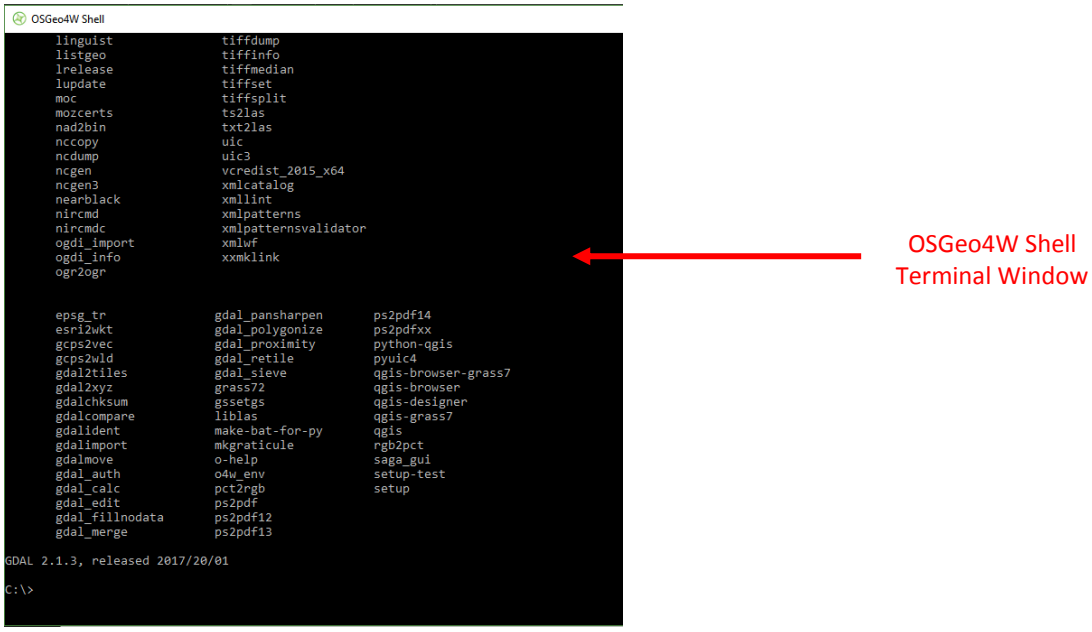
Table 2.

Before generating the random samples, copy the Python Scripts ***sample_map.py*** and ***docopt.py*** to your working folder i.e. folder containing the '*ajk_luse.tif*' classified land use raster file in in this tutorial "..// Ex4_ImageClassification/Output/". *[Note: Make sure Python Programming environment is installed in your system]*

4. Goto **OSGeo4W Shell** from *Windows>OSGeo4 Shell* or from Desktop Icon to open OSGeo4W Terminal Window



[1] https://github.com/ceholden/accuracy_sampler/tree/master/script. Minor modification is done in the code for exception handling.



5. Using MS-DOS commands, navigate to your working directory containing the '*ajk_luse.tif*' file, in this tutorial "..// Ex4_ImageClassification/Output/".

```
D:\CD Module2_SLMS
D:\Module2_SLMS\CD Ex4_ImageClassification
D:\Module2_SLMS\Ex4_ImageClassification>CD Output
D:\Module2_SLMS\Ex4_ImageClassification>Output
```

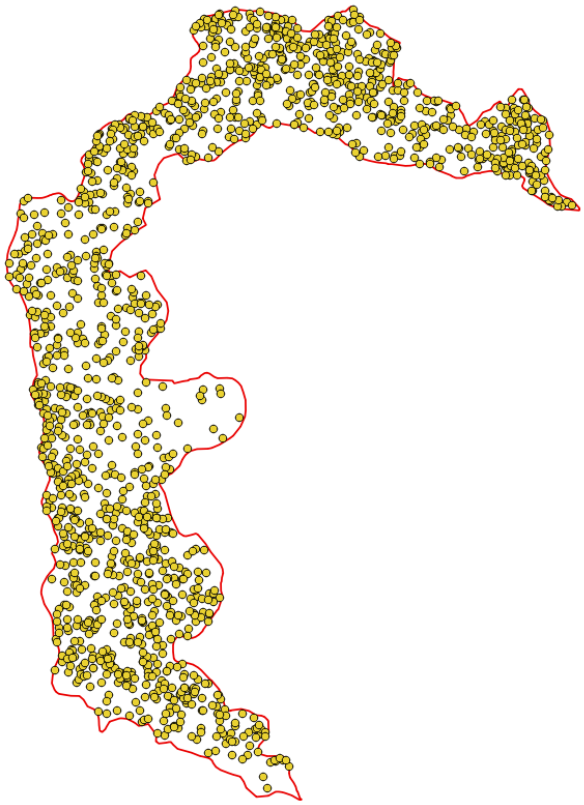
6. In the working folder path type python script code
   python sample_map.py -v --size 1600 --allocation "100 100 100 82 575 643" --mask "0 7" --ndv 255 --vector stratifiedsample.shp stratified AJK_ClassificationClipped.tif

```
D:\Module2_SLMS\Ex4_ImageClassification>Output>python sample_map.py -v --size 1600 --allocation "100 100 100 82 575 643" --mask "0 7" --ndv 255 --vector stratifiedsample.shp stratified AJK_ClassificationClipped.tif
```

This script defines sample size of 1600 and allocation of samples for Scheme 1 (in Table 2) with sample size of [100 100 100 82 575 643] for Forest, Cropland, Grassland, Wetland, Settlements and Other Land respectively. Mask [0 7] defines masking or exclusion of 0 pixels and unclassified pixel defined by value 7. A vector polygon shapefile *'stratifiedsample.shp'* representing sampled 1600 individual pixels is generated.

7. Centroid points of these stratified random sample polygons are generated on QGIS using Vector*>Geometry Tools>polygon Centroids.*
8. For interpretation of these stratified random sample points, OpenForis Collect Earth Process defined in **Exercise 2B Interpretation of LULC Classes using OpenForis CollectEarth** can be followed. Ground truthing is done is the selected sample subset from these samples.

5



Caution!

To complete Chapter 7.1.2, make sure you have Python environment installed in your system (https://www.python.org/).

OSGeo4W Shell

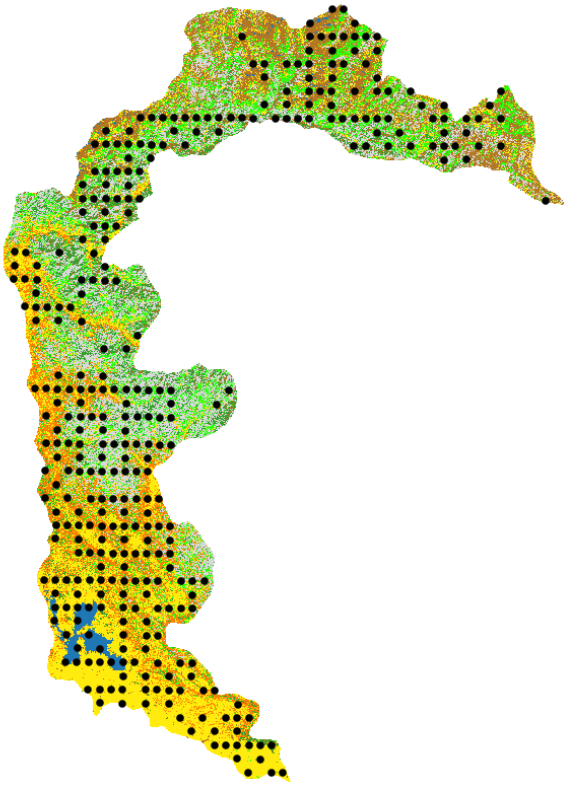
```
23:19:51 DEBUG:     collected samples
23:19:51 DEBUG: Randomizing order of samples
23:19:51 DEBUG: Finished collecting samples
23:19:51 DEBUG: Writing raster output to sample.tif
23:19:52 DEBUG: Writing vector output to stratifiedsample.shp
23:19:52 DEBUG: Sampling complete

F:\MRV_Pakistan\CLASSIFICATION\AJK>python sample_map.py -v --size 1600 --allocation "100 100 100 82 575 643" --mask "0 7
" --ndv 255 --vector stratifiedsample2.shp stratified AJK_ClassificationClipped.tif
00:26:03 DEBUG: Using map image AJK_ClassificationClipped.tif
00:26:03 DEBUG: Sampling method is stratified
00:26:03 DEBUG: Sample size is 1600
00:26:03 DEBUG: Allocation is [100 100 100  82 575 643]
00:26:03 DEBUG: Mask values are [0 7]
00:26:03 DEBUG: Writing output raster to sample.gtif (GTiff)
00:26:03 DEBUG: Writing output vector to stratifiedsample2.shp (ESRI Shapefile)
00:26:03 DEBUG: Read in map image to be sampled
00:26:04 DEBUG: Found 6 classes
00:26:04 DEBUG:     class 1.0 - 1535127px (2.76%)
00:26:05 DEBUG:     class 2.0 - 3940740px (7.09%)
00:26:05 DEBUG:     class 3.0 - 2279402px (4.1%)
00:26:05 DEBUG:     class 4.0 - 280542px (0.5%)
00:26:05 DEBUG:     class 5.0 - 1971933px (3.55%)
00:26:05 DEBUG:     class 6.0 - 2205148px (3.96%)
00:26:05 DEBUG: Performing sampling
00:26:05 DEBUG: Sampling class 1.0
00:26:05 DEBUG:     collected samples
00:26:05 DEBUG: Sampling class 2.0
00:26:06 DEBUG:     collected samples
00:26:06 DEBUG: Sampling class 3.0
00:26:06 DEBUG:     collected samples
00:26:06 DEBUG: Sampling class 4.0
00:26:06 DEBUG:     collected samples
00:26:06 DEBUG: Sampling class 5.0
00:26:06 DEBUG:     collected samples
00:26:06 DEBUG: Sampling class 6.0
00:26:07 DEBUG:     collected samples
00:26:07 DEBUG: Randomizing order of samples
00:26:07 DEBUG: Finished collecting samples
00:26:07 DEBUG: Writing raster output to sample.gtif
00:26:07 DEBUG: Writing vector output to stratifiedsample2.shp
00:26:07 DEBUG: Sampling complete

F:\MRV_Pakistan\CLASSIFICATION\AJK>
```

7.1.3. Exercise 6B Construction Error Matrix

The designed and allocated reference samples in the previous step is NOT used for this exercise due to time constraints. To facilitate this exercise, a separate set random of samples have been generated (with filename *'AJK_ValidationRandomPlots.shp')* and visually interpreted in Collect Earth. These were generated at 2.5'systematic grids (n=434) (excluding the ones used for classification and ones with No Data) and then randomly selected considering 80% intensity (i.e. n= 347) as shown here.



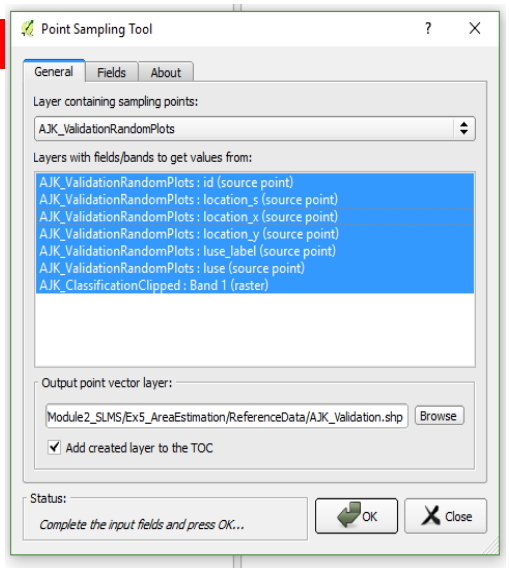
Step 1 Extracting Mapped Land Use Classes to Stratified Random Sample Points

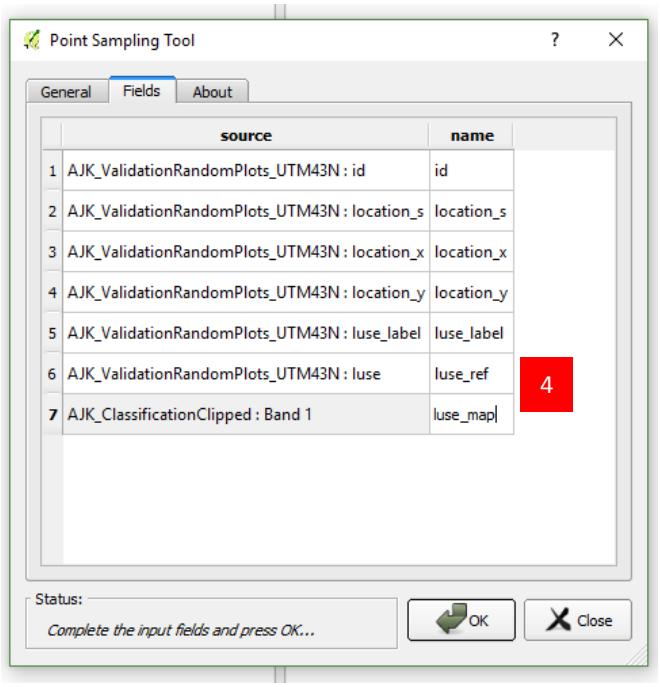
1. Add classified raster layer *'ajk_luse.tif'* and validation random samples point layer *'AJK_ValidationRandomPlots.shp'* in QGIS.
2. Using *Plugins>Analysis>Point sampling tool* extract mapped land use raster values to the validation random points.

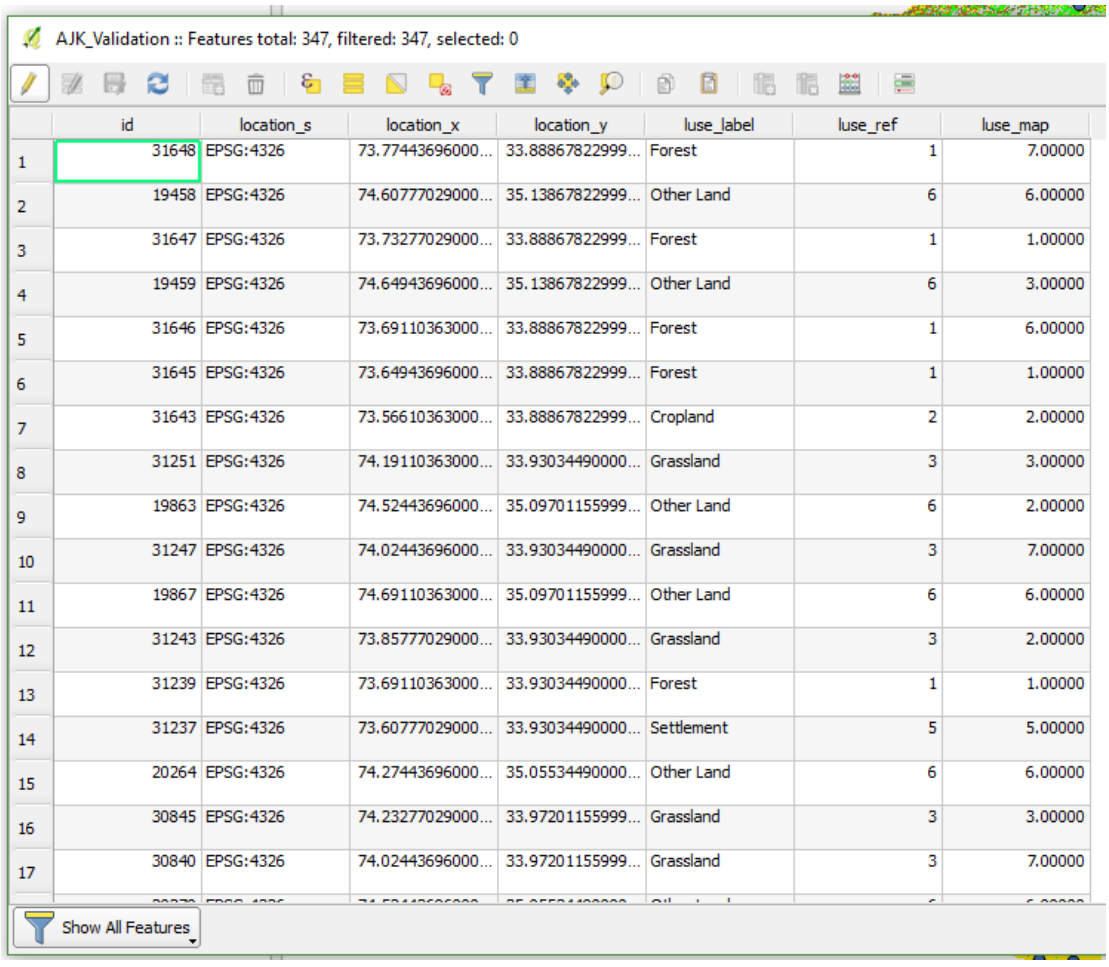
[Note: Make sure the SRS of plots and the image raster is the same i.e. UTM 43N in this exercise]

3. Select all the attributes of AJK_validationRandomPlots and Band 1 (raster) from 'ajk_luse' to create a point file named 'AJK_Validation.shp'.
4. You can choose to rename the fields of newly generated point file, naming 'luse_ref' field for reference land use interpreted using Collect Earth and 'luse_map' for extracted land use classes from the classified land use raster 'ajk_luse.tif' as shown.

3







	id	location_s	location_x	location_y	luse_label	luse_ref	luse_map
1	31648	EPSG:4326	73.77443696000...	33.88867822999...	Forest	1	7.00000
2	19458	EPSG:4326	74.60777029000...	35.13867822999...	Other Land	6	6.00000
3	31647	EPSG:4326	73.73277029000...	33.88867822999...	Forest	1	1.00000
4	19459	EPSG:4326	74.64943696000...	35.13867822999...	Other Land	6	3.00000
5	31646	EPSG:4326	73.69110363000...	33.88867822999...	Forest	1	6.00000
6	31645	EPSG:4326	73.64943696000...	33.88867822999...	Forest	1	1.00000
7	31643	EPSG:4326	73.56610363000...	33.88867822999...	Cropland	2	2.00000
8	31251	EPSG:4326	74.19110363000...	33.93034490000...	Grassland	3	3.00000
9	19863	EPSG:4326	74.52443696000...	35.09701155999...	Other Land	6	2.00000
10	31247	EPSG:4326	74.02443696000...	33.93034490000...	Grassland	3	7.00000
11	19867	EPSG:4326	74.69110363000...	35.09701155999...	Other Land	6	6.00000
12	31243	EPSG:4326	73.85777029000...	33.93034490000...	Grassland	3	2.00000
13	31239	EPSG:4326	73.69110363000...	33.93034490000...	Forest	1	1.00000
14	31237	EPSG:4326	73.60777029000...	33.93034490000...	Settlement	5	5.00000
15	20264	EPSG:4326	74.27443696000...	35.05534490000...	Other Land	6	6.00000
16	30845	EPSG:4326	74.23277029000...	33.97201155999...	Grassland	3	3.00000
17	30840	EPSG:4326	74.02443696000...	33.97201155999...	Grassland	3	7.00000

Step 2 Construction of Error Matrix

Error matrix is constructed by **Cross-Tabulating** mapped samples Vs reference samples in a tabular

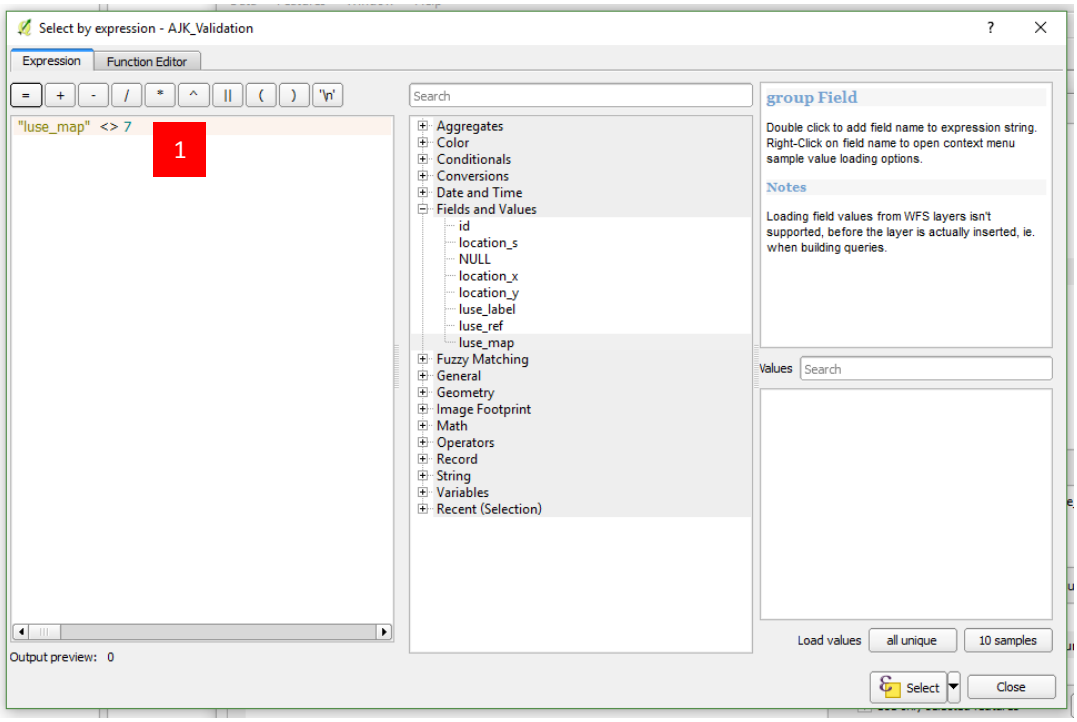
format as shown in Table 3 (adopted from (Olofsson et al., 2014)).

Table 3 Population error matrix of four classes with cell entries (p_{ij})

		Reference				
		Class 1	Class 2	Class 3	Class 4	Total
Map	Class 1	p_{11}	p_{12}	p_{13}	p_{14}	$p_{1.}$
	Class 2	P_{21}	P_{22}	P_{23}	P_{24}	$p_{2.}$
	Class 3	p_{31}	p_{31}	p_{31}	p_{31}	$p_{3.}$
	Class 4	p_{41}	p_{41}	p_{41}	p_{41}	$p_{4.}$
	Total	$p_{.1}$	$p_{.2}$	$p_{.3}$	$p_{.4}$	1

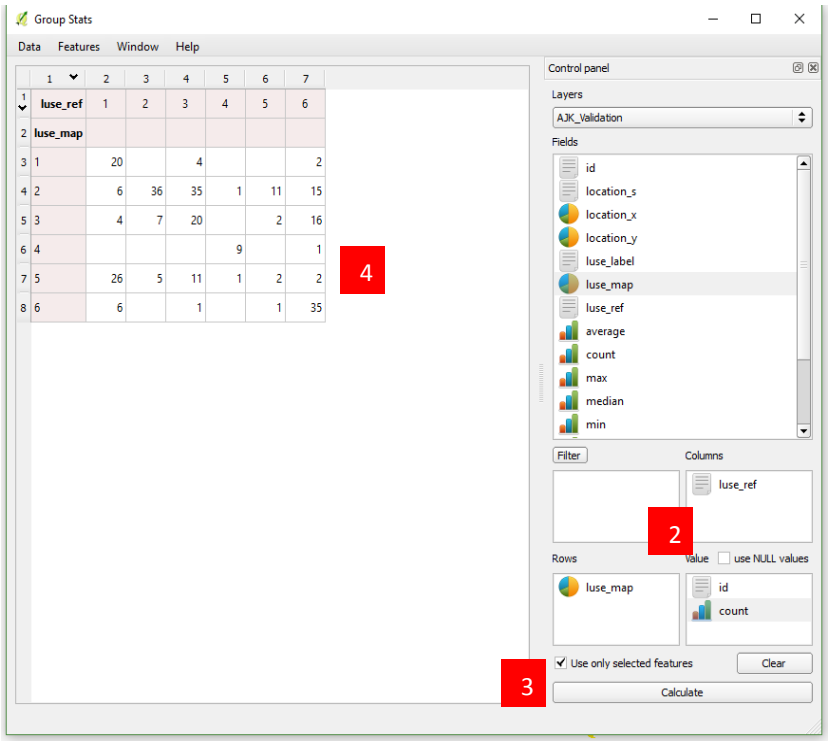
To generate the error matrix in QGIS, follow these steps.

1. In QGIS, open the attribute table of AJK_Validation point layer and using *Select Features using an expression,* filter out all the points with No Data in *luse_map field* i.e. select only those with land use classes 1 - 6. Out of 347, 279 plots are selected, excluding those 68 plots with 'luse_map'=7 (No data)

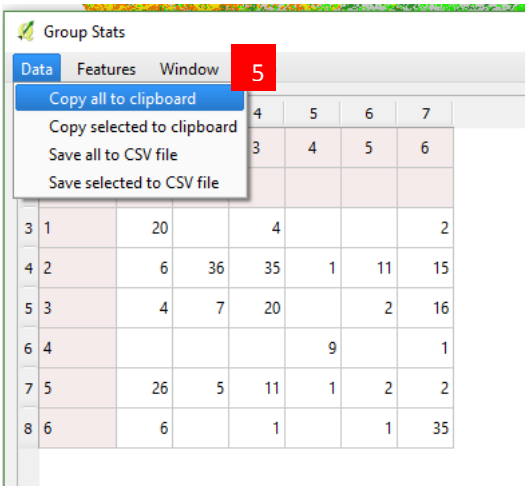


2. In QGIS, use Group Stats Plugin in *Vector>Group Stats*. In the *Columns* drag and drop *luse_ref* field, in the *Rows* drag and drop *luse_map* field, in the *Value* drag and drop *id* field and *Count* parameter to generate cross-tabulation of the fields.
3. Select *Use only selected features* option to tabulate those selected 279 plots (excluding luse_map=7 values). Calculate the cross-tab.

4. Cross-tabulated matrix is shown.



5. Using *Data>Copy all to clipboard* menu, the matrix can be copied to MS Excel/Libre Office Calc worksheet.
6. You can choose to rename the columns and rows appropriately with land use classes instead of 1-6 raster integer values. Calculate the totals of Rows and Columns in the worksheet.
7. Copy the matrix to the *'AJK_AreaEstimation.xlsx' Area Estimation* worksheet for further calculations.



6

	Forest	Cropland	Grassland	Wetland	Settlemer	Other Land	Total
Forest	20		4			2	26
Cropland	6	36	35	1	11	15	104
Grassland	4	7	20		2	16	49
Wetland				9		1	10
Settlement	26	5	11	1	2	2	47
Other Land	6		1		1	35	43
Total	62	48	71	11	16	71	279

7.1.4. Calculating User's, Producer's, Overall Accuracies, Variances and Confidence Interval

Error matrix of sample counts for randomly selected stratified validation points n_{ij} are presented in Table 4. The total map samples in a Class i is represented by $n_{i.}$ The proportion of mapped area $A_{m,i}$ for Class i is represented by W_i, where

$$W_i = \frac{A_{m,i}}{\sum A_{m,i}}$$

Table 4 Error matrix of sample counts for randomly stratified validation points for LUC 2005

LULC	Forest	Cropland	Grassland	Wetland	Settlement	Other Land	$n_{i.}$	$A_{m,i}$ (ha)	W_i
Forest	20		4			2	26	138,164.79	0.12570
Cropland	6	36	35	1	11	15	104	354,675.22	0.32267
Grassland	4	7	20		2	16	49	205,151.17	0.18664
Wetland				9		1	10	25,249.39	0.02297
Settlement	26	5	11	1	2	2	47	177,478.28	0.16146
Other Land	6		1		1	35	43	198,468.14	0.18056
Total	62	48	71	11	16	71	279	1,099,186.99	

The sample based estimator for area proportion p_{ij} is given by $\hat{p}_{ij} = W_i \frac{n_{ij}}{n_{i.}}$, where i denotes the row and j denotes the column in the error matrix. For stratified random sampling designs, where the strata correspond to the map classes, error matrix may be represented in terms of estimated area proportions $\hat{p}_{ij}$ as in Table 2.

User's accuracy of class i is $\hat{U}_i = \frac{\hat{p}_{ii}}{\hat{p}_{i.}}$, and its complimentary measure commission error of class i is $1 - \frac{\hat{p}_{ii}}{\hat{p}_{i.}}$

Producer's accuracy of class j is $\hat{P}_j = \frac{\hat{p}_{jj}}{\hat{p}_{j.}}$, and its complimentary measure omission error of class j is $1 - \frac{\hat{p}_{jj}}{\hat{p}_{j.}}$

LUC	Forest	Cropland	Grassland	Wetland	Settlement	Other Land	$A_{m,i}$ (ha)	W_i	$\hat{\mathbf{U}}_\mathbf{i}$	$1-\hat{\mathbf{U}}_\mathbf{i}$	$\hat{\mathbf{V}}_\mathbf{i}$	$\hat{\mathbf{V}}\left(\hat{\mathbf{U}}_i\right)$	95% Confidence Interval
Forest	0.0967	0.0000	0.0193	0.0000	0.0000	0.0097	138,164.8	0.12570	**0.7692**	**0.2308**	0.000112	0.007101	±0.165
Cropland	0.0186	0.1117	0.1086	0.0031	0.0341	0.0465	354,675.2	0.32267	**0.3462**	**0.6538**	0.000229	0.002197	±0.092
Grassland	0.0152	0.0267	0.0762	0.0000	0.0076	0.0609	205,151.2	0.18664	**0.4082**	**0.5918**	0.000175	0.005033	±0.139
Wetland	0.0000	0.0000	0.0000	0.0207	0.0000	0.0023	25,249.4	0.02297	**0.9000**	**0.1000**	0.000005	0.010000	±0.196
Settlement	0.0893	0.0172	0.0378	0.0034	0.0069	0.0069	177,478.3	0.16146	**0.0426**	**0.9574**	0.000023	0.000886	±0.058
Other Land	0.0252	0.0000	0.0042	0.0000	0.0042	0.1470	198,468.1	0.18056	**0.8140**	**0.1860**	0.000118	0.003606	±0.118
Total	**0.2451**	**0.1555**	**0.2461**	**0.0272**	**0.0528**	**0.2733**	**1,099,187.0**	**1.00000**			0.000662		
$\hat{\mathbf{P}}_\mathbf{j}$	**0.39**	**0.72**	**0.31**	**0.76**	**0.13**	**0.54**							
$\mathbf{1}-\hat{\mathbf{P}}_\mathbf{j}$	**0.61**	**0.28**	**0.69**	**0.24**	**0.87**	**0.46**							

The Overall Accuracy is given by $\hat{\mathrm{O}} = \sum_{j=1}^{q} \hat{p}_{jj}$ and its estimated variance is given by $\hat{\mathrm{V}}\left(\hat{\mathrm{O}}\right) = \frac{\sum_{i=1}^{q} W_i^2\, \hat{\mathrm{U}}_i(1-\hat{\mathrm{U}}_i)}{n_i - 1}$

The estimated variance for User's Accuracy by $\hat{\mathrm{V}}\left(\hat{\mathrm{U}}_i\right) = \frac{\hat{\mathrm{U}}_i(1-\hat{\mathrm{U}}_i)}{n_i - 1}$.

The 95% confidence intervals are estimated as $\pm 1.96\sqrt{\hat{\mathrm{V}}\left(\hat{\mathrm{O}}_i\right)}$ for Overall Accuracy and $\pm 1.96\sqrt{\hat{\mathrm{V}}\left(\hat{\mathrm{U}}_i\right)}$ for Users' Accuracy.

For the estimated area proportions,

Overall accuracy $\hat{\mathbf{O}} = \mathbf{0.46}$ and its variance $\hat{\mathbf{V}}\left(\hat{\mathbf{O}}\right) = \mathbf{0.000662}$ and 95% confidence intervals is **0.46 ±0.050. Which is very low due to the sampling used in this exercise. Also, the accuracy of the classification could be improved by removing the shadow areas/minimizing the effect of terrain shadow by topographical corrections. However, for satisfactory result the sample designed and allocated in Exercise 7A is recommended. Nonetheless, this calculation exercise demonstrates the process for accuracy assessment and area estimation.**

7.1.5. Calculating Standard Error

The estimate of confidence interval is given by Standard Error $\mathrm{SE}\left(\hat{\mathrm{p}}_{.j}\right) = \sqrt{\sum_i \frac{W_i\ \hat{p}_{ij} - \hat{p}_{ij}{}^2}{n_{i.} - 1}}$ of the stratified estimator of estimated proportion area. The

$\mathrm{SE}\ (\hat{\mathrm{p}}_{.j})$ calculations are presented in the Table 5.

Table 5 Estimate of confidence interval by Standard Error

LUC	Forest	Cropland	Grassland	Wetland	Settlement	Other Land
Forest	0.000112	0.000000	0.000082	0.000000	0.000000	0.000045
Cropland	0.000055	0.000229	0.000226	0.000010	0.000096	0.000125
Grassland	0.000054	0.000089	0.000175	0.000000	0.000028	0.000160
Wetland	0.000000	0.000000	0.000000	0.000005	0.000000	0.000005
Settlement	0.000140	0.000054	0.000102	0.000012	0.000023	0.000023
Other Land	0.000093	0.000000	0.000018	0.000000	0.000018	0.000118
Total	**0.021327**	**0.019275**	**0.024546**	**0.005168**	**0.012835**	**0.021798**

7.1.6. Calculating Area Estimation and Confidence Interval

Area estimates and confidence intervals are calculated as shown in the Table4.

LUC	Proportion Area		Estimated Area(ha)			
	$\hat{\mathbf{p}}_{.j}$	$\mathbf{SE}(\hat{\mathbf{p}}_{.j})$	$\hat{\mathbf{A}}_{.j} = \hat{\mathbf{p}}_{.j} \times A_{tot}$	$\mathbf{SE}(\hat{\mathbf{A}}_{.j}) = \mathbf{SE}(\hat{\mathbf{p}}_{.j}) \times A_{tot}$	95% Confidence interval	Confidence Interval
Forest	0.24506	0.02133	269,362.00	23442.00	45,946	269362±45946
Cropland	0.15553	0.01927	170,960.00	21186.00	41,524	170960±41524
Grassland	0.24610	0.02455	270,506.00	26980.00	52,880	270506±52880
Wetland	0.02721	0.00517	29,910.00	5680.00	11,132	29910±11132
Settlement	0.05282	0.01284	58,055.00	14108.00	27,651	58055±27651
Other Land	0.27329	0.02180	300,392.00	23959.00	46,959	300392±46959

7.1.7. Summary Result

Land Use	User's Accuracy	Commission Error	Variance	At 95% confidence interval	Area Estimation (ha)
Forest	0.7692	0.2308	0.000112	± 0.165	269362 ± 45946
Cropland	0.3462	0.6538	0.000229	± 0.092	170960 ± 41524
Grassland	0.4082	0.5918	0.000175	± 0.139	270506 ± 52880
Wetland	0.9000	0.1000	0.000005	± 0.196	29910 ± 11132
Settlement	0.0426	0.9574	0.000023	± 0.058	58055 ± 27651
Other Land	0.8140	0.1860	0.000118	± 0.118	300392 ± 46959

Caution!

This exercise should be used for learning process in this training. However, for satisfactory result, the sample designed and allocated in Exercise 6A is recommended. Nonetheless, this calculation exercise demonstrates the process for accuracy assessment and area estimation.

References

Congalton, R. G., & Green, K. (2009). *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*: CRC Press/Taylor & Francis.

FAO. (2016). *Map Accuracy Assessment and Area Estimation: A Practical Guide*. Retrieved from Rome:

Olofsson, P., Foody, G. M., Herold, M., Stehman, S. V., Woodcock, C. E., & Wulder, M. A. (2014). Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment, 148*, 42-57. doi:http://dx.doi.org/10.1016/j.rse.2014.02.015