Module 3: Introduction to QGIS and Land Cover Classification

The main goals of this Module are to become familiar with QGIS, an open source GIS software; construct a single-date land cover map by classification of a cloud-free composite generated from Landsat images; and complete an accuracy assessment of the map output. The tools for completing this work will be done using a suite of open-source tools, mostly focusing on QGIS. The land cover map will be created by training a machine learning algorithm, random forests, to predict land cover across the landscape. The random forests model is trained from a user generated reference data set – collected either in the field or manually through examination of remotely sensed data sources. The resulting model is then applied across the landscape. Finally you will assess agreement with a second reference data set generated using a stratified random sampling process and high resolution aerial imagery. The reference data set will be compared to the classified map image to determine the accuracy estimates.

Modules 3 and 4 have been adapted from Exercises and material developed by Dr. Pontus Olofsson, Christopher E. Holden, and Eric L. Bullock at the Boston Education in Earth Observation Data Analysis in the Department of Earth & Environment, Boston University. To learn more about their materials and their work, visit their github site at **https://github.com/beeoda**.

Exercise 6: Introduction to QGIS

Introduction

The tools for completing the workflow in this module are all open-source; QGIS is the primary tool used to complete both the land cover map and land cover change map workflow. A QGIS install was created from the OSGeo4W and is included on the website for download. It includes these additional packages:

- GDAL
- Orfeo ToolBox
- QGIS
- Python

Objectives

- Explore the QGIS Terminal
- Create a false color image from the SWIR, NIR, and Red bands from the cloud free Landsat composite image,
- Stack image bands, and
- Do some basic image band arithmetic.

Part 1: Software Install and Data Organization

A. Installing QGIS

1. Download the QGIS files from the SERVIR website, located here: https://s3.amazonaws.com/bucket.servirglobal.net/trainingmaterials/OSGeo4W64.zip
2. Right click and select to extract the files, save them on the C drive in the following location **C:\OSGeo4W64.** This is a large file, so the transfer might take around 30 minutes.

Note: *If the path name looks different than* ***C:\OSGeo4W64****, your QGIS tools will not work properly.*

B. Download QGIS supplemental scripts and shortcuts

1. Download the **QGIS_Scripts** from https://s3.amazonaws.com/bucket.servirglobal.net/trainingmaterials/QGIS_Scripts.zip.
2. Save these files on your C drive. The path name should read as **C:\QGIS_Scripts.**
3. Open the **C:\QGIS_Scripts** file and copy the **QGIS shortcut** and the **OSGeo4W command line shell shortcut.** You can select both at the same time by holding down the Ctrl key on your keyboard.
4. Paste these shortcuts onto your desktop. If you double click either of these, it will initiate a QGIS session or an OSGeo4 command line session.

Note: *There is also the option to set up QGIS with a Virtual Machine (VM) install provided by researchers and instructors at the Boston Education in Earth Observation Data Analysis, Boston University. The files to get this set up are larger than downloading and installing QGIS via the steps above, so you will need access to high speed internet to complete the Virtual Machine setup. The benefit of setting up QGIS on a Virtual Machine is to avoid some of the issues that may be encountered with different operating systems.*

Visit the following website to learn how to set up a Virtual Machine and download the necessary files here: https://github.com/beeoda/tutorials/tree/master/1_Introduction

C. Setting up your file folder structure

1. Keeping your data organized is important to your workflow. Download and extract the data files from https://s3.amazonaws.com/bucket.servirglobal.net/trainingmaterials/Change_detection.zip. Extract the files onto your C drive, **C:\Change_detection**.
2. You now should have a folder structure **C:\ Change_detection** with subfolders called Data and QGIS_Projects. You will refer to this directory structure throughout the training, so we recommend that you use this setup as you work through these exercises.

Part 2: Setting up QGIS

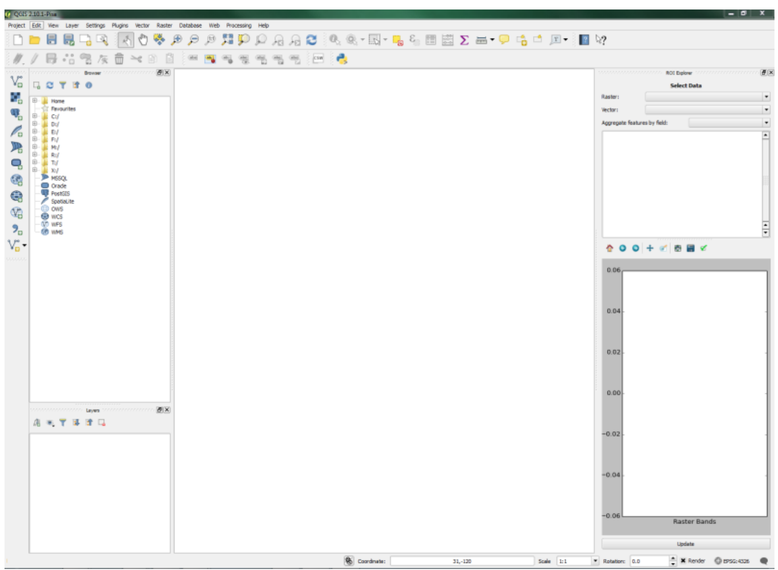
A. Start QGIS

1. Start QGIS by clicking on the **QGIS shortcut** file that you have just copied onto your desktop.
2. If QGIS doesn't open, but instead you get an error message about a missing 'vcruntime140.dll' file then you will need to copy and paste this file into this location: C:\Windows\System32.
   i. A copy of the missing vcruntime140.dll file is available in the QGIS_Scripts folder.

ii. Note: you may need administrative privileges to copy files into the C:\Windows\System32 folder.

***Note**: this shortcut will open a version of QGIS with all the plugins, packages, and associated files that you will be using in this online training course. If you would like to learn more about the QGIS download options, information is available online at https://www.qgis.org/en/site/forusers/download.html and http://trac.osgeo.org/osgeo4w/.*

3. If the QGIS Tips window opens, close it by clicking OK.
4. Install the ROI Explorer plugin.
   i. Click on Plugins, then select 'Manage and Install Plugins…'
   ii. Go to the 'Settings' tab.
   iii. Turn on the box next to 'Show also experimental plugins'.
   iv. Click the 'Reload repository' button.
   v. Go to the 'All' tab on the upper lefthand corner of the dialogue box.
   vi. Search for ROIExplorer.
   vii. Select it, then click on 'Install plugin'.
   viii. Close dialogue box.
5. Your screen should look like the image below – with the ROI Explorer window open on the right hand side of the QGIS interface.



B. Setting up the Default Coordinate Reference System

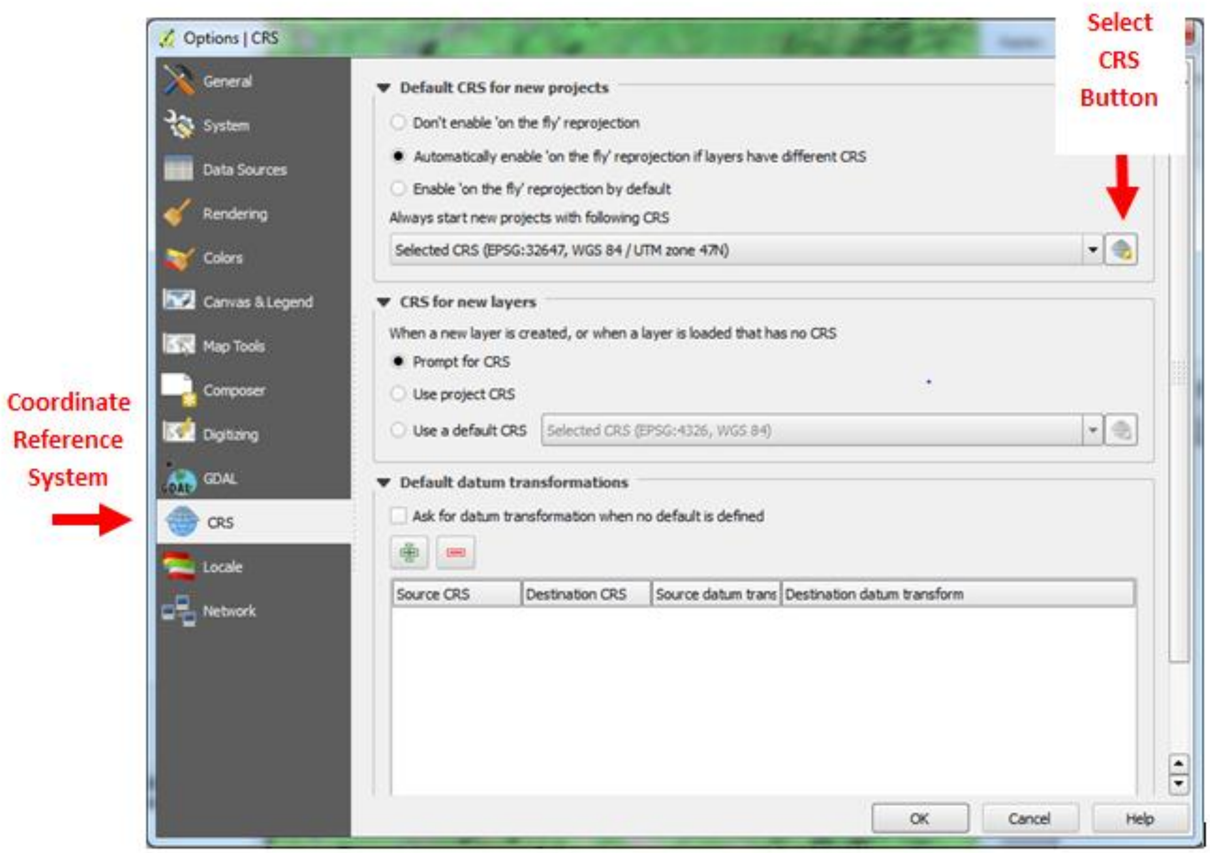
1. Set the default Coordinate Reference System (CRS) to match the cloud free composite data set for the study area. The case study for this Module is from Thailand, and has a coordinate reference system of: WGS 84, UTM zone 47N (EPSG: 32647).

To find the UTM zone and the associated EPSG code for your home country, refer to the Spatial reference website: http://spatialreference.org/ref/epsg/wgs-84-utm-zone-47n/.

   i. From the menu bar select **Settings > Options….** This opens the QGIS Options dialogue.

ii. On the left side of the Options dialogue select **CRS**, which stands for coordinate reference system.

iii. In the **Default CRS for new projects section** click the **Select CRS** button for the **Always start new projects with this CRS** field.

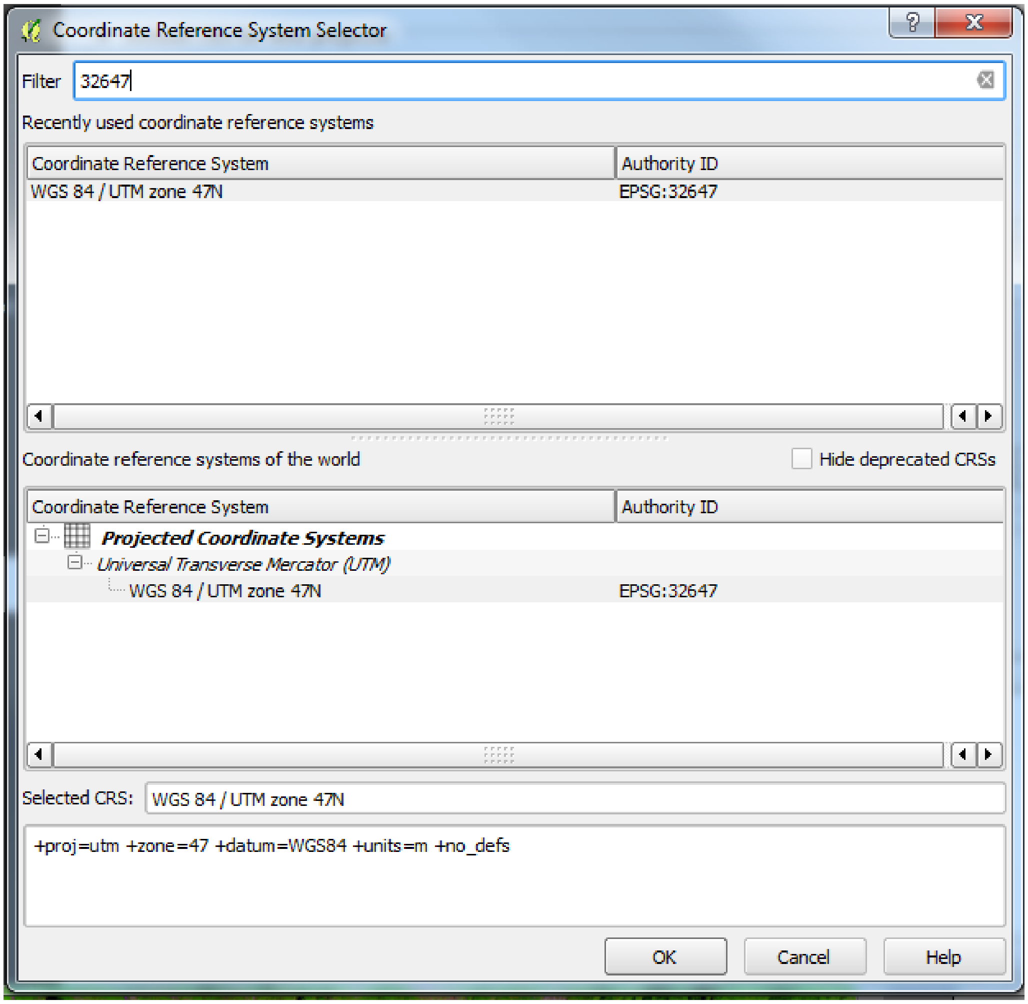


2. This opens the Coordinate Reference System Selector dialogue.

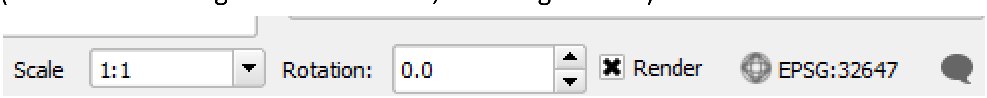
   i. Type **32647** in the **Filter** field at the top of the dialogue.

3. From the Coordinate reference systems of the world section, select **WGS 84/UTM zone 47N** (see image below).

4. Click OK to close the CRS dialogue and OK again to close the QGIS dialogue.



5. Close QGIS entirely and re-open to allow these changes to take effect. Once reopened the default CRS (shown in lower right of the window, see image below) should be EPSG: 32647.



***Note:** to learn more about coordinate reference systems in QGIS, visit their online tutorial information:* *https://docs.qgis.org/2.2/en/docs/user_manual/working_with_projections/working_with_projections.html*

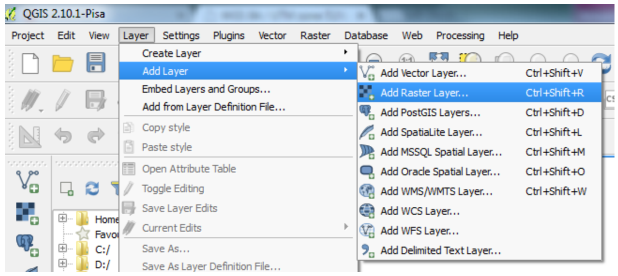
C. Setting up the Toolbars

1. Add the Advanced Digitizing Toolbar
   i. At the top of the QGIS window, click on **View > Toolbars**.
   ii. Click on **Advanced Digitizing Toolbar** to add it.

Part 3: Loading Rasters and Setting Display

A. Open the 2008-2010 cloud free composite image

1. Open the provided 2008-2010 cloud free composite tile. This is a tile from southern Thailand (red box in image below).
   i. From the menu bar select **Layer > Add Layer > Add Raster Layer**.



   ii. or click to the left of the Layer pane.

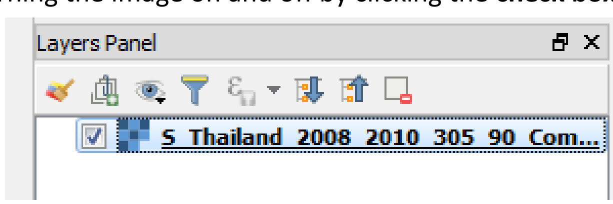
   iii. Navigate to **C:\Change_detection\Data\Composite** and select the image **S_Thailand_2008_2010_305_90_Composite.tif**.

Note: *The image shows up as a black and white image. You will need to adjust the stretch values and color palette to begin to see the details in the data. You will also need to set the no data values, as currently these values are set to show up as black pixels (e.g., the large black patch in the northeast and southwest). You will change the display values in the following section.*

2. Take a few minutes to explore the **Map Navigation tools** to pan and zoom around the image. QGIS has many different tools for navigating around an image. A few are highlighted in the table on the following page to get you started:

Button	Name	Purpose
	Zoom in	Zoom in to a selected area
	Zoom out	Zoom out of a selected area
	Zoom full	Zoom to the full extent of the highlighted layer
	Zoom back	Zoom to the previous extent
	Edit	Edit the highlighted vector layer. This can be used to edit attributes or add and delete features.
	Identify Features	View the attributes of the highlighted layer at the specified location. This works for either raster or vector layers.
	Pan Map	Pan around the map without selecting any layers.
	Save	Save the current QGIS project.

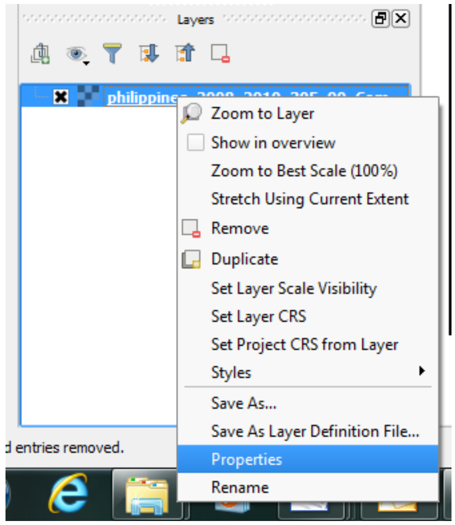
3. Experiment with turning the image on and off by clicking the **check box** next to the layer name.



B. Setting No Data values

Before adjusting the display of the cloud free composite image, you need to first specify the values used to represent cells with no data. The 'no data' value of the images downloaded from the Google Earth Engine script is set to -32768.

1. Right click on the composite layer in the layers window and select **Properties**.



2. Go to **Transparency > No data value > Additional no data value** and type in -32768 (note, don't forget the negative sign).
3. Select OK.

***Note**: now you will notice that the black patches in the northeast and southwest are set to transparent. These areas are outside of our study region, they cover the ocean, so the pixel values are masked and set as no data during the Google Earth Engine export process.*

After setting the No Data values in QGIS, the image displays as a box of light grey – almost white – with a white corner where the ocean is (lower lefthand corner).

Next you will adjust the display values so that the display illustrates the information in the image you are interested in.

C. Setting the display image

You will set the display of the Landsat composite image to a stretched false color composite using bands 6 (shortwave infrared, SWIR), 4 (near infrared, NIR) and 3 (red). A false color image where the SWIR, NIR, and Red bands are represented as red, green, blue (RGB), respectively, highlights vegetated areas as green and bare soil or impervious surfaces show up as pink.

1. Right click on the S_Thailand_2008_2010_305_90_composite image again, then click on **Properties**.
2. In the properties dialog select the **Style** tab on the left of the dialog box.
3. The window will be populated with new options to set.

   i. Set **Band rendering > Render type** to Multiband color

   ii. **Red band** to Band 06

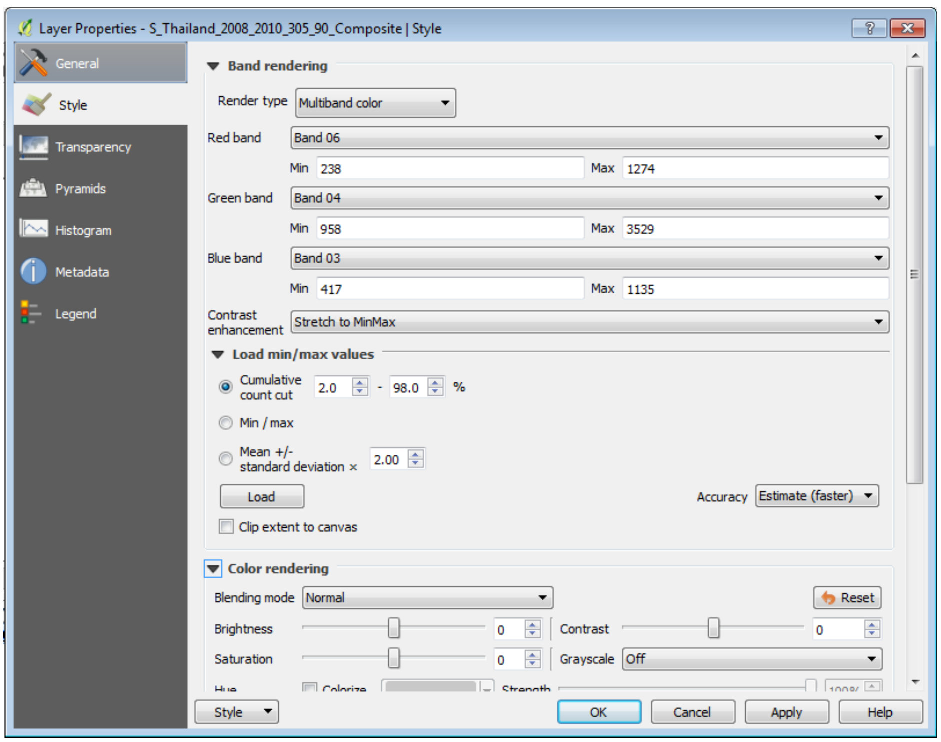
   iii. **Green band** to Band 04

   iv. **Blue band** to Band 03

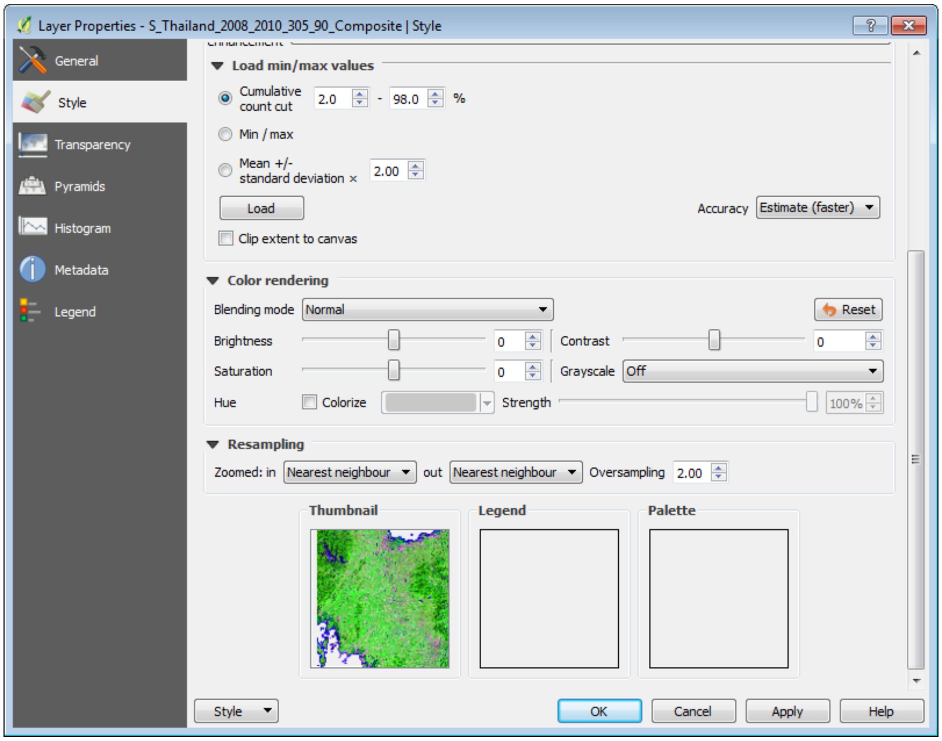
   v. Under **Load min/max values** set **Accuracy** to Estimate and make sure the box next to **Clip extent to canvas** is unchecked.

   vi. Click the **Load** button in the **Load min/max values** section. You'll notice that the Min/Max values under all the color band options have been updated.

   vii. Click **Apply** in the bottom right of the properties dialog box

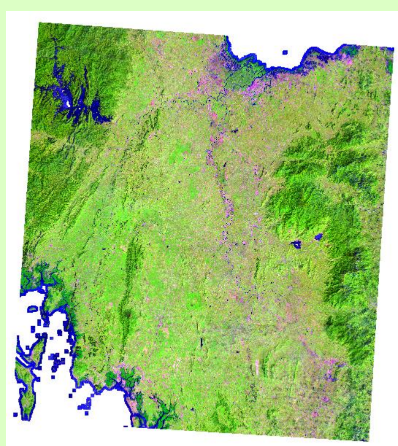


   viii. Scroll down using the side bar slider on the right and at the bottom of the dialog box the image with these display settings has been loaded in the properties panel.



ix. Then click **OK** to save the changes.

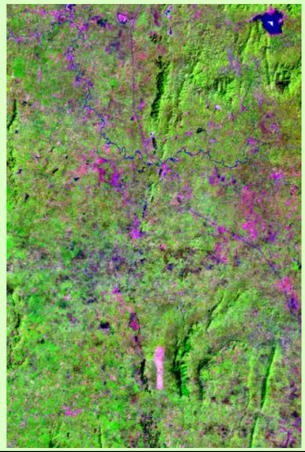
***Note**: The false color SWIR, NIR, Red image is now displayed on the map canvas. Vegetated areas appear in shades of green; areas with exposed soils or impervious surfaces appear in shades of pink and red; water as blue.*



If the colors still look a little muted, you can adjust the display settings again.

4. This time zoom into a part of region that is fully covered by land.
5. Now open the style box again (right click on the composite image - S_Thailand_2008_2010_305_90_composite - in the *layer pane* > *Properties* > *Style).*
   i. Under **Load min/max values** place a check in the box next to **Clip extent to canvas**.
   ii. Select **Load**.
   iii. Then **OK**.

Now your colors should appear more vibrant. See the image below of my adjusted, zoomed in area. Perhaps this is too much contrast – play around with the settings until you have a display you are pleased with.



D. Save your project

1. Save your project by clicking on the save icon.

2. Save the project in the QGIS_Projects file folder in the C:\Change_detection directory. Name it Intro_to_QGIS. You'll notice the default file extension for QGIS projects is '.qgs'.

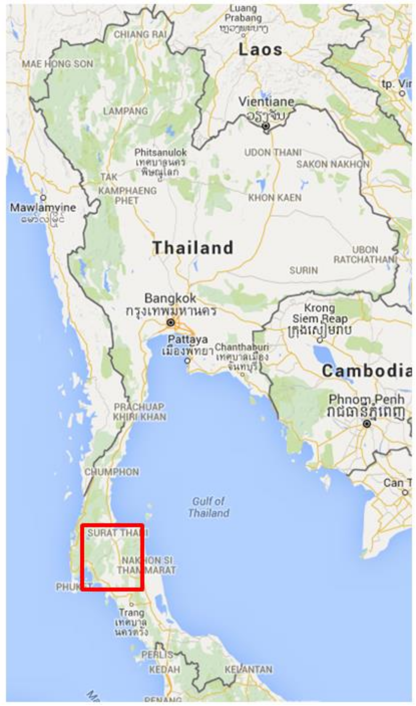
E. Load 2013-2015 composite image and adjust display

1. Repeat steps A to D to load the 2013 cloud free composite. Specify **S_Thailand_2013_2015_305_90_Composite.tif** as the raster to load.
2. After you have the 2013 image loaded and set the display, toggle between the two dates of imagery by clicking on (and off) the top image layer. Click the box to the left of the layer name in the Layers Panel.

3. Do you see any differences between the images?

Part 4: Exploring forest cover change and sub-setting data

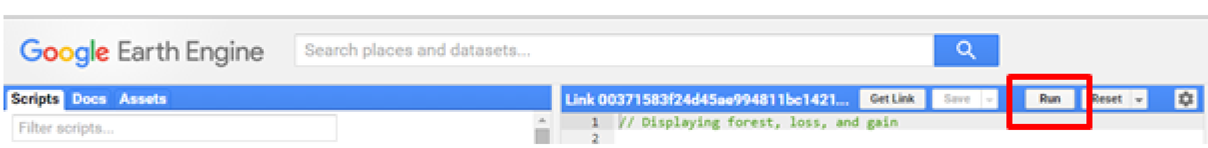
The study region is in the southern part of Thailand (red box in image below); this region was selected because it exhibits changes in forest cover due to plantation activities. In this section, you will do some preliminary data exploration to see which areas within the study region have undergone changes in forest cover between 2008 and 2013.



A. Investigate areas of forest change in Google Earth Engine (Hansen's global forest data products)

1. Open the Google Earth Engine Script by clicking on the link below. This script has been created to load and display forest cover as green, forest loss as red, and forest gain in blue.
   https://ee-api.appspot.com/00371583f24d45ae994811bc1421fdeb
2. If the code window doesn't automatically display after the site has finished loading, select the **Show Code option found** in the center of the screen, towards the top and drag it down.

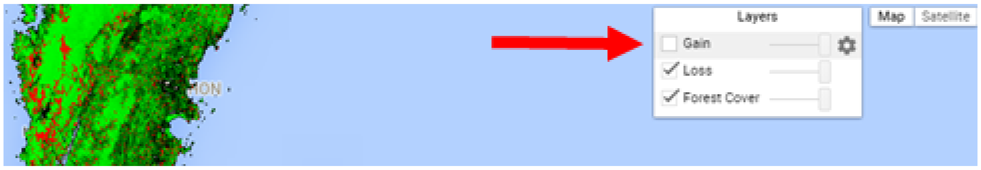
3. The script should run when you open the link. Although if it does not, click **Run** to execute the script. The script is very simple – it loads the tree cover, loss and gain image into the window display. Loss appears red, forest gain is blue.



***Note**: if you'd like to learn more about Hansen's Forest Data Products or examples of additional Google Earth Engine scripting exercises, visit the tutorials available here https://developers.google.com/earth-engine/tutorial_hansen_01*

4. Zoom in and scroll around the region – where do you see changes occurring over time? Also feel free to look around for changes in your home country.

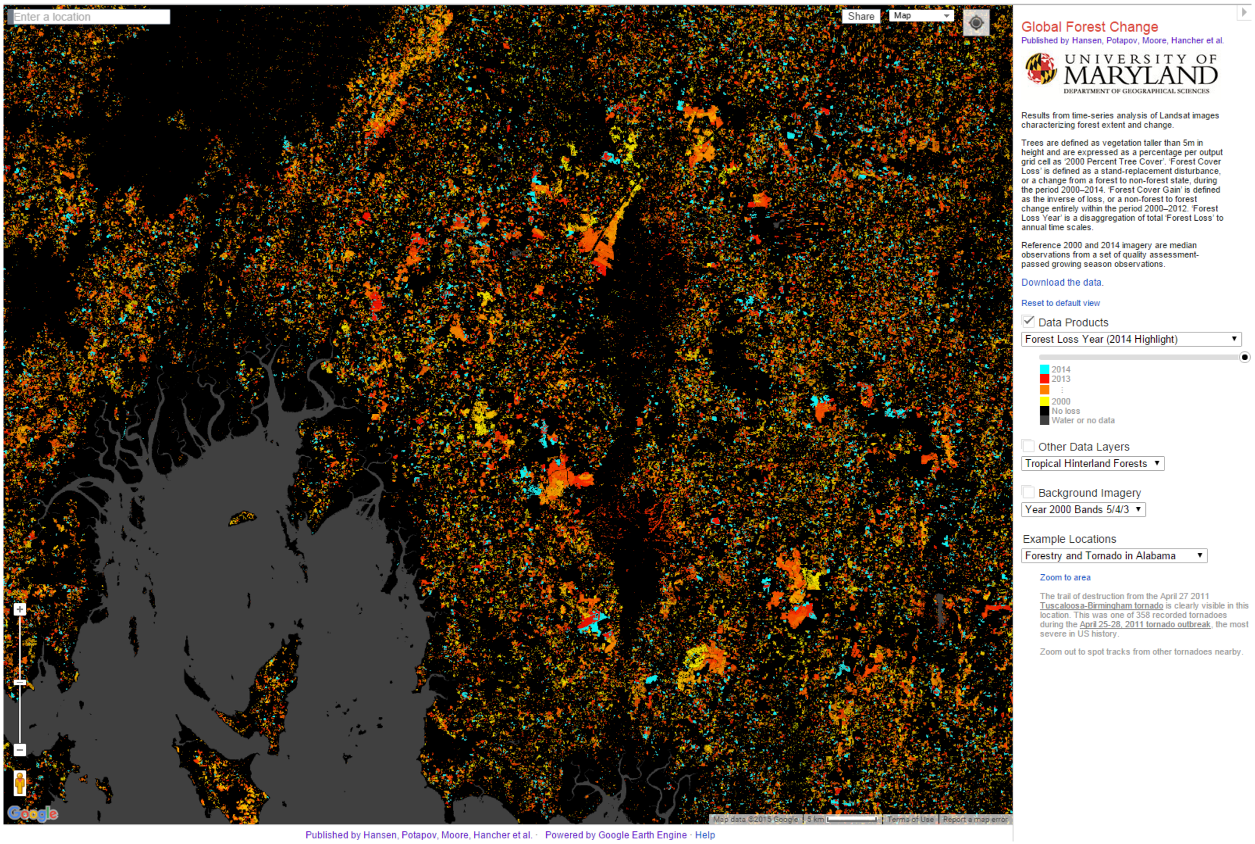
5. Turn the gain layer on and off – you'll notice that the same pixel can be classified as loss and gain. Why is this?



You can also visit the University of Maryland Global Forest Change website, which uses Google Earth Engine to display just the areas of forest loss since 2000. These areas are color coded by the year deforestation occurred.

6. Open this link in your web browser: http://earthenginepartners.appspot.com/science-2013-global-forest
   i. Zoom into the study region (near Phuket in southern Thailand, see image below).

   ii. To view the map symbols (city names, roads, etc), you can adjust the transparency of the forest loss layer by sliding the button along the continuum just above the forest loss year legend.
   iii. Where are the most recent, large losses in forest cover located within the study region? This will be the larger blue, red, and orange polygons in the map.



B. Add forest loss data and zoom to that area

Next you will load a shapefile in QGIS with some previously identified locations that may have undergone changes in land cover between the two study dates. You will investigate these areas so that you can get a sense of what an area undergoing a land cover transition looks like using a false color composite (SWIR, NIR, and red bands).

1. Load the shapefile that marks locations that have likely undergone a transition in land cover between the two study dates. From the menu bar select **Layer > Add Layer > Add Vector Layer** (or click on the Add Vector Layer icon on the left hand panel).

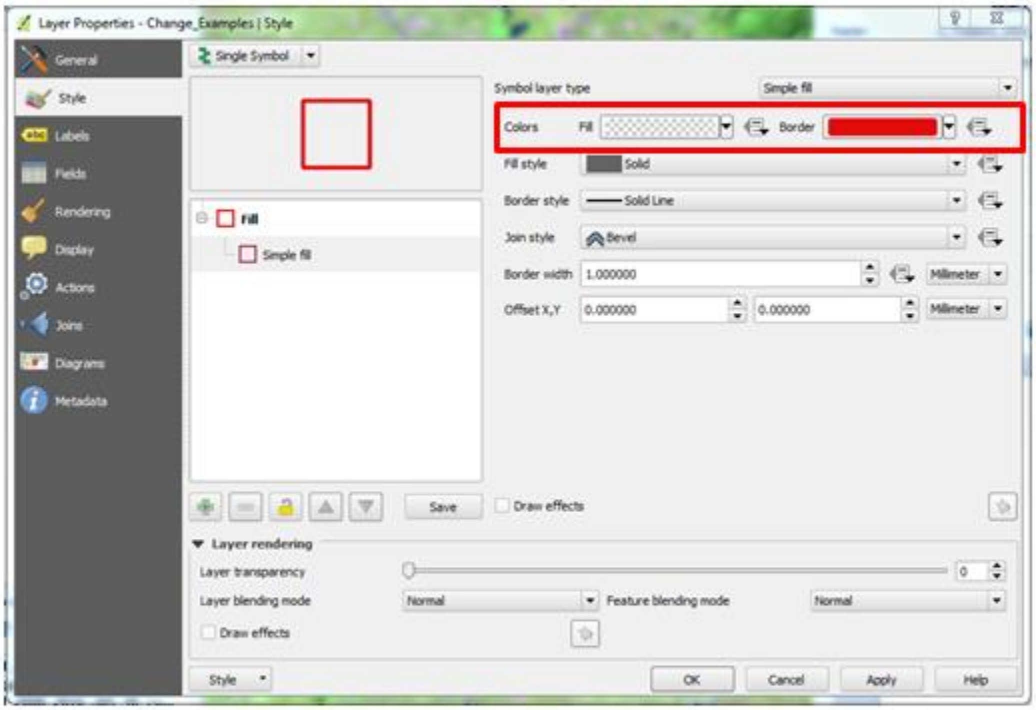
   

   i. Click Browse and navigate to the following shapefile: C:\Change_detection\Data\Shapefiles\Data_Exploration\Change_Examples.shp.
   ii. Click Open.
2. In the Layers panel, right click on the newly loaded layer, Change_Examples, and select **Properties**. Note – you can also just double click on the layer in the layer panel to open the Layer Properties dialog.
   i. Select the **Style** tag from the left panel.
   ii. Click on the Simple fill box.
   iii. Use the dropdown to set the **Colors Fill** to transparent fill.

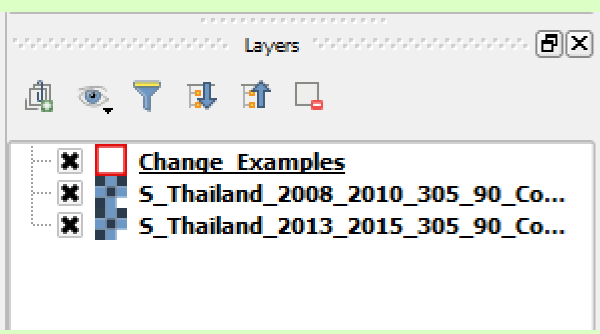
iv. Set the Colors Border to red.

v. Set the Border width to 1.

vi. Select OK to close the dialog.



Your Layers panel should now look something like the figure below (you may need to drag layers up or down to replicate the order shown here).



3. Right click on the **Change_Examples** layer and select Zoom to Layer.
4. The features in this shapefile delineate examples of areas that appear to have a change in land cover between the two time steps.
5. Zoom in closer to one of the change example areas.
6. Toggle the 2008-2010 layer on and off to compare differences between the two images.

***Note**: These areas appear pink in one time period and green in another. This is due to the relative increase in reflectance of the SWIR (displayed red) and relative decrease in NIR reflectance (displayed as green) from the removal of vegetation and exposure of soil.*

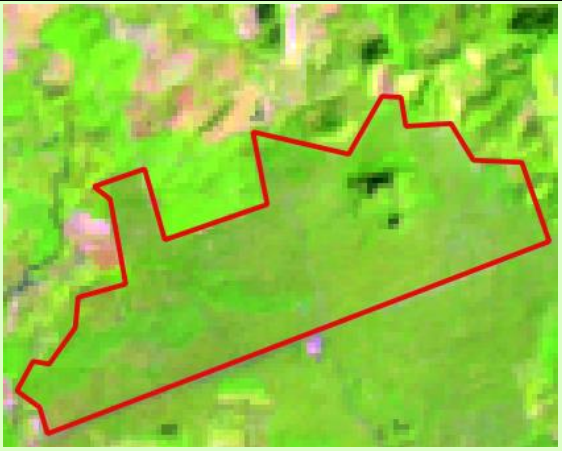


Image from 2008-2010 Composite

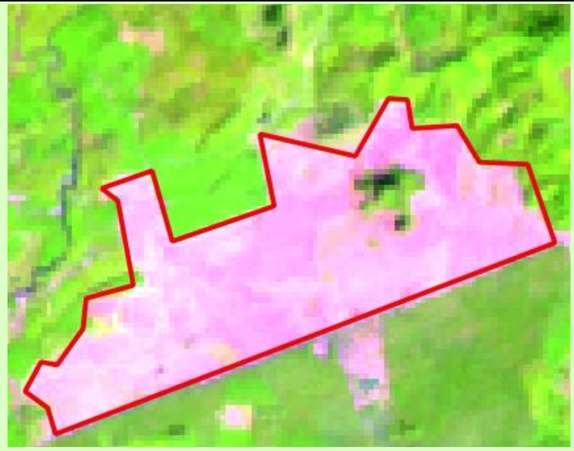


Image from 2013-2015 Composite

C. (Optional) Explore these areas in Google Earth

1. If you have Google Earth installed on your laptop, navigate to the kmz file stored here: C:\Change_detection\Data\Shapefiles\Data_Exploration
2. Double-click the file called **Change_Examples.kmz**.
3. This will open up the file of these same areas in Google Earth. You can now explore the changes in land cover using the imagery in Google Earth.
   i. Scroll the slider along the timeline of available imagery (in the upper left hand corner of Google Earth) to view older imagery.

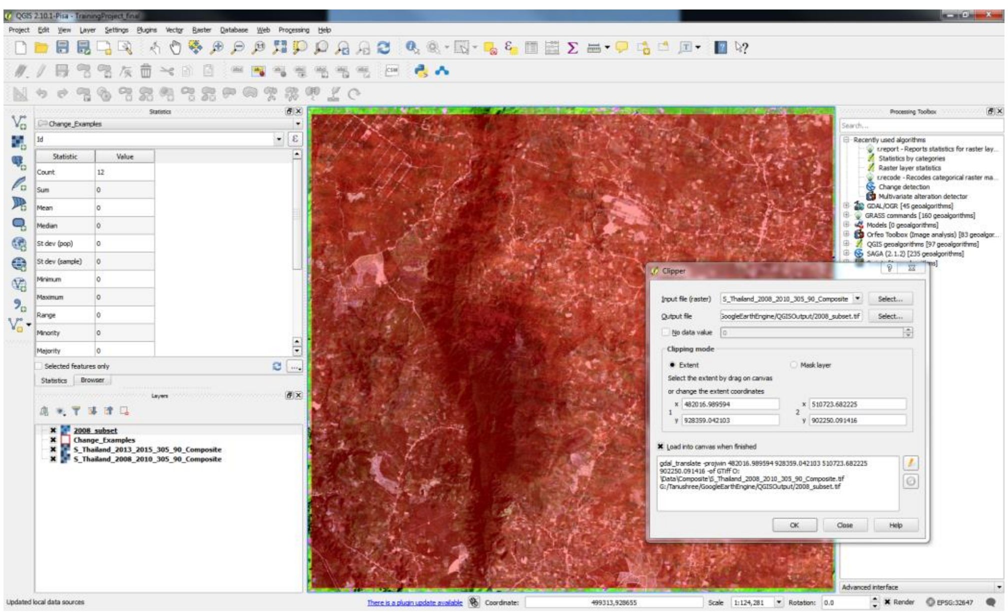
4. The imagery in BING maps is another useful data source to use to explore your study region. BING doesn't include an option to load shapefiles into the interface, but you can scroll around and zoom in to regions manually. You can access the BING imagery with the imagery date information

displayed at this website: http://mvexel.dev.openstreetmap.org/bing/. They provide quite clear imagery around the 2008 time period for this study region.

D. Creating a subset

When testing new algorithms or forms of analysis, computationally it makes sense to start with an image subset. This makes it faster to repeat the analysis with different parameters or options. You will use the extent of the example shapefile to clip the cloud free composite images.

1. Right click on the **Change_Examples** layer in the layer pane window and select **Zoom to Layer**.
2. With your raster image open in QGIS, go to **Raster > Extraction > Clipper**.
3. For the Input file choose your 2008-2010 cloud free composite image.
4. Save the Output file as C:\Change_detection\Data\Composite\imagery_subset\2008_draw_Subset.tif.
5. For the **Clipping mode** choose “Extent”.
6. Now draw on the raster layer the extent you want to subset (hint – trace the extent of the display window to get the region of the change example polygons). If you click and drag with the mouse it will automatically draw a bounding box. Notice that the XY values will be updated. Click **OK** to run the subset.



7. Leave the clipper dialog open and browse to your 2013-2015 cloud free composite layer and create your subset (remember to change the output name so you don’t overwrite your previous subset) – because the dialog box is still open you can use the same box region that you drew for the 2008-2010 image. Click **OK** to run, and then **Close** to close the dialog.
8. The resulting image subsets load with the blue, green, red bands displayed. Use what you learned earlier in the exercise to change the display to the SWIR, NIR, red false color composite for each subset image.

E. Exploring NDVI with a single band gray scale display

Vegetation indices can be calculated from multispectral imagery; these indices enhance features of interest. The normalized difference vegetation index, NDVI, provides a rough estimate of the abundance of healthy vegetation and provides a means of monitoring changes in vegetation over time. It remains the most well-known and oft-used index to detect live green plant canopies.

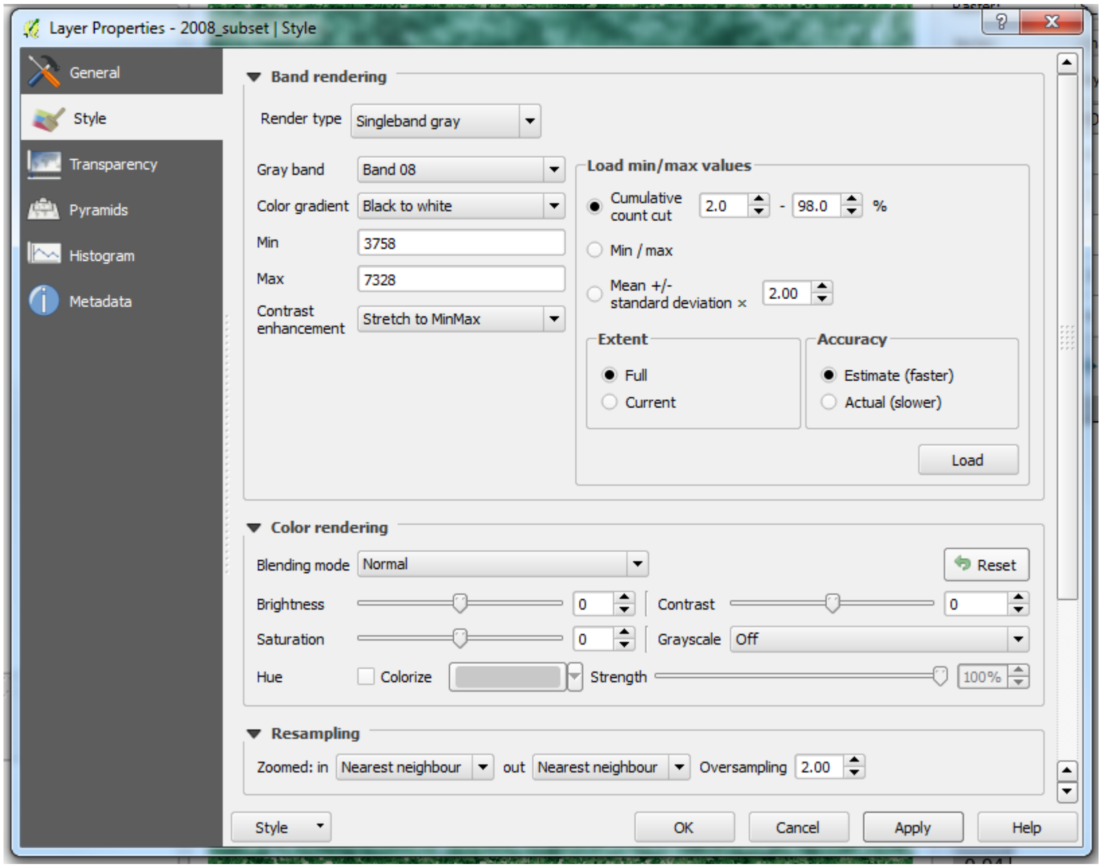
The NDVI ratio is calculated for each pixel in the image by dividing the difference of the near-infrared (NIR) and red color bands by the sum of the NIR and red color bands. The equation is as follows:

$$NDVI = (NIR - red)/(NIR + red)$$

There are three NDVI bands included in the Google Earth Engine cloud-free composite output: the 10^{th} percentile, the median, and the 90^{th} percentile over the composite time range. Including the information about the upper and lower values over time is useful when classifying agricultural land – as these areas often vacillate between high values at the peak of the growing season to low values after harvest.

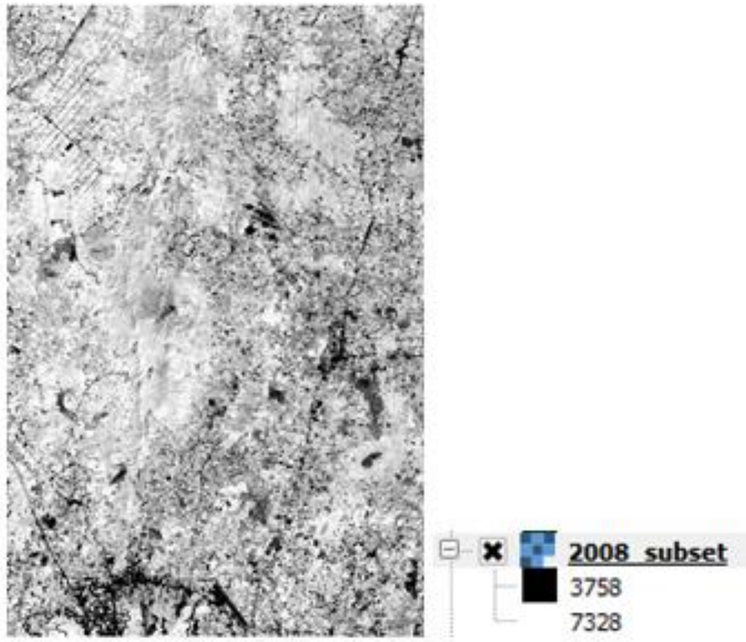
NDVI has already been calculated and included in the cloud free composite image stack from the Google Earth Engine script. Bands 7-9 in the cloud free composite are the 10^{th} percentile NDVI value, the median NDVI value, and the 90^{th} percentile NDVI value. You will explore the median value next.

1. Double click on the **2008_draw_subset** in the layers panel.
2. Select the **Style** tab from the left hand column.
    i. Next use the drop down box to select **Singleband gray** as the **Render type**.
    ii. Use the drop down box to select **Band 08** (this is the median NDVI band) as the **Gray band**.
    iii. Set the Contrast enhancement as **Stretch to MinMax**.
    iv. Keep the other options set as the default value.
    v. Click **Load** to update these settings, the properties dialog should look similar to the following graphic on the next page, click OK to apply the changes.



3. Do the same for 2013 subset data and explore the change in NDVI values in the polygons within the Change_Examples.shp

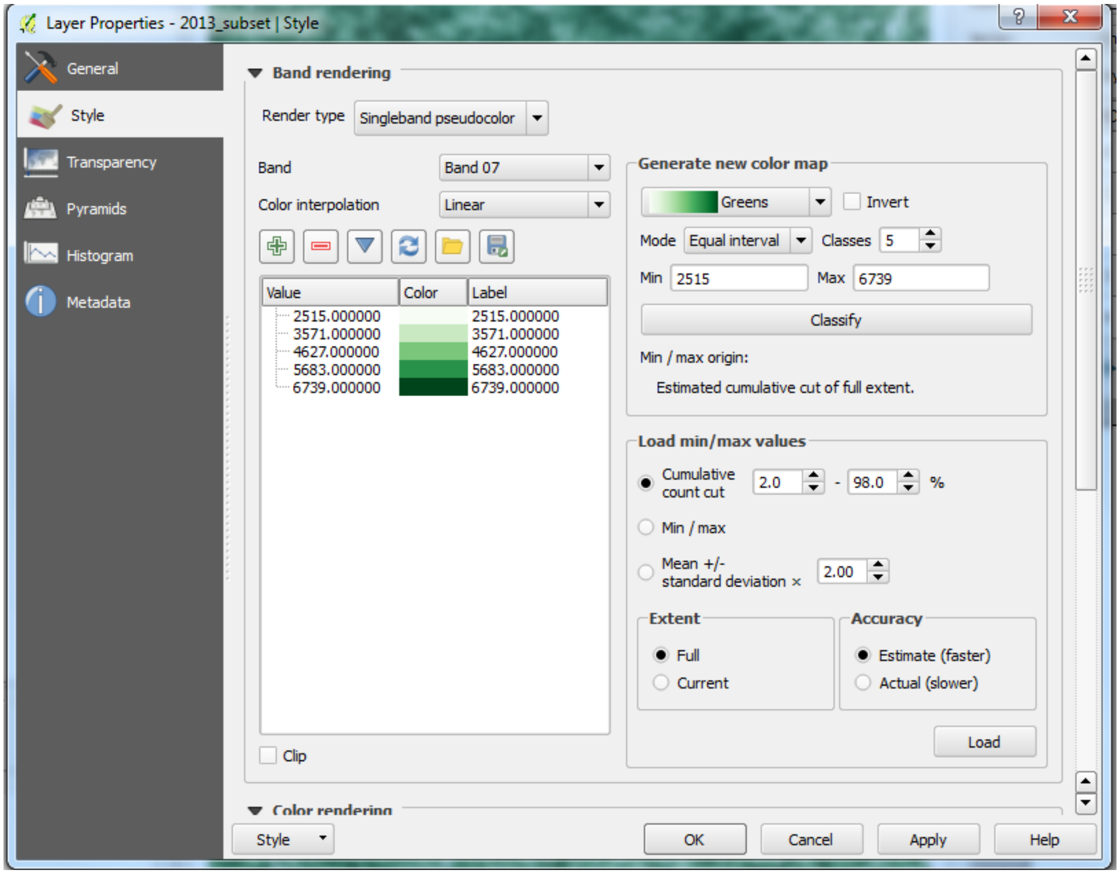
Your image should look similar to the graphic on the following page. In this single band gray display, the high NDVI values are white and the low values are black. Your min and max values might be little different than the snapshot above because of different extents drawn during extracting.



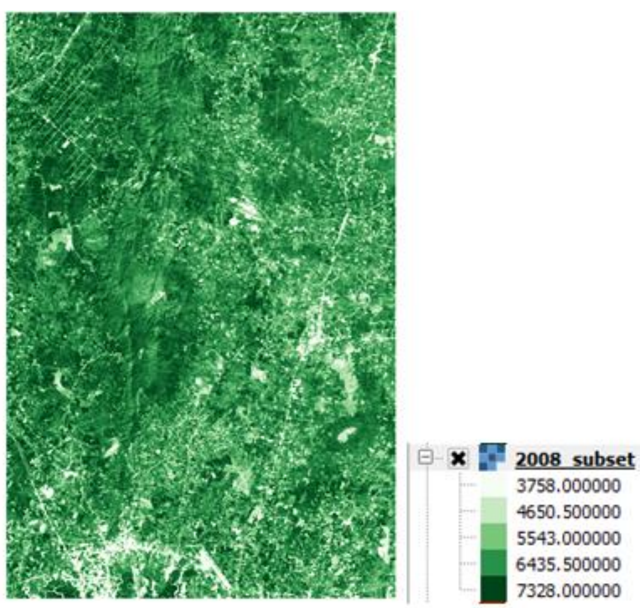
F. Exploring NDVI with a single band pseudo color display

Since NDVI represents how green an area is, change the display to a green color scale.

1. Double click on the **2008_draw_subset** in the layers panel.
2. Select the **Style** tab from the left hand column.
   i. Next use the drop down box to select **Singleband pseudocolor** as the **Render type**.
   ii. Use the drop down box to select **Band 08,** the median NDVI band, as the **Band** (to display).
   iii. Specify **Greens** from the dropdown under **Generate new color map** (ramp).
   iv. Click **Load** to update the min/max values.
   v. Change the **Mode** to **Equal Interval** (or another option of interest to you).
   vi. Click the **Classify** button.
   vii. Then Select OK, the properties dialog should look similar to the graphic on the next page.
3. Do the same for 2013 subset data and explore the change in NDVI values in the polygons within the Change_Examples.shp.



Now the pixels with a high NDVI value will be displayed as a dark green (refer to image below). You can more clearly see areas where there is an abundance of vegetation (green areas) and areas with sparse vegetation (white areas).



4. Save your project.

Exercise 7: Land Cover Classification

Introduction

For this assignment, the requirement is to make a single-date thematic map of some kind using image classification. Learning to do image classification well is extremely important and requires experience. So here is your chance to build some experience.

Objectives

- Create an appropriate list of land cover classes that can be used to create your classification map
- Create regions of interest (ROIs) for your classes that can be used to train a machine learning algorithm
- Classify your image using a Random forests classifier

Part 5: Project Set up

A. Open your QGIS project

1. Start QGIS by clicking on the **QGIS** shortcut on your desktop
2. From the toolbar, go to **Project > Open**. Navigate to the QGIS project you created in the previous exercise (**C:\Change_detection\QGIS_Projects\Intro_to_QGIS.qgs**).
3. Save this as a new project in the same folder (**C:\Change_detection\QGIS_Projects**). Name it **classification.qgs**.

B. Subset images

Since the size of the image you will be working on is not relevant to learning the land cover classification concepts, you will classify a subset of the 2008-2010 cloud free composite image. This will reduce processing time and allow you to redo steps more efficiently. You created a subset in Exercise 6, however you will repeat this step to make sure you are working with the same spatial extent. This time you will subset using a mask layer, a boundary shapefile that has been provided for you in the Data folder.

1. Sub-set the 2008-2010 cloud free composite.
   i. Go to Raster > Extraction > Clipper.
   ii. Set S_Thailand_2008_2010_305_90_Composite.tif as the input file.

iii. Set the output file as C:\Change_detection\Data\Composite\imagery_subset\Subset_2008.tif.

iv. Set the Clipping mode to Mask layer.

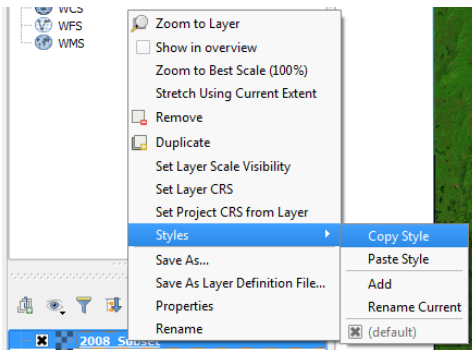
v. Set the Mask layer as Processing_Subset (found in C:\Change_detection\Data\Shapefiles\Thailand\).

vi. Hit OK to execute.

vii. Close the Clipper dialog box.

C. Display subset

1. Use what you learned in Exercise 6 to apply a swir, nir, red (bands 6, 4, 3 in the Landsat cloud free composite image) false color composite stretch to the subset image. You might need to also set no data in this step.
2. Or if you want to copy the style from the full cloud free composite, right click on the image name (**S_Thailand_2008_2010_305_90_Composite.tif**) under Layers and go to **Styles > Copy Style**.



3. To apply this matching stretch to your subset, right click on the **Subset_2008** image in the Layers panel. Select **Styles > Paste Style**.
4. Turn the other cloud free composite images off (click on the box to the left of the layer name in the layers panel).
5. Zoom to the extent of the new subset layer by right clicking on the Subset_2008 layer in the Layer panel and select **Zoom to Layer**.

Part 6: Define land cover classes

The first thing you need to do is define the legend for the map, or the list of classes to be included in your map. What classes do you want to map? For this exercise try to keep it simple, but perhaps more interesting than just forest and non-forest. 3 to 5 land cover classes might be a good choice.

It is also important to have definitions for each of the classes. A lack of clear definitions of the land cover classes can make the resulting maps difficult for others to use. These will probably be working definitions as you move through the land cover classification process. As you become more familiar with the landscape, data limitations, and the ability of the land cover classification methods to discriminate some classes better than others, you will undoubtedly need to update your definitions.

A. Defining forest land cover

*The image on the following page is an excerpt of text from the **Methods and Guidance from the Global Forest Observations Initiative (GFOI)** document that describes the Intergovernmental Panel on Climate Change (IPCC) 2003 Good Practice Guidance (GPG) forest definition and suggestions to consider when drafting your forest definition.*

GFOI Global Forest Observations Initiative

GEO

1.4.4 Forest definition

A forest definition is needed to be able to determine whether deforestation or afforestation or reforestation has taken place, and to define the areas within which degradation and the other REDD+ activities may occur.

The IPCC 2003 GPG defines Forest Land as *including all land with woody vegetation consistent with thresholds used to define forest land in the national GHG inventory, sub-divided into managed and unmanaged, and also by ecosystem type as specified in the IPCC Guidelines. It also includes systems with vegetation that currently fall below, but are expected to exceed, the threshold of the forest land category*. The Forest Land definition in the 2006GL refers to *threshold values*. IPCC therefore anticipates that countries will have a forest definition with quantitative thresholds.

No single definition has been agreed under the UNFCCC for REDD+ purposes. Countries will often have an existing forest definition in place, and the COP has decided that, as part of the guidelines for submission of information on forest reference levels, Parties should provide the definition of forest used, and if there is a difference with the definition of forest used in the national greenhouse gas inventory or in reporting to other international organizations, an explanation of why and how the definition used in the construction of forest reference emission levels and/or forest reference levels was chosen[30].

Countries that do not already have a forest definition may wish to note that for Kyoto Protocol (KP) purposes *Forest …is a minimum area of land of 0.05–1.0 hectare with tree crown cover (or equivalent stocking level) of more than 10–30 per cent with trees with the potential to reach a minimum height of 2–5 metres at maturity. A forest may consist either of closed forest formations where trees of various storeys and undergrowth cover a high proportion of the ground or open forest. Young natural stands and all plantations which have yet to reach a crown density of 10–30 per cent or tree height of 2–5 metres are included under forest, as are areas normally forming part of the forest area which are temporarily unstocked as a result of human intervention such as harvesting or natural causes but which are expected to revert to forest*[31].

In developing an NFMS, countries will need to establish whether there is an existing forest definition, and if not to put one in place. Definitions can differ in ecosystem coverage, which can have a significant effect on the estimate of emissions or removals associated with REDD+ activities, and the allocation to activity (Box 2). Definitions should therefore be used consistently over time, and the definition used to establish the FRL or FREL should be the same as that used for subsequently for MRV.

---

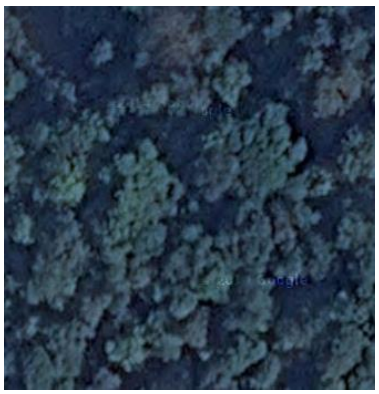
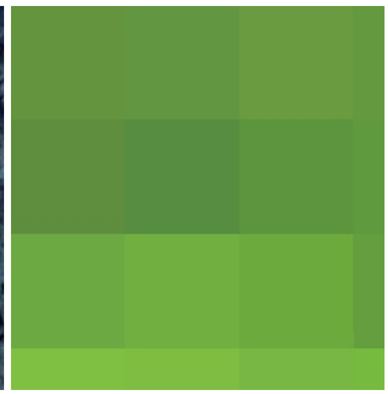
[30] See the Annex to decision 12/CP.17, Guidelines for submissions of information on reference levels

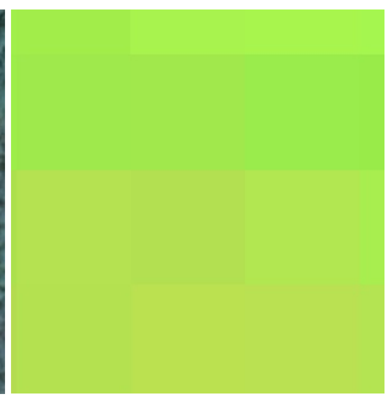
[31] In the Forest Resource Assessment 2010 FAO defines Forest as Land spanning more than 0.5 hectares with trees higher than 5 meters and a canopy cover of more than 10 percent, or trees able to reach these thresholds in situ. It does not include land that is predominantly under agricultural or urban land use. The area threshold falls within the range in the KP definition and the height threshold is at the upper end of the KP range.

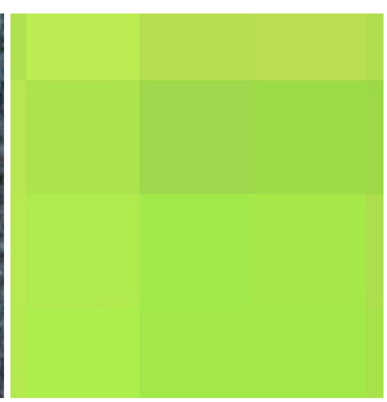
29 GFOI Methods and Guidance

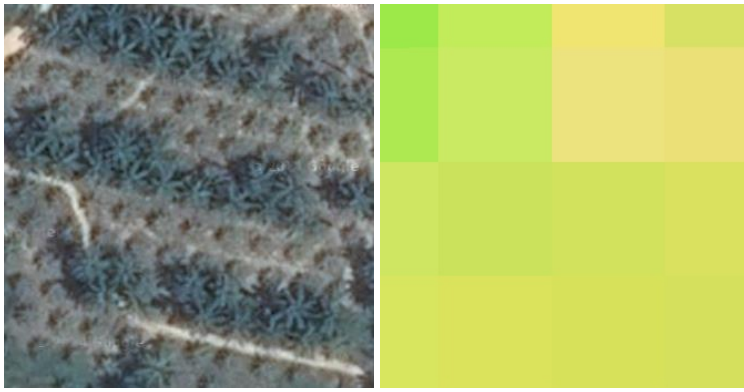
1. Think about what criteria determine if you have a forest vs. another land cover type? An area of trees that are conserved as a national park, an area that is sustainably logged, and an area managed for palm oil all are characterized by tree cover – but they all have very different land uses. Will their spectral signatures differ? Look at the images below – would you characterize all of the images below as forest?

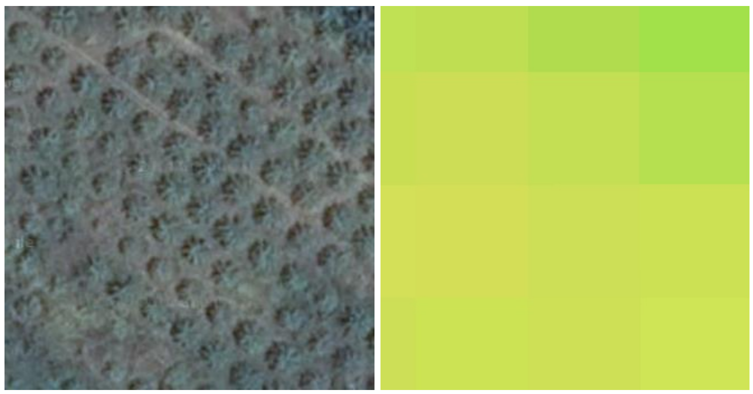
Note: *The snapshots below are from both high resolution aerial images (left hand side) and the Landsat cloud free image composite (right hand side). During this online training course, you will be mapping land cover across the landscape using the Landsat composite, a moderate resolution data set. You may develop definitions based upon your knowledge from the field or from investigating high resolution imagery. However, when deriving your land cover class definitions, it's also important to be aware of how the definitions relate to the data used to model the land cover. You will continue to explore this relationship throughout the exercise.*





2. Draft a definition for the forest land cover class.

Note: *the previous series of images are a good illustration of the connections between land use and land cover. You can map land cover with remote sensing data and tools. However most of the time, you are equally, if not more so, interested in land use–for example, reporting the land use, land-use change and forestry activities for REDD+. Below is an excerpt about land-use and land-cover relationships from the Intergovenmental Panel on Climate Change (IPCC) Working Group II.*

3.3. Land-Use and Land-Cover Change Scenarios 3.3.1. Purpose

The land cover of the Earth has a central role in many important biophysical and socioeconomic processes of global environmental change. Contemporary land cover is changed mostly by human use; therefore, understanding of land-use change is essential in understanding land-cover change (Turner et al., 1995). Land use is defined through its purpose and is characterized by management practices such as logging, ranching, and cropping. Land cover is the actual manifestation of land use (i.e., forest, grassland, cropland) (IPCC, 2000). Land-use change and land-cover change (LUC-LCC) involve several processes that are central to the estimation of climate change and its impacts (Turner et al., 1995). Read the full publication here: http://www.ipcc.ch/ipccreports/tar/wg2/index.php?idp=132

Why does land use matter to REDD+ monitoring and reporting activities, and what can you capture with remote sensing tools?

IPCC greenhouse gas inventory methodologies are used to estimate changes in 5 carbon pools for six categories of land use. The six land use categories include

1. Forest land

2. Cropland

3. Grassland

4. Wetland

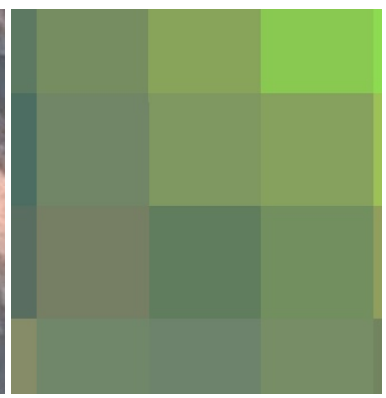
5. Settlements

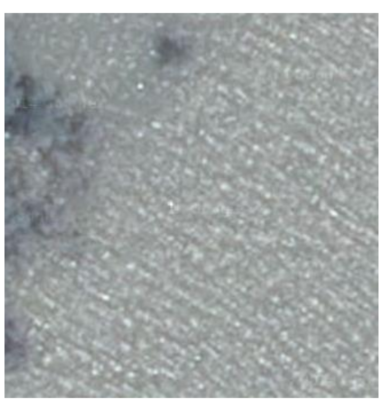
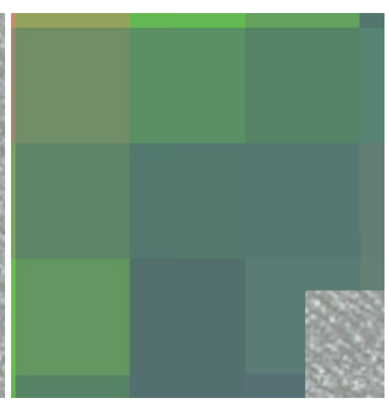
6. Other land

For more information on these land use definitions and on converting land cover to land use, see here: [*http://www.ipcc.ch/ipccreports/sres/land_use/index.php?idp=44#s2-2-1*](http://www.ipcc.ch/ipccreports/sres/land_use/index.php?idp=44#s2-2-1)

B. Defining water land cover

1. Now look at water. Think about what criteria determine if you have water vs. another land cover type? Is your data sensitive to these criteria?
2. Are fish ponds in the same class as the ocean along the shoreline? Again, here is an example of similar land cover, but with different land uses – and subsequently different carbon processing rates and characteristics (carbon sources and sinks). Would you classify both images below as water? Will their spectral signatures differ?

3. Using these questions and images as a guide, draft a definition for your water land cover class.

C. Define the other land cover categories you intend to map

1. Think about the other land cover classes you would like to map. Use what you have learned above and the Landsat data and aerial imagery from Google Earth Engine to draft definitions of your other land cover categories.

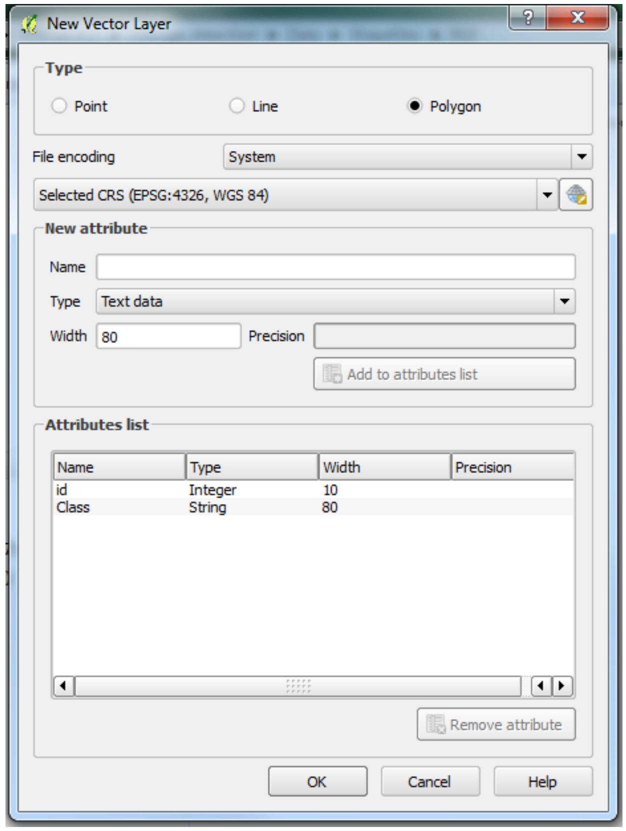
Part 7: Creating Regions of Interest for the 2008 image classification

Any supervised classification method requires prior identification of training samples, also referred to as regions of interest (ROI). High quality training data is necessary to get good land cover map results. In the most ideal situation, training data is collected in the field by visiting each of the land cover types to be mapped and collecting attributes. When field collection is not an option, the second best choice is to digitize training data from high resolution imagery, or at the very least from the imagery to be classified.

In this assignment, you will define regions of interest (ROIs) through a combination of high resolution imagery from online sources and the Landsat cloud free composite. These will be used to train the classifier in a supervised classification using the Orfeo tool and random forests algorithms. The goal of training the classifier is to provide examples of the variety of spectral signatures associated with each class in the map. This can be done in the form of training polygons, which are digitized on an image as Regions of Interest (ROIs). It is important to take time when collecting your training data. This is very important!

A. Create your regions of interest shapefile

1. To collect ROIs in QGIS you need to create a shapefile with features corresponding to your different ROIs. Go to **Layer > Create Layer > New Shapefile Layer**.
   i. Under Type, click **Polygon**.
   ii. Specify the same coordinate reference system as the image you intend to classify – **EPSG: 32647 – WGS 84/UTM zone 47N**.
   iii. Now add an attribute to your Polygon with the
      (a) Name: **Class**
      (b) Type: **Text data**
      (c) Click the **Add to attribute list** button



iv. Your box should look like the image above.

v. Click OK to create the shapefile.

vi. It will prompt you to save the ROI. Save it as **C:\Change_detection\Data\Shapefiles\ROI\ROI_1.shp**.

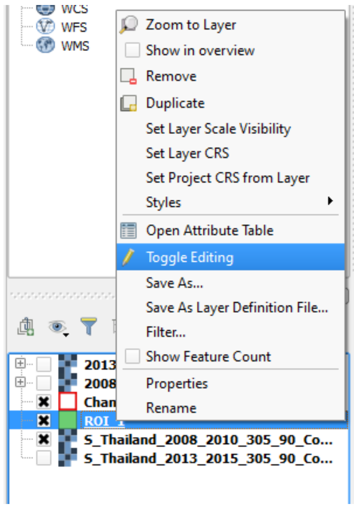
B. Digitize your regions of interest

When you draw your ROIs in the next step you will need to represent them as integer values. If you have 5 classes, then your classes will be labeled 1-5.

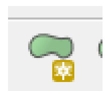
1. In the table below, fill in the corresponding number for each class. Make sure you add in any additional land cover categories you have decided to map. This table will help you keep track of which number corresponds to which class label.

Land Cover Class	Numeric Value
Forest	1
Water	2
Other	3

2. Start drawing your ROIs by right clicking the newly created shapefile in the Layers panel > **Layers panel > Toggle Editing**.

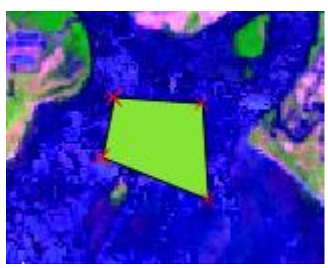


3. Click the Add Feature button (or Layers > Add Feature).



   i. Left click in the image to draw a polygon, and right click to complete the polygon.

4. When you finish drawing an ROI, it will ask you to fill in a value for ‘id’ and ‘Class’. Enter the class and number that correspond to the class you’ve just drawn. Revisit the classification scheme in the table you populated a moment ago to keep track of which number corresponds to each land cover class.
5. Below is an example of an ROI for a water class.
6. After you enter the number for the class, click OK.



Note: *To create the ROIs today, you can also refer to the imagery data sets available in Google Earth. Refer to Appendix 2 to learn how to work with Google Earth. Make sure you are looking at imagery collected around the 2008 to 2010 time period.*

Forest – there is a good example of a patch of natural forest in the center of the subset (in the Khao Phanombencha National Park). There are also a number of forest plantations. Make sure that you create forest regions of interest that represent both areas if you have included both in your land cover classification definition.

7. Repeat this multiple times for each land cover class.
8. When collecting ROIs, pay attention to the tips in the notes below.

9. When you are done creating ROIs, right click on your shapefile in the Layers panel and click Toggle Editing to stop editing.
    i. Save your edits.

> ***Note:*** *You can have more than one polygon that represents the same class. In fact, you should typically have many polygons for each class. Many small ROIs are better than a few large ones (a few hundred pixels tops for each ROI). This can help account for the spectral variations in each land cover class. If you provide just one example of forest, it will be hard to recognize the variety of spectral signatures associated with that class. If you have multiple sub-classes under the same class (such as different tree species all being considered forest), your ROIs should reflect the full breadth of possible spectral signatures.*
>
> *You will find that you will need to adjust your training data a number of times (maybe many times!) in order to produce a quality classification. The actual classification part is just pressing a button, what most influences the accuracy of your results is the quality of your training data. Therefore, make sure to take your time when collecting ROIs.*
>
> *The amount of ROIs for each class should be proportional to the total area of that class in the image. If you have more forest than water, then you should have more ROIs for forest than you have for water.*

C. Analyze your ROIs

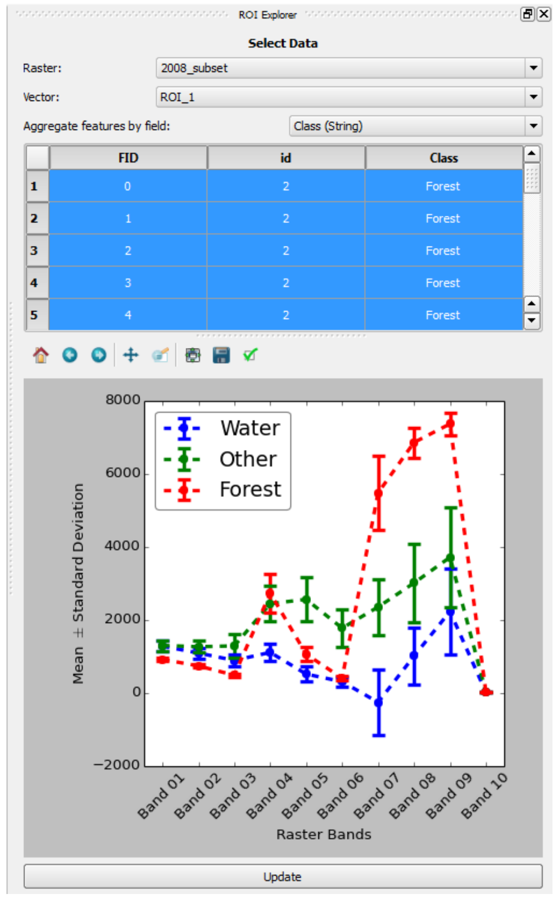
To help you understand why some categories you are trying to map might get confused with others, you might consider plotting the spectral signatures of your different land cover classes. In this section, you should notice a few things when exploring the spectral signatures of your land cover classes. First, some classes are more distinct spectrally than others. For example, water is consistently dark in the NIR and MIR wavelengths, and much darker than the other classes. This means that it shouldn't be difficult to classify water correctly. Also, not all pixels in the same classes have the exact same values, there is some natural variability! This following steps will help you begin to understand the inherent variability of your land cover classes.

Another factor that will strongly influence the results of your classification is the map class variance. For example, classes with high variance will tend to include more places in the map than you'd like. Similarly, if you use only very small and limited sites as an ROI for a class, you may get extremely low variances and the class will be underrepresented in your map.

1. The user interface of the ROI Explorer plugin should appear docked in the right side of QGIS.
    i. If the ROI Explorer interface is not open, right click on any QGIS toolbar and activate the ROI Explorer from the toolbar menu dropdown list.

> **Note***: you have loaded the ROI Explorer plugin within the supplied QGIS install version. If you need to load this on your own at some later date (or it's not showing up on the installation) – simply open the Plugins menu and click* ***Manage and Install Plugins****. Under the* ***Installed*** *tab, enable the* ***ROI Explorer*** *plugin. If this plugin is not currently installed, you may need to install it from the* ***All*** *tab. If it is not showing up as an option, go to Settings and choose to show also experimental plug ins.*

2. The ROI Explorer user interface tracks all raster and vector data opened within QGIS. Select from the Raster drop down box the **Subset_2008**, the raster image you want to use for the classification.
3. Using the Vector drop down box, select ROI_1 - the vector layer containing your training data regions of interest.
4. Use the ROI "Aggregate features by field" drop down box to select the field (column) within your training data vector file that contains the training data labels- **Class**. This selects the field that is used to aggregate features into class labels for plotting graphs in the bottom window.



5. Now that your raster and vector files are selected, you may view each training data feature within the included attribute table. Selecting features from this attribute table will select them for visualization within the plot. You may select more than one feature by using the Control key and clicking additional features, by using the Shift key to select a range of features, or by using the Control + A shortcut to select all features.
6. Click the Update button below the plot to calculate the mean and standard deviation of all raster bands for each unique class label. The plot will update and display the mean as a point on the plot with the error bars around the point representing one standard deviation in either direction.
7. Continue to analyze your ROIs for separability. Highlighting ROIs from different classes in the attribute table provides an understanding of differences in signatures between land cover classes.
8. You might also want to analyze the within-class variability of your ROIs. (Optional) create another field within your vector layer that provides additional class information, such as degraded,

growing, or mature labels for examples of forests, or simply provide a unique identifier for each example of forest.

i. If you select this additional field within the ROI "aggregate features by field" down box, then you could analyze the differences among these forest sub-class examples by highlighting only features labeled forest and by clicking Update.

Part 8: Classification using Machine Learning Algorithms (random forests)

There are a number of supervised classification algorithms that can be used to assign the pixels in the image to the various map classes. The one you will be using today is Random forests. The selection of features used to train any statistical model should be well thought out and informed by your knowledge of the phenomenon of interest. Here you kept the selection pretty simple, focusing on metrics that are associated with water, impervious surface, and vegetation spectral signatures.

> ***Note on setting Features (predictor variables to train random forests model)****: The land cover classification pilot project in the Philippines undertaken by Deutsche Gesellschaft für Internationale Zusammenarbeit Gmbh (GIZ)[1] (located in* ***Supporting_Documentation*** *folder) showed that agricultural lands were better classified if the lower and upper range of NDVI values over the growing season were included in the classification process. These two values are represented in this project as the 10th and 90th percentile value of NDVI over a two year period (this was specified in the Earth Engine script that generated the cloud free composite).*
>
> *Also, if there are other ancillary data sets available that can help to differentiate land cover classes, it would be good to load this into the project and include them in the model. Examples of additional data sets that would probably be quite helpful to differentiate classes include climatic and topographic (aspect, elevation) information.*

There are many ways to try and improve land cover mapping efforts beyond setting high quality regions of interest, reference data, used to train the model. Using improved or more appropriate classification algorithms, exploring object based approaches opposed to pixel based approaches, or being more creative with specifying the model predictor variables. In the case of being more creative with model predictor variables you can try using multiple dates of data (instead of a single date), or try using texture bands. The possibilities are many and should relate back to the nature of the classes you hope to map. Last but certainly not least is to improve the quality of your training data.

One way of performing a supervised classification is to utilize a Machine Learning algorithm. Machine Learning algorithms utilize training data to efficiently learn how to classify pixels. Using ROIs, these algorithms can train a classifier, and then use the relationships identified in the training process to classify the rest of the pixels in the map.

[1] 2013 Deutsche Gesellschaft für Internationale Zusammenarbeit Gmbh (GIZ) Landsat Land Cover Classification Leyte Island, Eastern Visayas, Philippines. GIZ: Manila, Philippines.

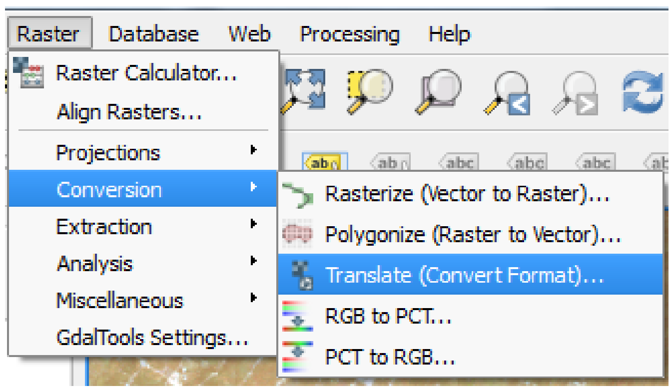
One Machine Learning algorithm that is particularly popular in remote sensing is Random forests. A Random forests algorithm creates numerous decision trees for each pixel. Each of these decision trees votes on what the pixel should be classified as. The land cover class that receives the most votes is then assigned as the map class for that pixel. Random forests are efficient on large data and accurate when compared to other classification algorithms.

To complete the classification of our subset image you are going to use a Random forests classifier contained within the Orfeo toolbox in QGIS. But first, you must train the classifier using the regions of interest you just collected.

A. Remove bands from the composite stack

You will use the cloud free composite bands as predictor variables in the land cover classification model. First you have to remove the ancillary bands – those that you don't expect to supply predictive power for the land cover classification process. You will make two new stacked images. The first will have only the Landsat bands (bands 1-6 in the cloud free composite); the three NDVI bands and count will be dropped. Then you will repeat this but keep the Landsat bands and the three NDVI bands- you will just be dropping the count band.

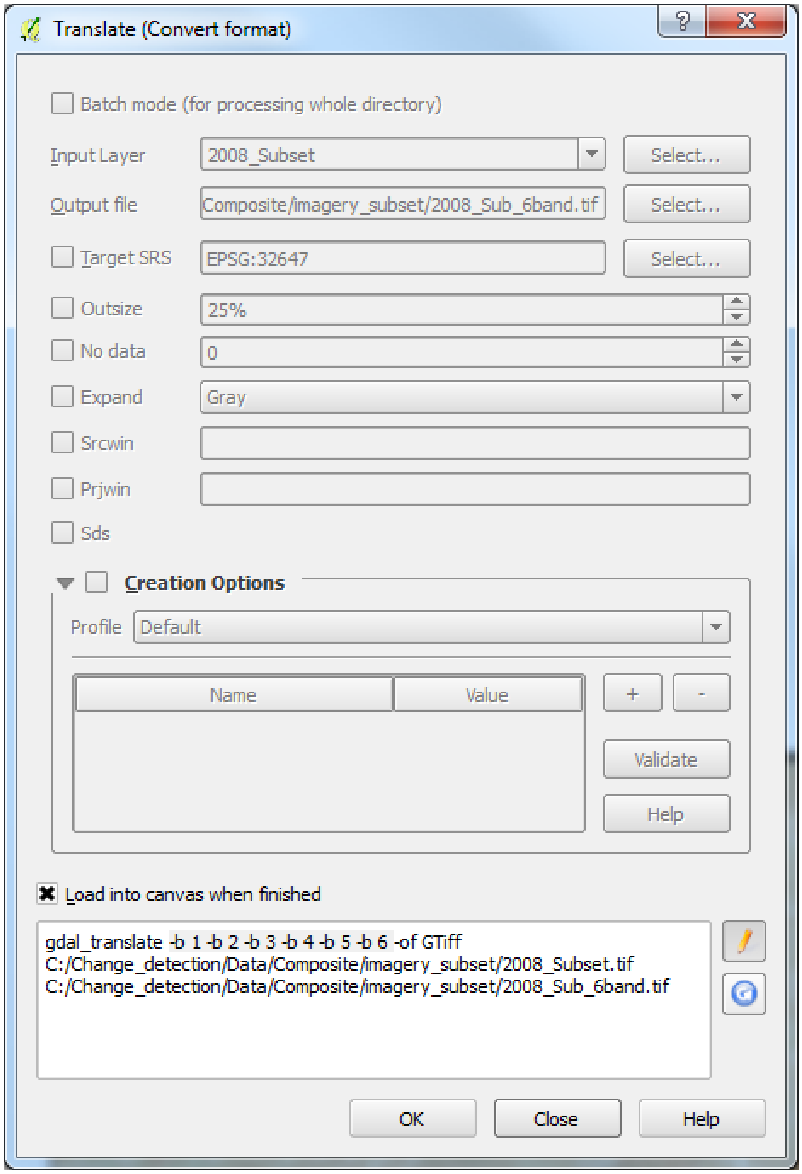
1. Open the Translate tool by clicking on Raster > Conversion > Translate (Convert Format)…



2. Select the Subset_2008 layer as the Input Layer.
3. Save the Output layer in C:\Change_detection\Data\Composite\imagery_subset. Name it 2008_Sub_6band.tif.
4. Now select the Edit pencil at the bottom of the screen

5. You will type in the bands you are interested in saving to the raster stack in the script in the lower panel of the screen. Type the text, found on the yellow line below, in between **gdal_translate** and **–of** in the script window. It is important to type it in (or copy and paste) exactly as it appears below, as the spaces are important. When you are done, you script box should appear as the image below. Make sure there is a space between the **-b 6** and the **–of**.

-b 1 -b 2 -b 3 -b 4 -b 5 -b 6



6. Then click OK to create the new file.
7. Repeat these steps to get a stacked raster with the 6 Landsat bands and the three ndvi bands.
   i. Name this file 2008_Sub_9band.tif.
   ii. The code you will enter is found below.

```
-b 1 -b 2 -b 3 -b 4 -b 5 -b 6 -b 7 -b 8 -b 9
```

8. Close the Translate window.

B. Train the classifier

1. Make sure that the image you intend to classify and the vector file containing the ROIs are added to the Layers panel.
2. To find the tool needed, go to **Processing > Toolbox** (this may already be docked on the right side of the screen with the ROI Explorer).

3. In the toolbox that pops up on the right side of the screen double click **Orfeo Toolbox > Learning > TrainImagesClassifier(rf)**.
   i. If the **Orfeo Toolbox** isn't an option in the **Processing Toolbox**, you'll need to use the drop down menu at the bottom of the **Processing Toolbox** to select the **Advanced interface**.

> ***Note***: *Orfeo has many different types of machine learning classification algorithms. The algorithm can be changed from random forests to another by choosing a different classifier from the Orfeo toolbox in this step.*

4. For the **Input Image List**, select your 2008-2010 Subset Landsat composite image, **2008_Sub_6band**.
5. For **Input Vector Data List** select your ROI shapefile, **ROI_1.shp**.
6. Make sure the **Name of the discrimination field** is the name of the integer attribute you used to differentiate the classes (it should be **id**).
7. For **Output Model**, click **Save to file…** and save it as a **run1.txt** file in **C:\Change_detection\Data\Models**. Note, make sure you type the .txt extension when typing in the name.
8. For now, leave the other fields as they are. These are different parameters to specify how the random forest model is specified. If you'd like, play around with different parameters and see how they affect the end results. When doing so, you will need to retrain the classifier with new output models.
9. To train the classifier click **Run**.

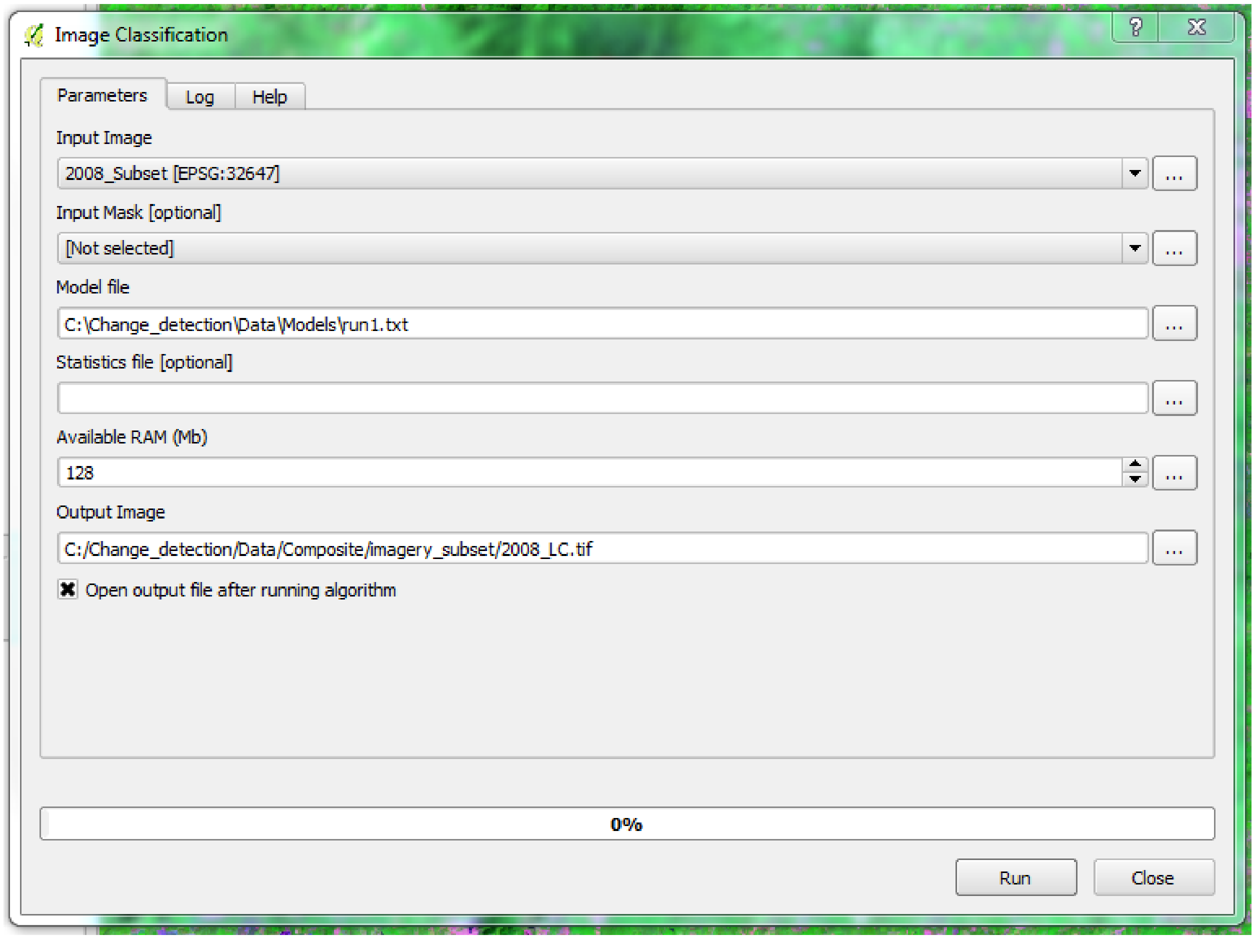
> ***Note:*** *There were some issues running Orfeo on some operating systems. If you run into problems, you can try to set up QGIS with the Virtual Machine (VM) option supported by researchers and instructors at the Boston Education in Earth Observation Data Analysis at Boston University. The files to get this set up are large, so you will need access to high speed internet to complete the Virtual Machine setup.*
>
> *Visit the following website to learn how to set up a Virtual Machine and download the necessary files here: https://github.com/beeoda/tutorials/tree/master/1_Introduction*

C. Apply the random forests model to classify the image

You should now have a text (**.txt**) file saved that contains all the information needed to classify your image. Now the Random forests classifier can use the information in this model and apply it to the rest of the pixels in the image.

1. To find the tool needed for classification, return to the **Processing Toolbox** and go to **Orfeo toolbox** > **Learning > Image Classification**.
   i. For **Input Image** select your Landsat image (**2008_Sub_6band.tif**);
   ii. For **Model file** select the text file you created in the step above (**C:\Change_detection\Data\Models\run1.txt**); and
   iii. For **Output Image** select **Save to file…** and save as **2008_LC.tif** in the **C:\Change_detection\Data\Composite\imagery_subset** folder.
   iv. Leave the rest of the fields as they are. Your Image Classification box should look something like the image below.
   v. Select **Run**.

vi. You may notice that the name present in the Layers Panel is not the same as the file name you just specified. However if you examine your output folder you will notice the file exists.



2. Right click the new layer in the **Layers Panel > Properties > Style** to change the name and color of the classes.

   i. Render type: **singleband pseudocolor**

   ii. At the bottom of the Band Rendering box:

      (a) Mode: **Equal interval**

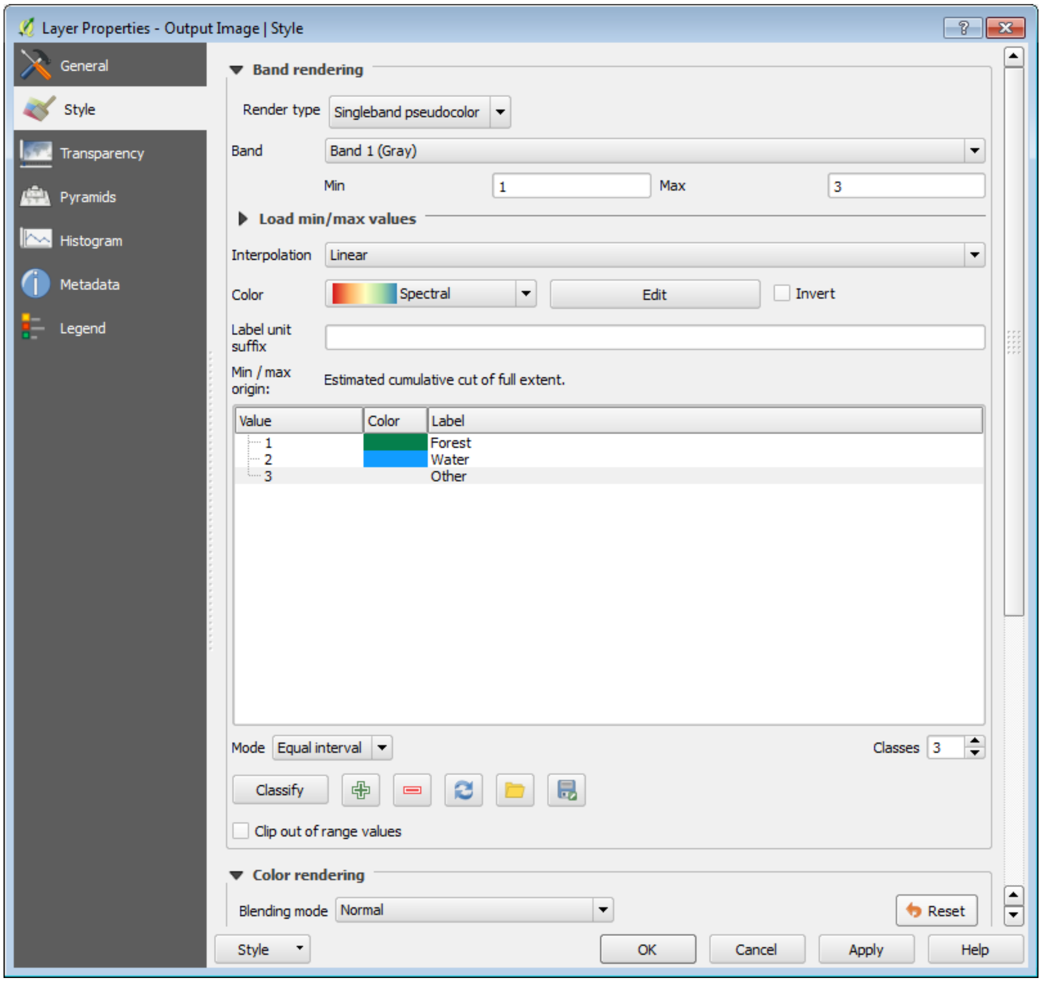
      (b) Classes: the number of land cover categories you included in your classification scheme - I had three (forest, water, other)

      (c) Select **Classify**

   iii. Now in the box on the left (with the values, color swatches and labels) you can double click on any of these to edit them.

      (a) Edit the labels to be text values – e.g., change '1' to 'forest', etc.

      (b) Double click on each color swatch to change it to a color that matches land cover expectations (e.g., water is usually set to a blue display, etc).



(c) Select **OK** to apply the updated settings to view your classification.

> ***Note**: Most likely, the map will not be satisfactory and you will need to go back to **Part 3** to revisit and edit your ROIs. Remember, image classification is an iterative process with a lot of learning by trial and error. You may find you have to add/remove/edit your ROIs several times before you have a map that you are satisfied with.*
>
> *How do you know if your map is satisfactory? Good question, you will delve into that next.*

Part 9: Examine your map

Now your classified image should be displayed as a layer in QGIS. Each different class will correspond to the number you gave that class when creating the ROIs. Look around at the classified image – are you happy with the results? If not (and you won't be the first time), you need to revisit your training data. Some classes are bound to be "too big" and others "too small". If a class is too small, it will help to add a few more training sites for better characterization of the within-class variability. As a rule of thumb, if you have lots of pixels in a class, it will tend to have higher variances, and tend to be "too big" in your results. You will need to spend quite a bit of time looking over your results and modifying your sites to get a good result. One common problem is that there may be land covers in the image that you haven't included in your legend, and those areas will be poorly classified. So you might need to add a class or two.

If the results of the classification are not satisfactory, you may want to investigate the quality of your training samples. Things to consider include:

Quality: *Are the existing training samples that were created from the training shapefile correct? For example, if your ROI is mostly forest but includes a small section of river it will be better to remove it and replace it with one that is composed of pixels that represent only forest. Eliminate and replace any problem samples.*

Homogeneity: *is the variance of spectral signatures high in any of your land cover classes? Or is your data distributed with a bi-modal distribution (or multi-modal)? If so, maybe consider splitting a broad category into a couple of subcategories (e.g., splitting a single* ***agriculture*** *class into* ***vegetated crops*** *and* ***fallow / flooded****). This might be important to do with commonly confused subclasses, such as mangroves or rice paddies.*

Quantity and representation: *Do you have enough training points? Do your training data points cover the full range of variability of a land cover class? For example, does your water class include information for the sea, lakes, and rivers? If not, you may need to create additional training data. You can do this by editing the training data shapefile.*

A. (Optional) Repeat the classification using the layer stack with the three NDVI bands

1. Use what you learned in the previous steps to run a land cover classification using the random forests algorithm. This time, specify **2008_Sub_9band** as the input file instead of the **2008_Sub_6band. Note you will have to create a new random forests model using the same ROIs but with the 9band image, Part 4 step B.**
2. Name the output **2008_LC_ndvi.tif**.
3. Do you notice any differences in the land cover classification output?

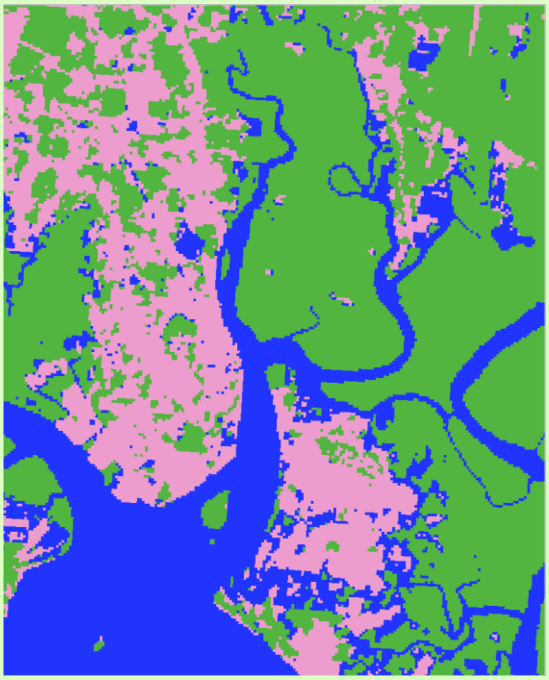
Part 10: Refining the map using a Sieve

When examining your map, you may notice a "salt and pepper" effect in some of your classes. This refers to lone pixels of a class inside a large, otherwise homogenous class that act to make the map look "noisy". Chances are these pixels should be classified as the majority class that surrounds it. If this is occurring in your map and you are confident these pixels are the result of classification errors then they can easily be removed. A technique often used in remote sensing is to 'sieve' isolated pixels and replace them with the classification of the majority class that surrounds it.

A. Map smoothing/clean-up using a Sieve

1. In the general QGIS toolbar, go to **Raster > Analysis > Sieve**.
   i. For Input file select your classified image (**2008_LC.tif** or **2008_LC_ndvi.tif**).
   ii. For Output file save your file as **2008_LC_Sieve10.tif** in **C:\Change_detection\Data\Composite\imagery_subset**.
   iii. For Threshold put **10**. This defines the minimum size of pixel groupings. Adjust this if you would like to remove larger or smaller groups.
   iv. For Pixel connections select **8**. This will take the information from the 8 adjacent cells into account. Start with 8 now, then if you are interested in exploring how this parameter influences the results feel free to change the number and re-run.
   v. Press **OK** to run the sieving process.
   vi. Then **Close** the dialog box.
2. Copy the display style from the 2008_LC layer (in the layer panel, right click on the 2008 classified image and select **Styles > Copy Style**).
3. Paste this style onto the Sieved raster (in the layer panel, right click on the 2008 sieve classified image and select **Styles > Paste Style**).

Notice a change in your classification map? It should look less noisy. Below you can see a subset of a classification unsmoothed (left) and smoothed using the sieving process (right).




***Thought questions**:*

Recall the IPCC forest definition (refer to Part 2 for a refresher). How could you use the sieve function to modify the forest classification to match the minimum forest size threshold from the IPCC definition?

What are the implications of using the sieve function if your goal is to monitor encroachment of forestland by small scale settlements? Or the re-greening of small areas in an urban environment?

Part 11: Image Segmentation

One additional step you can do in the classification process is to classify on segments instead of individual pixels. This can be advantageous if you are finding your classification image to be overly “noisy”. Image segmentation algorithms look for clusters of similar pixels in an image and group them together based on pre-determined criteria. In doing so, you lose the spatial resolution of the image but can create a simpler image that’s easier to work with. In addition, the shape, size, and spectral variance of a segment can aid in the classification process.

The clusters reveal changes at the landscape level, which is why segmentation is often used in both manual interpretation and automated algorithms for change detection. There are numerous algorithms for image segmentation, but the one you are going to use is called Mean-Shift Clustering. Mean-shift analysis is a non-parametric clustering technique that is widely used in image processing. The algorithm assumes that the spectral values in the image are sampled from an underlying probability density function. The dense clusters of data are therefore assumed to correspond with the modes of the underlying density function. The algorithm makes no assumption about the number of modes or size of output clusters.

While the segmentation process can be computationally intensive, it can reveal additional information about the landscape that is not present at the pixel level. Primarily, it allows spectral variance to be utilized in the classification process. The output of the segmentation described in this tutorial is an image that is made up of segments determined by the Mean-Shift Algorithm. The segments contain both the mean and the variance of the pixels that they contain. This image can then be classified in the same way described in the previous sections.

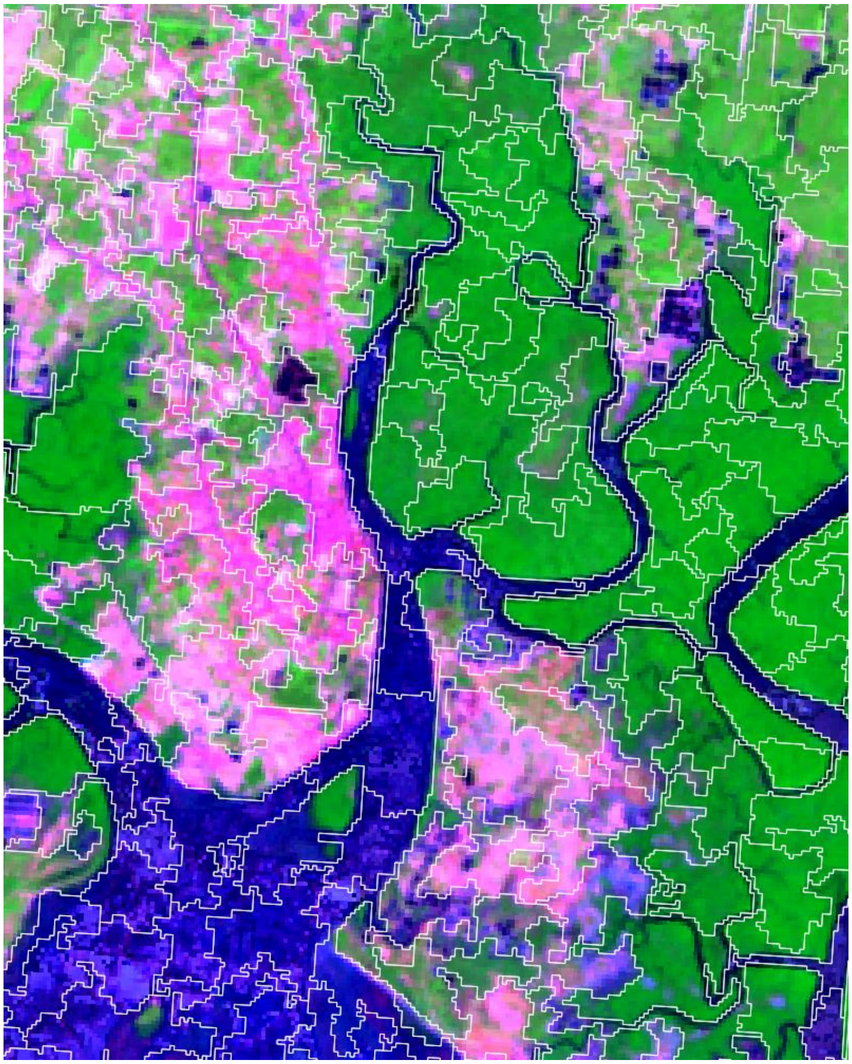
A. Image segmentation

Be sure that you have the subset image open in QGIS. Since image segmentation is a computationally expensive process, it is best to first work with a small subset to test the parameters you will use. At most, start with an image that is a few thousand pixels. If your image is significantly larger that, use **Raster > Extraction > Clipper** to make a smaller one. This will save you significant time in the long run. Once you have determined appropriate parameters then segmentation can be run on the full scene.

1. In the Processing Toolbox, go to **Orfeo Toolbox > Segmentation > Segmentation (meanshift)**.
2. Under **Input Image** select the subsetted image with the NDVI bands (**2008_Sub_9band.tif**).
3. You might have to tweak the parameters a bit to get the results you are looking for. Here are some suggestions:
   i. **Spatial radius**: the size of a window (in pixels) being considered in mean calculation; to start leave this at 5.
   ii. **Range radius**: specifies how close, in terms of Euclidian distance, neighboring pixels need to be to be grouped together; to start, try keeping it at 15.
   iii. **Mode convergence threshold**: Neighboring pixels whose multi-spectral distance lies below this threshold with converge into one mode. Leave this at 0.1.
   iv. **Maximum number of iterations**: Algorithm will stop if convergence has not been reached at this amount of iterations. This can impact computation time. If you want to speed up the process at the expense of segment quality, decrease this value. To start, keep this at 100.
   v. **Minimum region size**: This is intuitively the smallest an output segment can be. This will depend on the types of regions you are trying to find. Increase this to 150.
   vi. **Tile size**: Since the algorithm is so computationally expensive, it will break up in the image into specified tiles to reduce the memory needed. This can result in the output segment image looking 'blocky' since segments are not accurately calculated at the lines between the segments. This will often not matter, but if the blocks are jeopardizing your results, increase this value. If you are running into memory issues, decrease this value. To start, keep it at 1024.
   vii. Save the **Output vector file** as **Segments_08v1.shp** in the folder: C:\Change_detection\Data\Composite\imagery_subset.
   viii. Keep everything else the same, and click **Run**.

Depending on the size of your input file, this process can take a while. For the **2008_Sub_9band** image, it took a little less than a minute to run on my desktop.

Once it is done running, you should get something like the following image, the white lines are the output segments (a vector file) overlaid on the original Landsat image. You may need to change the display style (see instructions below) to assess the output.



4. To set the display to just the segment outlines, right click the shapefile in **Layers Panel > Properties > Style > Simple Fill >** set **Symbol layer type** to **Outline: Simple line.**

i. Change the outline color to white, or another color that has high contrast with your underlying image.

One use of image segmentation is to identify spectrally varying regions in a normally homogenous land cover. Through visual interpretation, this can help detect areas of change.

B. Layer stack for input to classify land cover of objects (segments)

For classification, the vector file needs to be converted to a raster. Using zonal statistics, a raster can be created that contains the mean and variance of the pixels within each segment as the output bands. This 'object' image can then be classified.

1. To calculate mean and variance for the objects in your image you will use a script that must be run in the command line, outside of QGIS. Double click on the **OSGeo4W Shell Shortcut** that you copied onto your desktop earlier in the online training course to open the **OSGeo4W command line shell**.
2. Navigate to the folder where we'd like to save the data. You can do this a couple different ways.
   i. You can copy the code below. Paste it into the command line shell by right clicking your mouse.

```
Cd C:\Change_detection\Data\Composite\imagery_subset
```

   ii. Or this: you can type 'cd \' in the command line and enter the first characters of the path and use the tab key to autofill:

`Cd \Chang` (tab) … etc, etc

3. Next you will run the script. First you specify python. Then the full path name to access the script 'object_stats.py'. Then you enter the Landsat image, the segment vector, and the statistics to be calculated (mean and variance in this case).
   i. For this exercise, you can just copy the text below.
   ii. Right click in the command line shell to paste the text.

```
python C:\QGIS_Scripts\object_stats.py
C:\Change_detection\Data\Composite\imagery_subset\2008_Sub_9band.tif
C:\Change_detection\Data\Composite\imagery_subset\Segments_08v1.shp
C:\Change_detection\Data\Composite\imagery_subset\2008_seg_stats.tif mean var
```

***Note**: Instead of typing the full name in as you work with command line, you can also drag the files from Windows Explorer onto the command line to get the full path name of scripts and files (without typing them in by hand).*

4. A dialogue box will open that informs you that python.exe has stopped working. This is OK your layer was created, so Click OK to close the program and proceed.

The process you just initiated will generate a multiband image where each segment contains the mean and variance for the pixels in that segment in the following band order:

- Mean Band 1 (blue)

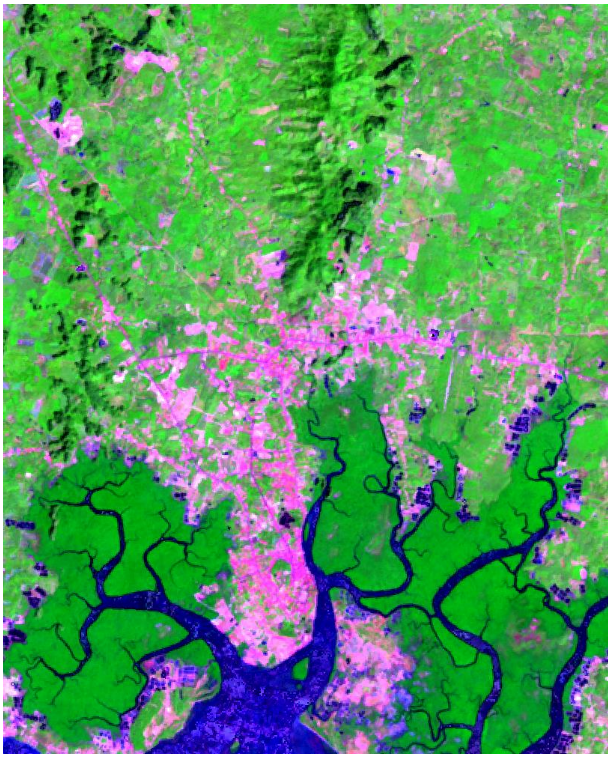
- Variance Band 1 (blue)

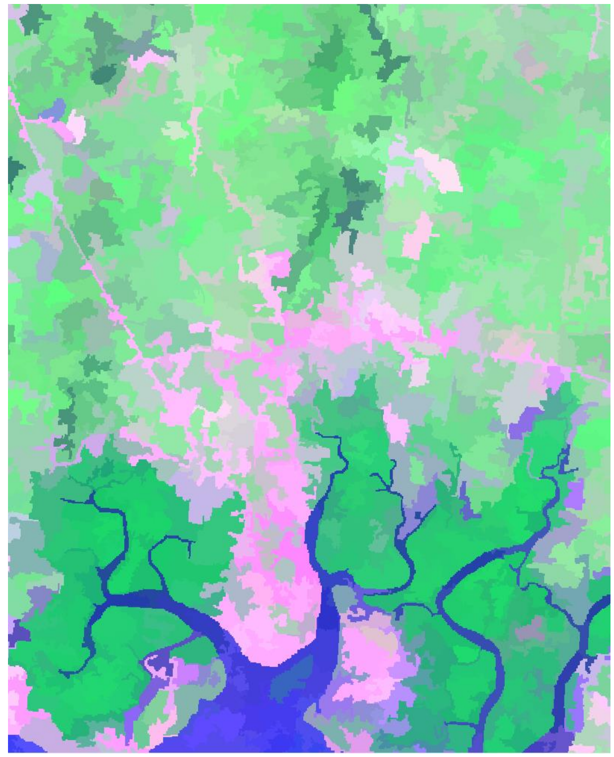
- Mean Band 2 (green)

- Variance Band 2 (green)

- etc…

5. Load the generated stacked segment statistics file (**2008_seg_stats.tif**) into your QGIS project (Hint: **Layer > Add Layer > Add Raster Layer…**).
6. This raster can now be used in an object-based classification approach; display it as false color composite of choice, it should look like the image on the right below. The original Landsat image (swir, nir, red {6, 4, 3} false color composite) appears on the left and mean segment image (swir, nir, red {11-7-5} false color composite) on the right.

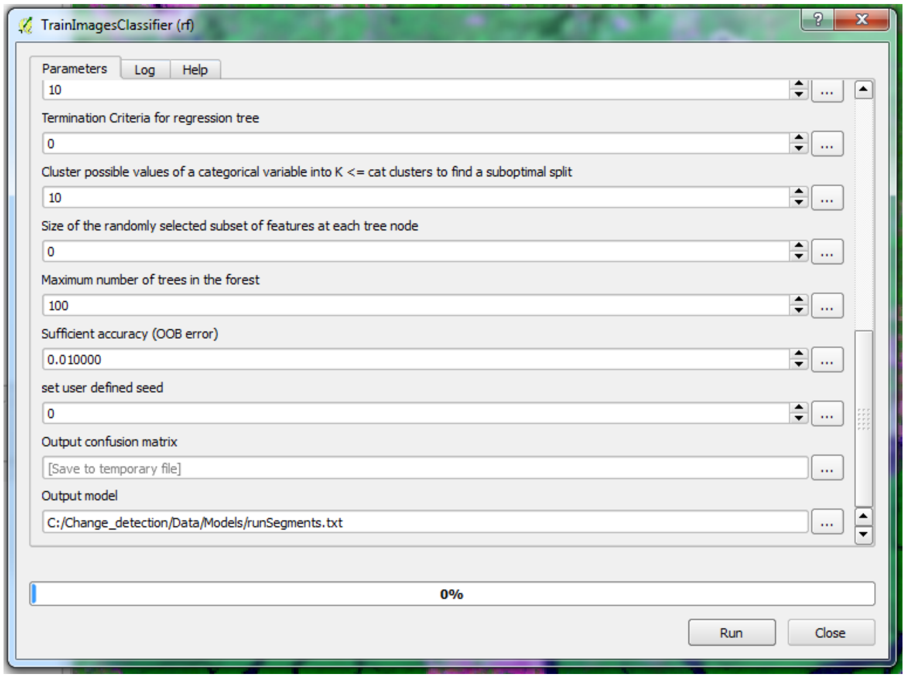




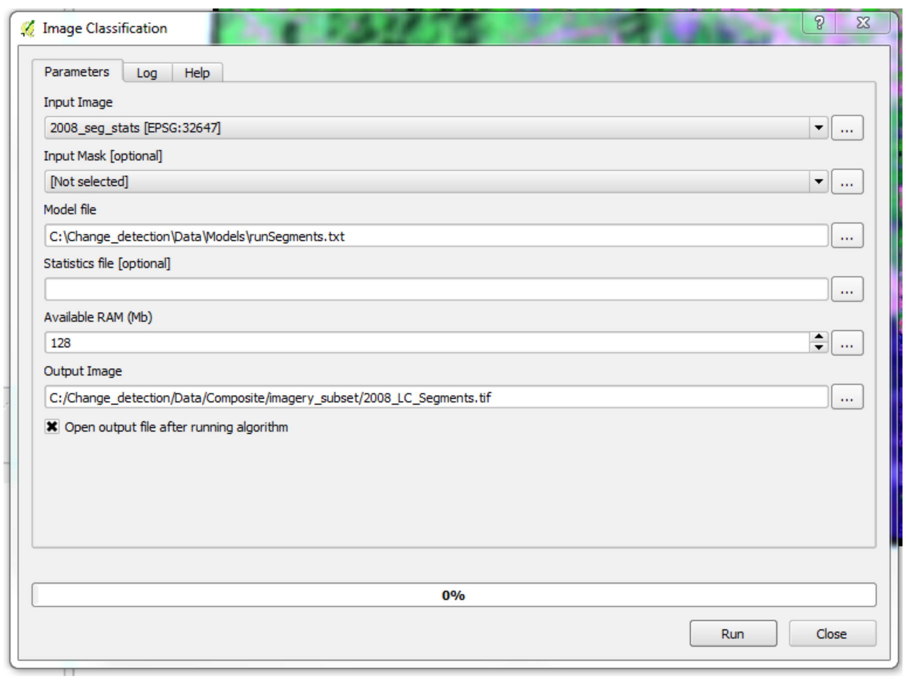
***Note**: for the stacked segment statistics file, the SWIR, NIR, red false color composite bands are 11-7-5 instead of 6, 4, 3 since band1 is the mean of Band1 (blue), band2 is Variance for Band1 (blue), etc.*

C. Run random forests on the segment layer

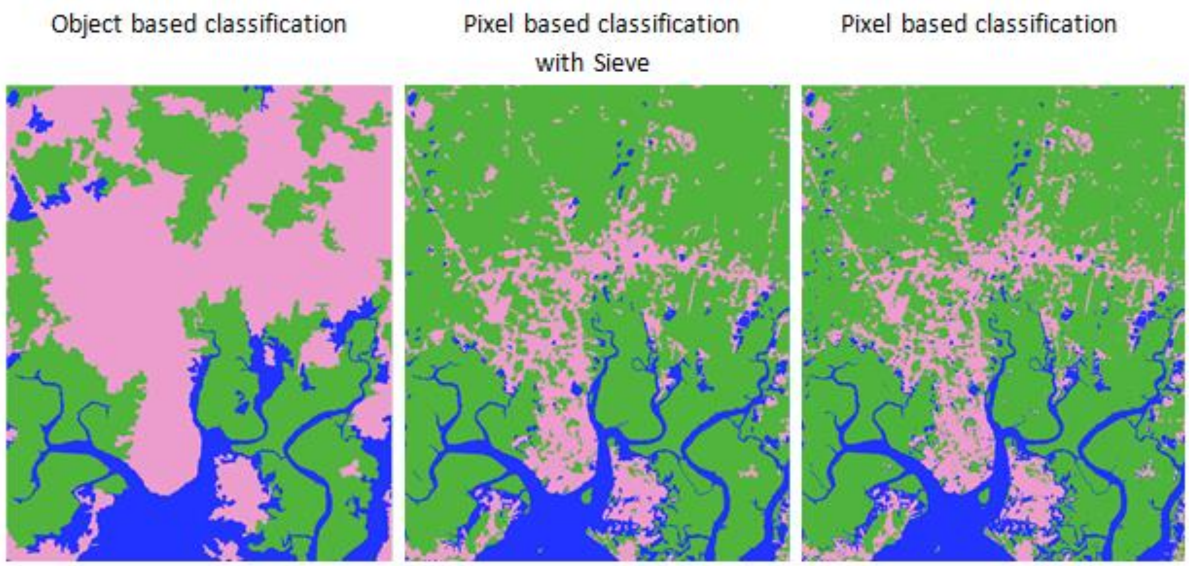
1. Use what you learned in Part 4 of this exercise to run a land cover classification on the stacked segment statistics data set (**2008_seg_stats.tif**).
   i. Name the output random forests model **runSegments.txt**.



ii. Name your classified image **2008_LC_Segments.tif.**



2. Use the same display style to view the output as you did for the pixel based classification (tip: you can copy and paste styles).
3. Toggle the new classification on and off to compare to the previous classification (the pixel based and the pixel based classification that was cleaned up using the sieve process). What are the major differences between these three methods?



Note: *The results of the classification of the segments may result in some over-classification errors (commission), as seen in the graphic above. Think back to when you were creating your ROIs – you were probably making very general shapes, and it is very likely that some of the polygons you created overlap multiple segments. If your ROI actually includes segments from multiple land cover types, this could negatively affect your classification results. Based on the output above, I may want to go back and revise my ROIs for the 'other' land cover class to make sure that they do not include forested areas or other land cover types.*

Alternatively, my minimum segment size may have been set too big. Some of my segments appear to include multiple land cover types. You can also adjust the segmentation process to create more homogenous segments before running the classification process, such as random forests.

D. Map clean-up through segmentation

In Part 6 you sieved out isolated groups of classified pixels and replaced them with the group majority of the pixels around it. This made the map appear less noisy while still retaining the geometry of the features in the map. An alternative method to cleaning up a map is to use image segmentation to assign classifications at the segment level instead of at the pixel level. In the previous step, zonal statistics were used to gain information about the segments and perform object-based classification. This is not the only way segmentation can be utilized in classification, however.

An alternative to classifying objects based on the original Landsat data is to classify the map at the pixel level and use the segmentation results to clean up that classification. In this process, no means or variances are calculated. Instead, each pixel is classified normally. The mode, or most likely classification, for each segment is then calculated. These modes are then assigned to each segment and used as the output classification map.

1. Return to the OSGeo4W Shell (double click on the **OSGeo4W Shell shortcut** on your desktop if the Shell isn't still open).

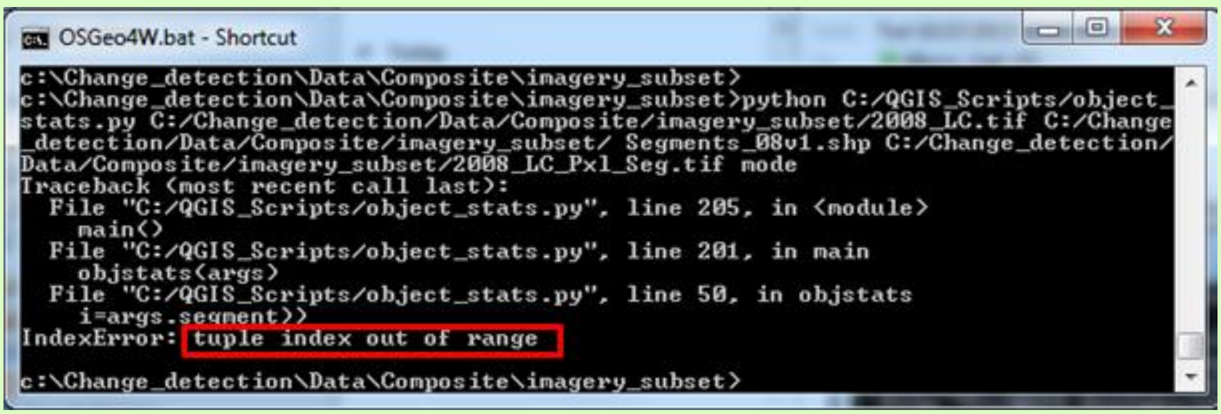
i. If the Shell isn't pointing to the **imagery_subset** folder, navigate to that directory (hint: refer to Part 7, Step B).

You will use the **object_stats.py** script again. This time, instead of extracting statistics from the Landsat image use the original (non-segmented) pixel-based classification map from Part 4.

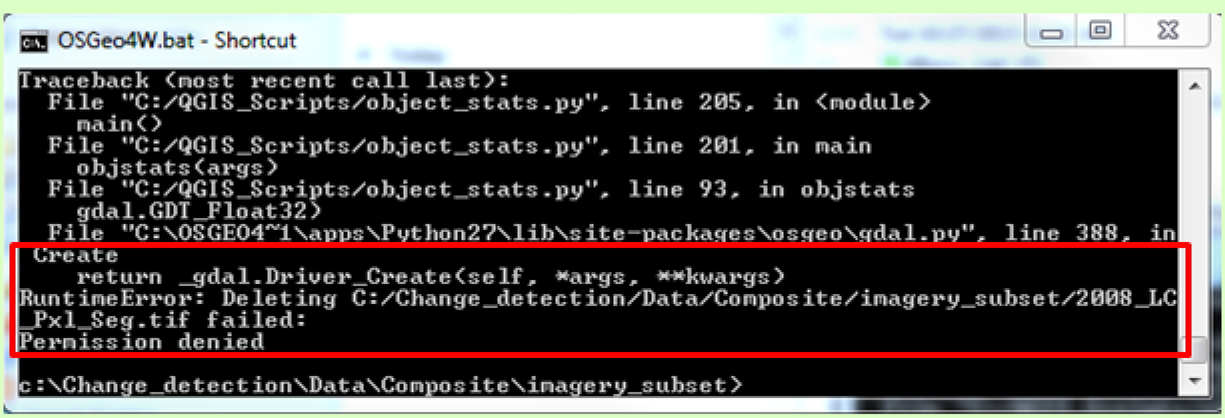
2. The segmentation vector will again be used, but this time specify 'mode' as the statistic. What this is doing is calculating the classification that has the highest likelihood of being assigned inside that segment and applying that classification to the entire segment. This has the effect of "smoothing" out the classification map based on calculated segments. An example command would be:

```
python C:\QGIS_Scripts\object_stats.py
C:\Change_detection\Data\Composite\imagery_subset\2008_LC.tif
C:\Change_detection\Data\Composite\imagery_subset\Segments_08v1.shp
C:\Change_detection\Data\Composite\imagery_subset\2008_LC_Pxl_Seg.tif mode
```

Note: *if you get a message about the tuple index being out of range (see example below), it is likely due to an error in the path name of one of your files. In the example below, there is an extra space in the path name where the segments shapefile is being called.*

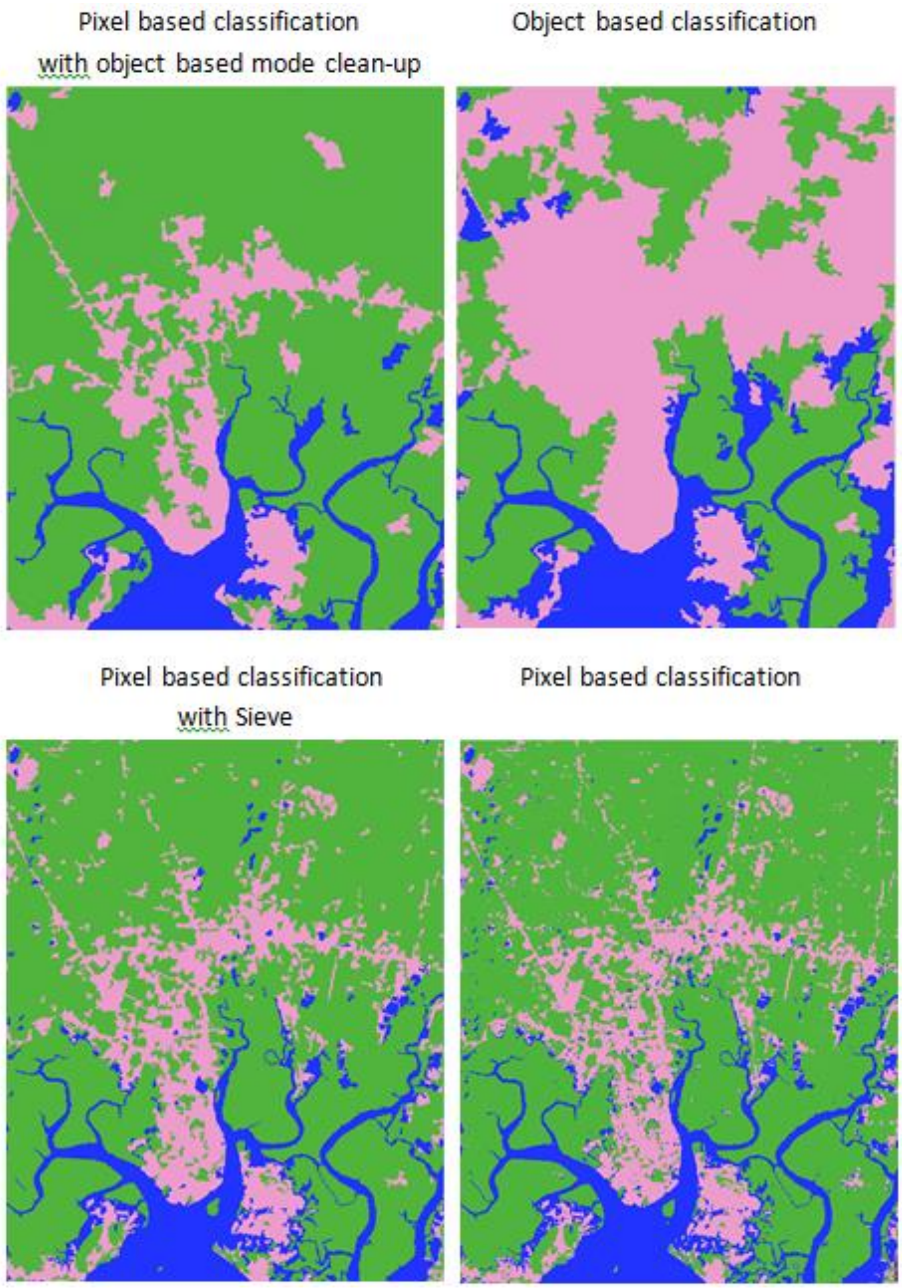


*If you get a warning message with a runtime error following a **Create** line about deleting the output file, followed with a permission denied line, you need to change the name of the output file in the script. You can't overwrite an already existing file.*



3. Add the new raster (**2008_LC_Pxl_Seg.tif**) to your QGIS project and adjust the display style to match the other land cover classification results.

4. The result is a “clean” map like the one on the right in the image on the following page, with the original unsegmented map to the left (note: if this removes too much detail, you will need to redo the segmentation and decrease the minimum region size).



5. Save your project.

Congratulations! You have successfully completed this exercise.

Exercise 8: Accuracy Assessment

Introduction

In this Module you will select a sample of reference observations from the study area with the aim of estimating accuracy of the land cover classification from Exercise 7.

Objectives

- Generate a stratified random sample.
- Learn about response designs and practice generating reference data from a combination of remotely sensed data sources including the cloud free composite and an archive of high resolution aerial imagery available in Google Earth.
- Analyze the agreement between the classified image output, your 2008 land cover change map, and the reference data to estimate the accuracy of your classification product.

Part 12: Project Set up

A. Open a QGIS project

1. Start QGIS by clicking on the **QGIS** shortcut on your desktop

B. Load the land cover and 2008-2010 cloud free composite data

1. Click on the **Add Raster Layer icon**.
2. Navigate to the **C:\Change_detection\Data\Composite\imagery_subset** folder.
3. Select the **2008_LC.tif** and **2008_Sub_9band.tif** rasters. You can hold down the Ctrl key while selecting the rasters with your mouse to select both.

C. Adjust display settings

1. Right click on the **2008_LC.tif** layer in the **Layers Panel**. Go to **Properties > Style** to change the name and color of the classes.
   i. Render type: **Singleband pseudocolor**
   ii. At the bottom of the Band Rendering box:
      (a) Mode: **Equal interval**
      (b) Classes: the number of land cover categories you included in your classification scheme - I had three (forest, water, other)
      (c) Select **Classify**
   iii. Now in the box on the left (with the values, color swatches and labels) you can double click on any of these to edit them.
      (a) Edit the labels to be text values – e.g., change ‘1’ to ‘forest’, etc.
      (b) Double click on each color swatch to change it to a color that matches land cover expectations (e.g., water is usually set to a blue display, etc).
      (c) Select **OK** to apply the updated settings to view your classification.

2. Right click on the **2008_Sub_9band.tif** layer in the **Layers panel**. Go to **Properties > Style**.
   i. Set Render type to **Multiband color**
   ii. Red band to **Band 06**
   iii. Green band to **Band 04**
   iv. Blue band to **Band 03**
   v. Accuracy to **Estimate**
   vi. Click **Load** in the **Load min/max values** section.
   vii. Then click **OK** to save the changes.

Part 13: Sample Design

The sampling design is the protocol for selecting the subset of spatial units (e.g., pixels or segments) that will form the basis of the analysis of area and accuracy. Sample design considerations are discussed in Exercise 10. For this exercise, you will just focus on the basics of how to conduct an accuracy assessment. You will specify a stratified random design, a total sample size of 30 plots, and the allocation of samples to each stratum as indicated in the table below.

	Forest	*Water*	*Other*
n_i	10	10	10

***Note**: This is not a robust sample design, but rather a simplified sample design for the purpose of learning about how to conduct an accuracy assessment.*

A. Select Sample

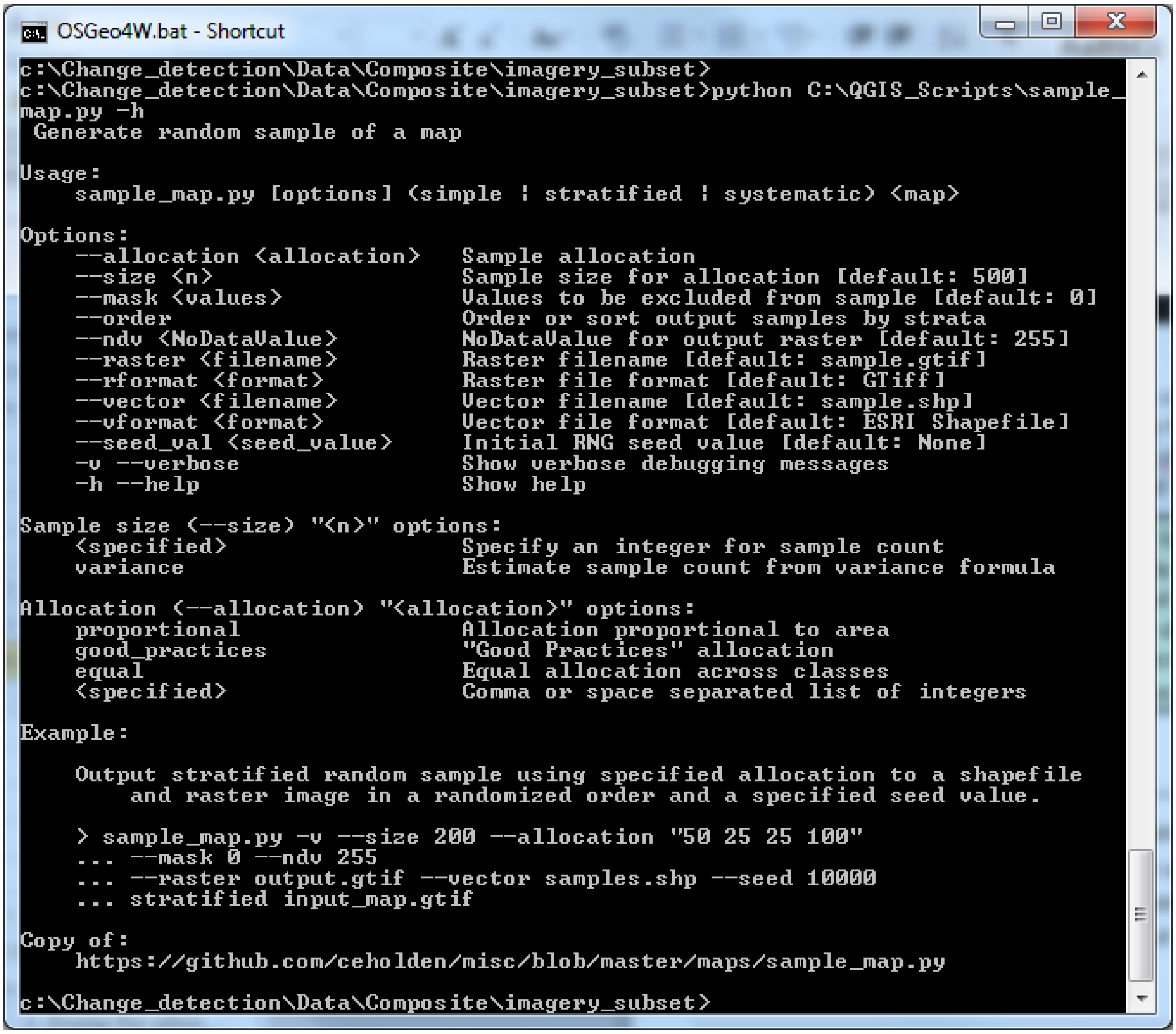
Next the sample needs to be selected. This can be done in several ways, but many software lack good support for selecting a sample. Therefore, you will use a program written by Earth and Environment scientists at Boston University.

1. QGIS does not have built-in tools for drawing samples so you need to make use of two Python scripts, sample_map.py and docopt.py, in the folder you've copied from the USB to the C drive (C:\QGIS_Scripts).
2. First you will open the OSGeo4W command line shell. Go to your desktop and double-click the OSGeo4W.bat – Shortcut.
3. Once the shell opens, navigate to the directory where the land cover classification output is stored.

```
cd C:\Change_detection\Data\Composite\imagery_subset
```

4. Type the command below. This will open a dialog with descriptions about the different options.

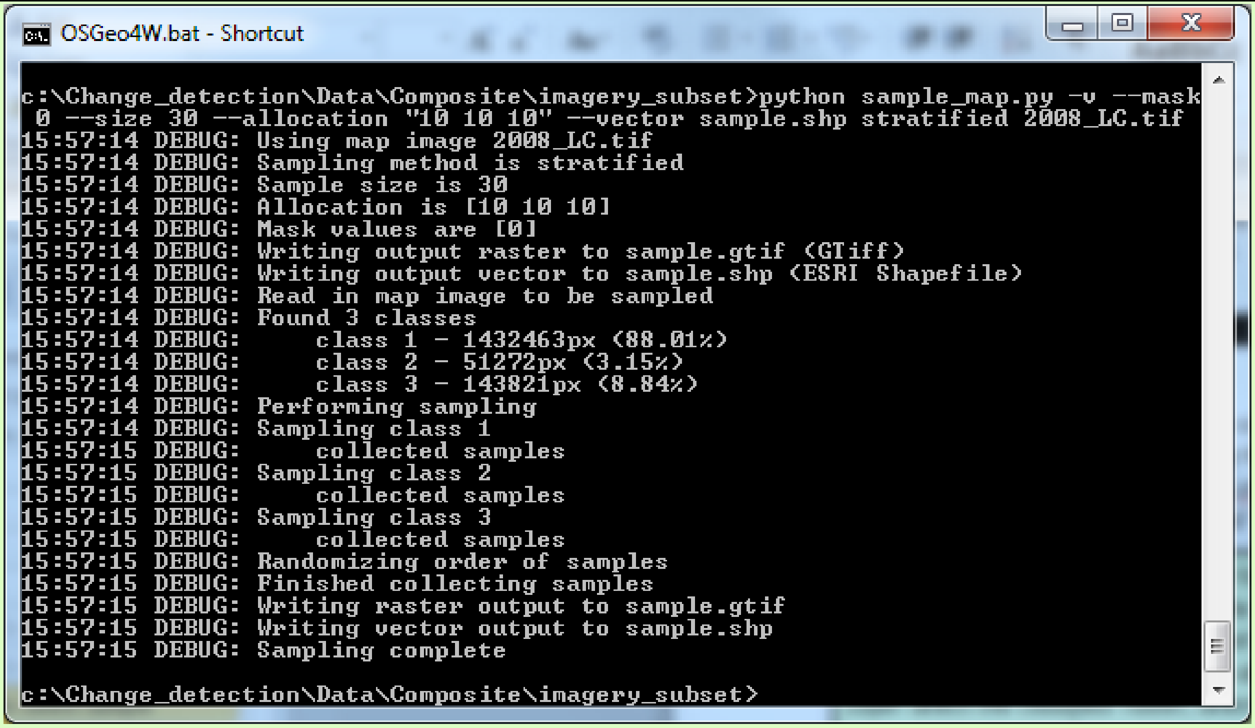
```
python C:\QGIS_Scripts\sample_map.py -h
```



5. To select a stratified random sample, type:

```
python C:\QGIS_Scripts\sample_map.py -v --mask 0 --size 30 --allocation "10 10 10" --vector sample.shp stratified C:\Change_detection\Data\Composite\imagery_subset\2008_LC.tif
```

***Note**: when it's done running, you will see a screen like the one on the next page which indicates the sampling has completed. This has now created a shapefile called **sample.shp** that contains the stratified random sample with 10 points per strata based on the **2008_LC.tif**. Note you might also get a popup that python.exe has stopped working, just click close program it has finished and everything is OK!*



```
OSGeo4W.bat - Shortcut
c:\Change_detection\Data\Composite\imagery_subset>python sample_map.py -v --mask
 0 --size 30 --allocation "10 10 10" --vector sample.shp stratified 2008_LC.tif
15:57:14 DEBUG: Using map image 2008_LC.tif
15:57:14 DEBUG: Sampling method is stratified
15:57:14 DEBUG: Sample size is 30
15:57:14 DEBUG: Allocation is [10 10 10]
15:57:14 DEBUG: Mask values are [0]
15:57:14 DEBUG: Writing output raster to sample.gtif (GTiff)
15:57:14 DEBUG: Writing output vector to sample.shp (ESRI Shapefile)
15:57:14 DEBUG: Read in map image to be sampled
15:57:14 DEBUG: Found 3 classes
15:57:14 DEBUG:     class 1 - 1432463px (88.01%)
15:57:14 DEBUG:     class 2 - 51272px (3.15%)
15:57:14 DEBUG:     class 3 - 143821px (8.84%)
15:57:14 DEBUG: Performing sampling
15:57:14 DEBUG: Sampling class 1
15:57:15 DEBUG:     collected samples
15:57:15 DEBUG: Sampling class 2
15:57:15 DEBUG:     collected samples
15:57:15 DEBUG: Sampling class 3
15:57:15 DEBUG:     collected samples
15:57:15 DEBUG: Randomizing order of samples
15:57:15 DEBUG: Finished collecting samples
15:57:15 DEBUG: Writing raster output to sample.gtif
15:57:15 DEBUG: Writing vector output to sample.shp
15:57:15 DEBUG: Sampling complete

c:\Change_detection\Data\Composite\imagery_subset>
```

Thought question: *If you wanted to create 10 points per strata using another land cover classification output, how would you change the call?*

B. Load the sample into the QGIS project

1. Back in the QGIS software, click the **Add Vector Layer** icon. Browse to the working directory (**C:\Change_detection\Data\Composite\imagery_subset**) and select the shapefile you just created, called **sample.shp**.

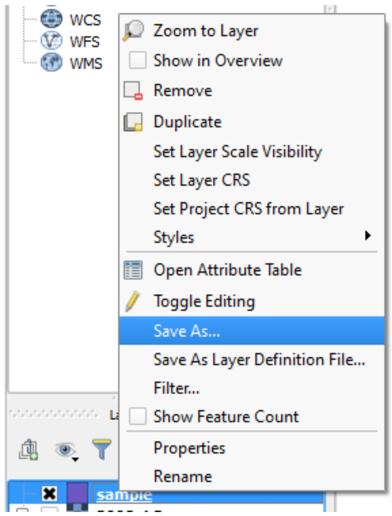
Part 14: Response Design

Once you design the sample and a stratified random sample is selected, it needs to be interpreted using a suitable source of reference data, and you need to decide if the map and reference observations agree. This step is referred to as the response design.

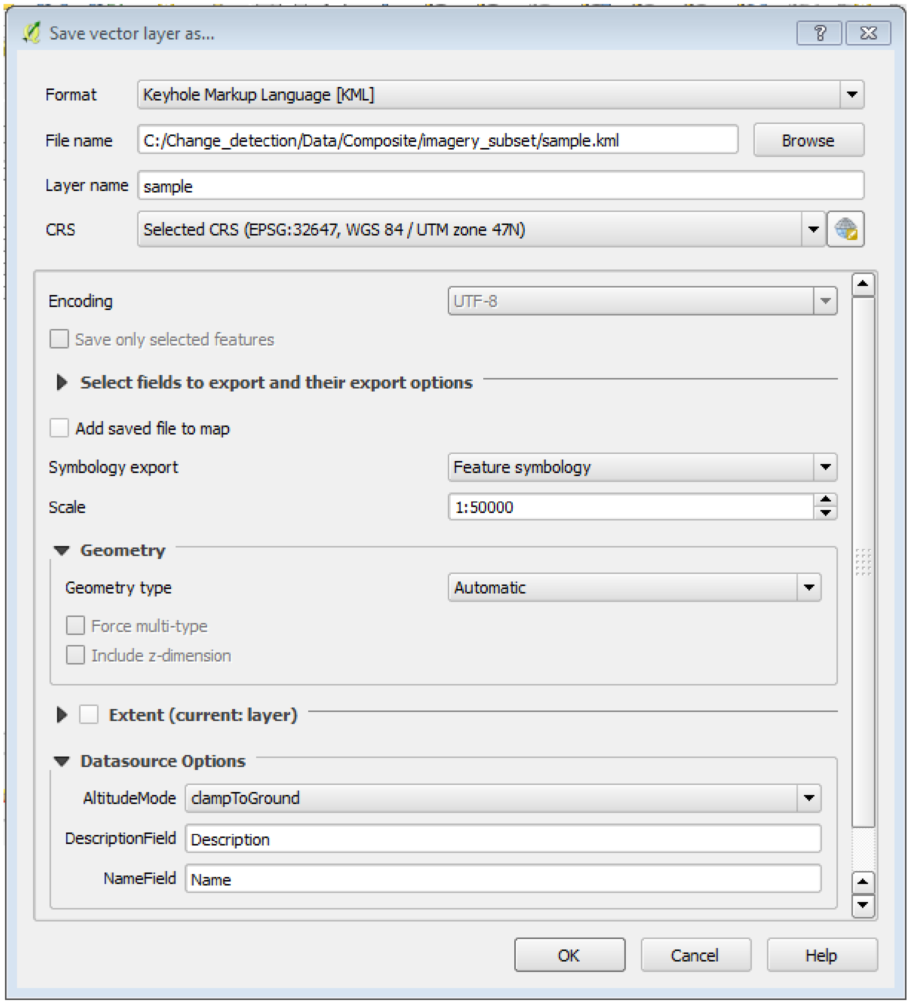
First, you need to identify the reference data sources. Ideally, you would have plots revisited in the field, but this is rarely attainable, so you will need to collect reference observations by careful examination of the sample units in satellite data. The more data you have at your disposal the better. If you have no additional data you can use the Landsat data for collecting reference observations but the process has to be more accurate than the process used to create the map being evaluated. Careful manual examination can be regarded as being a more accurate process than automated classification. In addition to Landsat data, you can use whatever data are available in Google Earth™. As the estimates are based on the sample, it is important that the labels are correct and it is recommended that three interpreters examine each unit independently.

A. Loading reference data (e.g., Landsat and aerial imagery)

1. You have already displayed the reference data in QGIS, i.e., display the data you will use to interpret the sample you just created. This data set is the cloud free composite data you have been working with throughout the online training course.
2. Load any additional data that is available, such as RapidEye, acquired around the same times as the data used to create the map (in this case 2008).
3. You can also load the sample points into Google Earth to access their online imagery library.
    i. Right click on **sample** in the layers panel. Select **Save As**.

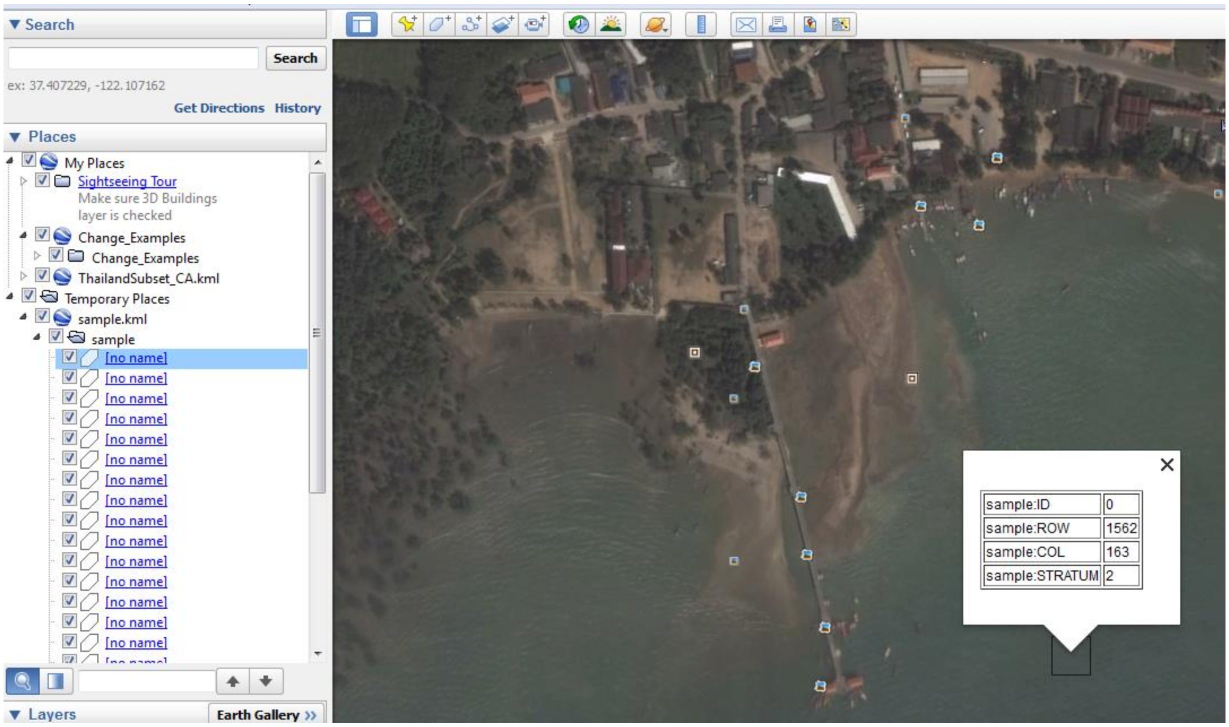


    ii. Use the drop down menu to change the Format to Keyhole Markup Language [KML].
    iii. Click Browse and save the file as sample in the imagery_subset folder.
    iv. Uncheck Add saved file to map.
    v. Change the Symbology export to Feature symbology.
    vi. Select **OK**.



Note: If you don't have Google Earth on your computer, you will need to install it for the next steps. You can download the installation files here: https://www.google.com/earth/.

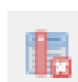
4. In Windows Explorer, navigate to the **imagery_subset** folder and double click on the newly created kml file, **sample.kml**. This will open the data set of plot locations in Google Earth.
   i. In the Places panel on the left hand side of the screen, you can expand the list of samples by clicking on arrows next to Temporary Places > sample.kml > sample.
   ii. Double click on any of the sample features in the list to zoom into that plot location.



B. Interpreting sample

1. Return to QGIS and right-click the **sample** shapefile in the **Layer pane.** Select **Open Attribute Table.**
   i. Then select the **Toggle Editing Mode** (the pencil icon):

   ii. Then the **Delete field** icon (table with a red box with an 'x').

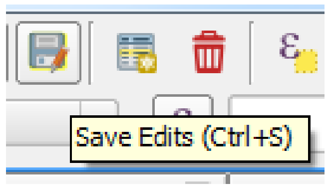


   iii. Highlight and delete the STRATUM column.
2. Click the **New field** button to add a column;

   i. Name it "reference";
   ii. Change length to 3;
   iii. Leave the other options as the default.
3. Now provide a label for each of the units in the sample by manually examining the reference data. Add labels that correspond to the grid codes of the map: for example, if the forest class has the grid code "3" in the map, then provide each sample unit exhibiting forest with the label "3".
   i. Refer to the cloud free composite values and the aerial imagery (from around 2008-2010) in Google Earth to determine what the land cover is at each plot location.

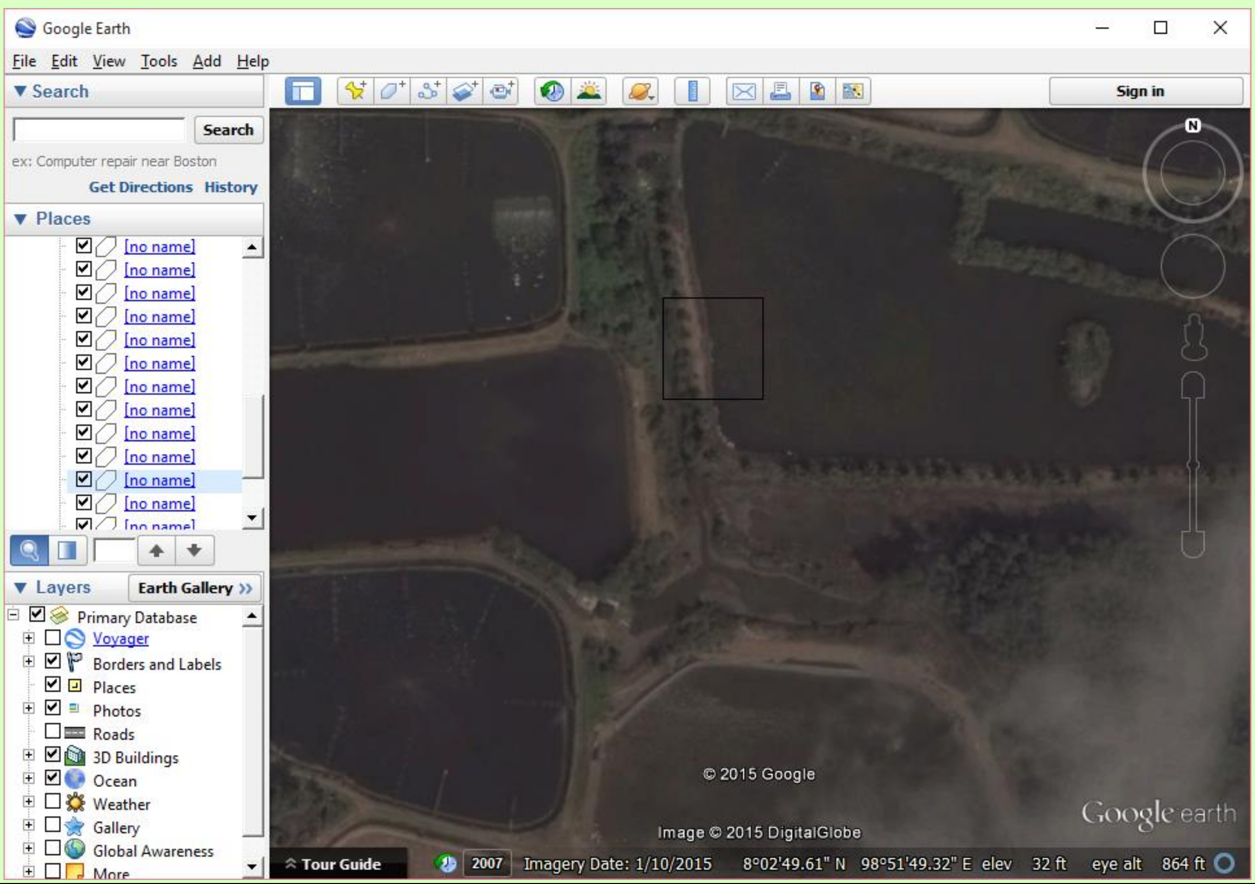
ii. You can click Zoom map to selected row button in the attribute table to jump to the highlighted plot.

iii. Make sure you save the edits to your shapefile regularly.



4. Click the Toggle Editing Mode button again to complete your editing session.
5. Since your final area estimates are based on the interpretation of this sample it is important that the labels are correct – if you can't provide a correct label then delete the unit rather than guessing.

Look at the image below, how would you classify this plot? Make sure that you have clear definitions on how you are determining land cover at each plot.



C. Construct the error matrix

Once the sample has been interpreted, the agreement between map and reference labels needs to decided. This could potentially be a complicated task, but in this case you are using the map classes as strata, which makes the decision straightforward. The agreement is preferably expressed in the form of an error matrix, which is a simple cross-tabulation of the map labels against the reference labels for the sample units. The error matrix organizes the acquired sample data in a way that summarizes key results and aids the quantification of accuracy and area. The main diagonal of the error matrix highlights correct classifications while the off-diagonal elements show omission and commission errors. The cell entries and marginal values of the error matrix are fundamental to both accuracy assessment and area estimation.

1. With each unit having a map label and a reference label you can construct an error matrix. This can be done in various ways but we recommend using a home-made script that executes in the terminal.
2. Return to the OSGeo4W Shell and if it's not still pointing to the imagery_subset folder, navigate to it (**C:\Change_detection\Data\Composite\imagery_subset**).
3. In the Shell, type **python**, the full pathname to the crosstab python script, **-v**, **-a**, the name of the column in the sample shapefile with the reference data, the name of the classified raster, the name of the sample shapefile, and finally the name of the file that you will be generating. See the example below.

```
python C:\QGIS_Scripts\crosstab.py -v -a reference 2008_LC.tif
sample.shp errormatrix.txt
```

Note: If you named the field something other than 'reference' when you added a new column previously in Part A, you will need to edit this line of code accordingly. It is case sensitive, so if you named the field 'Reference' update the casing accordingly.

i. This will create text file that contains the error matrix called "errormatrix.txt" in the **imagery_subset** folder.

Note: *If the script gives you an error regarding varying input shapes, check to make sure your raster file has the same number of classes as the shapefile. If the edges of your map contain 0s, you do not have any 0 values in your shapefile, and 0 is not the no data value of your raster, the script will not work.*

To fix this create a new raster with 0 as the no data value by going to Raster -> Conversion -> Translate. For Input Layer select the classified raster, select a name for the Output file, and for No data put 0 (or whatever you want to declare the no data value).

Part 15: Analysis

With the construction of an error matrix the estimation becomes straightforward. At the heart of the analysis is the implementation of an unbiased area estimator. Different estimators can be implemented but with a sample stratified by discrete map classes, stratified estimation has proven useful. A stratified estimator of area includes the area of omission but excludes the error of commission, and it is easily implemented from the data in the error matrix. Using the error matrix one can also estimate the accuracy of the map and the map classes.

Note that stratified estimation can be used with simple or systematic random samples too.

The error matrix (with the mapped areas of each map category) contains all the information needed to perform the analysis which includes stratified estimation of area and confidence intervals. Again, this can be done various ways, but we recommend implementation in a spreadsheet program to provide the user with an understanding of the estimation procedure.

A. Open the error matrix in a spreadsheet software

1. Open Microsoft Excel.
2. In Excel, go to File > Open > browse.
   i. Change the type from All Excel Files (*.xlsx…) to All Files (*.*).
   ii. Open the text file created in previous steps.
   iii. Choose Delimited. Click Next.
   iv. Select Comma as the Delimiter. Click Next.
   v. Then Finish.
   vi. The screen should like below:

A1

	A	B	C	D	E
1		Map-Class_1	Map-Class_2	Map-Class_3	
2	Ref-Class_1	8	4	5	
3	Ref-Class_2	1	6	0	
4	Ref-Class_3	1	0	5	
5					
6					

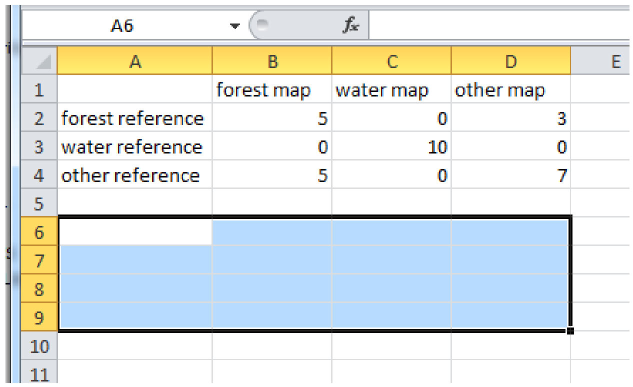
3. First, rename the columns and rows so that they are more intuitive to work with. In this example. Class 1 is forest, class 2 is water, and class 3 is other. Rename the table headers accordingly. See example below.
4. Save the file as an Excel Spreadsheet
   i. Click **File > Save As**
   ii. Change type **From Text (Tab delimited) (*.txt)** to Excel Workbook (*.xlsx)
   iii. Click **Save**

C11

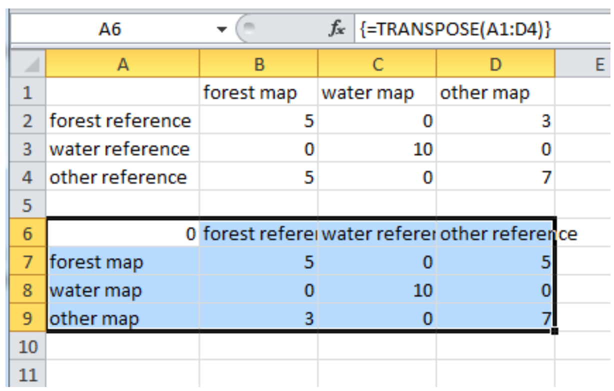
	A	B	C	D	E
1		forest map	water map	other map	
2	forest reference	8	4	5	
3	water reference	1	6	0	
4	other reference	1	0	5	
5					
6					

B. Below the error matrix output, invert the matrix

1. Below the sample error matrix information – duplicate the matrix, but invert the information (such that the reference information is stored as columns, and the map information is stored as rows).
   i. Starting with cell A6, click and drag your mouse down to cell D9.



   ii. In the function bar, type **=transpose(A1:D4)** and hold down the **Ctrl + Shift + Enter keys**.
   iii. You will see that cells A6:D9 are now populated with the information from your error matrix, but the rows and columns have been transposed.



***Note:** You may notice curly brackets in the screen capture above – these were populated automatically after pressing the Ctrl + Shift + Enter keys. Do not type these in yourself.*

C. Adding labels

1. Add **Map Total** as a column label on the right hand side of the lower matrix.
2. Add **Reference Total** as a row header at the bottom of the map labels.

	A	B	C	D	E
1		forest map	water map	other map	
2	forest reference	5	0	3	
3	water reference	0	10	0	
4	other reference	5	0	7	
5					
6		0	forest reference	water reference	other reference	**Map total**
7	forest map	5	0	5	
8	water map	0	10	0	
9	other map	3	0	7	
10	**Reference Total**				
11					

D. Calculate reference and map sums

1. In the cell below **Map Total** (E7), total the number of instances where the forest category was mapped by entering the following:

```
=sum(B7:D7)
```

SUM =SUM(B7:D7)

SUM(**number1**, [number2], ...)

	A	B	C	D	E	F
1		forest map	water map	other map		
2	forest reference	5	0	3		
3	water reference	0	10	0		
4	other reference	5	0	7		
5						
6	0	forest reference	water reference	other reference	**Map total**	
7	forest map	5	0	5	=SUM(B7:D	
8	water map	0	10	0		
9	other map	3	0	7		
10	**Reference Total**					
11						

2. Repeat for the other mapped land cover classes by highlighting cell E7, clicking the lower right-hand corner of the cell, and dragging it down to cell E9.

E7 =SUM(B7:D7)

	A	B	C	D	E	F
1		forest map	water map	other map		
2	forest reference	5	0	3		
3	water reference	0	10	0		
4	other reference	5	0	7		
5						
6	0	forest reference	water reference	other reference	**Map total**	
7	forest map	5	0	5	10	
8	water map	0	10	0		
9	other map	3	0	7		
10	**Reference Total**					
11						

3. Repeat the summing process for the **Reference Total** counts (see example below).

SUM | =SUM(D7:D9)

SUM(number1, [number2], ...)

	A	B	C	D	E
1		forest map	water map	other map	
2	forest reference	5	0	3	
3	water reference	0	10	0	
4	other reference	5	0	7	
5					
6	0	forest reference	water reference	other reference	**Map total**
7	forest map	5	0	5	10
8	water map	0	10	0	10
9	other map	3	0	7	10
10	**Reference Total**	8	10	=SUM(D7:D9)	
11					

E. Calculate the Accuracy of the map

Finally, you can estimate the accuracy of the map. Three different accuracy measures are of interest:

- **overall accuracy** which is simply the sum of the diagonals in the error matrix of estimated area proportions;
- **user's accuracy** which for a map category *i* is given by $\widehat{U}_i = \hat{p}_{ii} \div \hat{p}_{i\cdot}$ and
- **producer's accuracy** for map category *j* given by $\widehat{P}_i = \hat{p}_{jj} \div \hat{p}_{\cdot j}$ where $\hat{p}_{i\cdot}$ and $\hat{p}_{\cdot j}$ are the row and columns totals respectively.

1. To the right of the **Map Total** label, add **User's Accuracy**.
2. In the cell below the Reference **Total** label, add **Producer's Accuracy**.

F18

	A	B	C	D	E	F
1		forest map	water map	other map		
2	forest reference	8	4	5		
3	water reference	1	6	0		
4	other reference	1	0	5		
5						
6		forest reference	water reference	other reference	**Map Total**	**User's Accuracy**
7	forest map	5	0	5	**10**	
8	water map	0	10	0	**10**	
9	other map	3	0	7	**10**	
10	**Reference Total**	**8**	**10**	**12**		
11	**Producer's Accuracy**					

3. To calculate the **User's Accuracy**, you will divide the agreement (the diagonal along the matrix where, for example, forest reference and forest map intersect) by the **Map Total**. In the image below, the agreement cells are highlighted in green.
   i. In cell F7, type in the following:

```
=B7/E7
```

   ii. This equation will calculate the total number of forest samples that were mapped correctly (in this example, there are 5 correctly mapped samples) and divide by the total number of forest samples in the map (you placed 10 samples within each class, so there are a total of 10 forest samples).

iii. Repeat this process for each mapped category.

SUM =D9/E9

	A	B	C	D	E	F
1		forest map	water map	other map		
2	forest reference	8	4	5		
3	water reference	1	6	0		
4	other reference	1	0	5		
5						
6		forest reference	water reference	other reference	**Map Total**	**User's Accuracy**
7	forest map	5	0	5	**10**	0.5
8	water map	0	10	0	**10**	1.0
9	other map	3	0	7	**10**	=D9/E9
10	**Reference Total**	**8**	**10**	**12**		
11	**Producer's Accuracy**					

Note*: this means that the user's accuracy for the forest land cover is 50%; so the error of commission is 50%. The user's accuracy for the water land cover is 100%; so the error of commission is 0% - this means that according to these reference data, the map creator did not map water in any non-water classes.*

What is the user's accuracy and error of commission for the other class in this example (7/10)?

4. To calculate the **Producer's Accuracy**, you will divide the agreement (the diagonal along the matrix where, for example, forest reference and forest map intersect) by the **Reference Total**. In the image below, the agreement cells are highlighted in green.

i. In cell B11, type in the following:

```
=B7/B10
```

ii. This equation will calculate the total number of forest samples that were mapped correctly (in this example, there are 5 correctly mapped samples) and divide by the total number of forest samples in the reference data (in this example, 8 of the total samples were interpreted as forest).

iii. Repeat this process for each mapped category.

SUM =D9/D10

	A	B	C	D	E	F
1		forest map	water map	other map		
2	forest reference	8	4	5		
3	water reference	1	6	0		
4	other reference	1	0	5		
5						
6		forest reference	water reference	other reference	**Map Total**	**User's Accuracy**
7	forest map	5	0	5	**10**	0.5
8	water map	0	10	0	**10**	1.0
9	other map	3	0	7	**10**	0.7
10	**Reference Total**	**8**	**10**	**12**		
11	**Producer's Accuracy**	0.63	1.00	=D9/D10		

***Note**: this means that the producer's accuracy for the forest land cover is 62.5%; so the error of omission is 37.5% - that's kind of high! The producer's accuracy for the water land cover is 100%; so the error of omission is 0% - this means that of all of the water reference samples were successfully mapped as water, and none were omitted from the map.*

What is the producer's accuracy and of omission for the other class in this example (7/12)?

F. Calculate Overall Accuracy

1. To calculate the **Overall Accuracy**, first you need to determine the area of land that each of the strata cover. You will determine the areas of each map category using a gdal tool called from the OSGeo4W command line shell. Return to the command line shell and type in the following:

```
gdalinfo -hist 2008_LC.tif
```

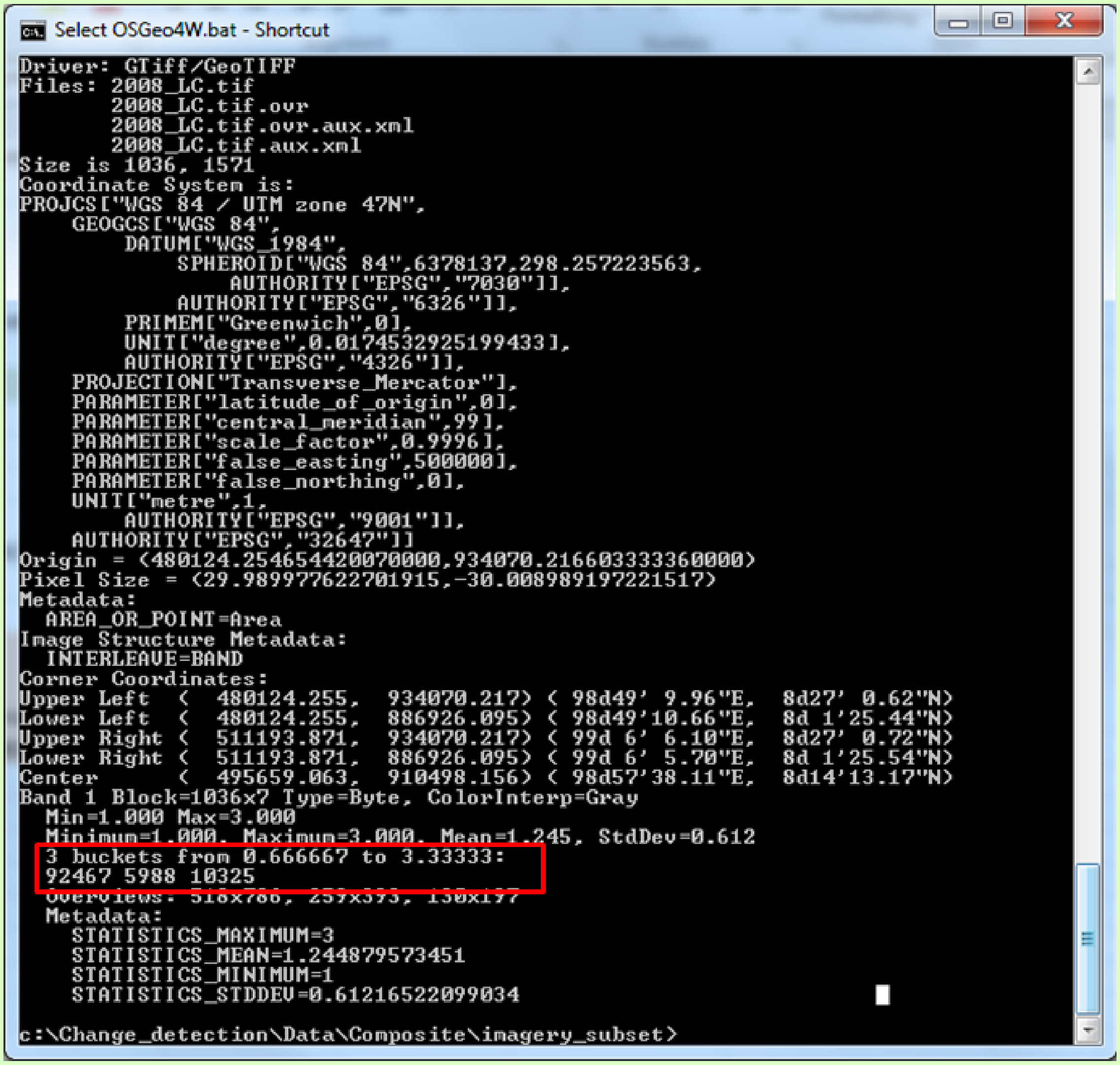
***Note**: The return looks a little messy. The information you are interested in is the summary below the bucket information line (highlighted in red in the image below). It gives you the count of the number of pixels in each 'histogram bucket'. In this example, the histogram is broken up into 3 buckets – starting at 0.67 to 3.3.*

This indicates that there are:

- 92,467 pixels with a value between ~0.6 to 1.5 (forest land cover);

- 5,988 pixels with a raster value between ~1.6 and 2.5 (water land cover);

- 10,325 pixels with a value between ~2.6 and 3.3 (other land cover);



2. In the column to the right of the User's Accuracy, type in the column header **Pixels**.
3. Use the information from the gdalinfo call in the OSGeo4W command line shell to update the pixel count of each land cover category.

	A	B	C	D	E	F	G
1		forest map	water map	other map			
2	forest reference	8	4	5			
3	water reference	1	6	0			
4	other reference	1	0	5			
5							
6		forest reference	water reference	other reference	**Map Total**	**User's Accuracy**	**Pixels**
7	forest map	5	0	5	**10**	0.5	92467
8	water map	0	10	0	**10**	1.0	5988
9	other map	3	0	7	**10**	0.7	10325
10	**Reference Total**	**8**	**10**	**12**			
11	**Producer's Accuracy**	0.63	1.00	0.58			

4. To the right of the **Pixels** column header, type in **Weights**.
5. Now calculate the weights by dividing the pixel value by the sum of pixel counts for all three land cover categories.

```
=G7/SUM($G$7:$G$9)
```

6. Copy and paste the equation for the other two land cover categories.

SUM | =G9/SUM($G$7:$G$9)

	A	B	C	D	E	F	G	H
1		forest map	water map	other map				
2	forest reference	8	4	5				
3	water reference	1	6	0				
4	other reference	1	0	5				
5								
6		forest reference	water reference	other reference	**Map Total**	**User's Accuracy**	**Pixels**	**Weights**
7	forest map	5	0	5	**10**	0.5	92467	0.850037
8	water map	0	10	0	**10**	1.0	5988	0.055047
9	other map	3	0	7	**10**	0.7	10325	=G9/SUM(
10	**Reference Total**	**8**	**10**	**12**				
11	**Producer's Accuracy**	0.63	1.00	0.58				

Next you will calculate the overall accuracy after taking into account the area weights. You will learn more about the equation, specifically the use of weighting your overall estimates by area, in Exercise 10

7. Underneath your estimate of producer's accuracy, type **Overall Accuracy, weighted by area**.
8. In the cell to the right of the new label, type in the following equation.

```
=H7*B7/E7+H8*C8/E8+H9*D9/E9
```

	A	B	C	D	E	F	G	H
1		forest map	water map	other map				
2	forest reference	8	4	5				
3	water reference	1	6	0				
4	other reference	1	0	5				
5								
6		forest reference	water reference	other reference	**Map Total**	**User's Accuracy**	**Pixels**	**Weight**
7	forest map	5	0	5	**10**	0.5	92467	0.8500
8	water map	0	10	0	**10**	1.0	5988	0.0550
9	other map	3	0	7	**10**	0.7	10325	0.0949
10	**Reference Total**	**8**	**10**	**12**				
11	**Producer's Accuracy**	0.63	1.00	0.58				
12								
13								
14								
15	**Overall Accuracy, weighted by area**	0.546506711						

***Note**: In this example, the overall accuracy is fairly low – it's 54.7%. This means that for any given pixel there is a 54.7% chance that it has been mapped correctly. Or you could reverse the focus to the likelihood of error – in which case for any given pixel there is a 45.3% chance that it is mapped incorrectly.*

Congratulations! You have successfully completed this exercise.