African Drone and Data Academy (ADDA)

Module Title: **GIS and Data Analytics**

Lab Title: **Supervised Image Classification in QGIS**

Lab content author: Emmanuel Chinkaka echinkaka@must.ac.mw)

Land cover/Land use Mapping

Introduction

Digital image classification techniques are used to group pixels with similar values in several image bands into land cover classes. Common approaches are unsupervised, supervised and object-based. This lab will focus on the supervised approach. In supervised classification, the user selects representative samples for each land cover class in the digital image. These sample land cover classes are called “training sites”. The image classification software uses the training sites to identify the land cover classes in the entire image. The classification of land cover is based on the spectral signature defined classification on a Landsat image. In this lab, you will use the QGIS Semi-Automatic Classification Plugin (SCP) version 6.0.0 (Greenbelt) to preprocess two satellite image datasets for classification.

Developed by Luca Congedo, the Semi-Automatic Classification Plugin (SCP) allows for the supervised classification of remote sensing images, providing tools for the download, the pre-processing and post-processing of images.

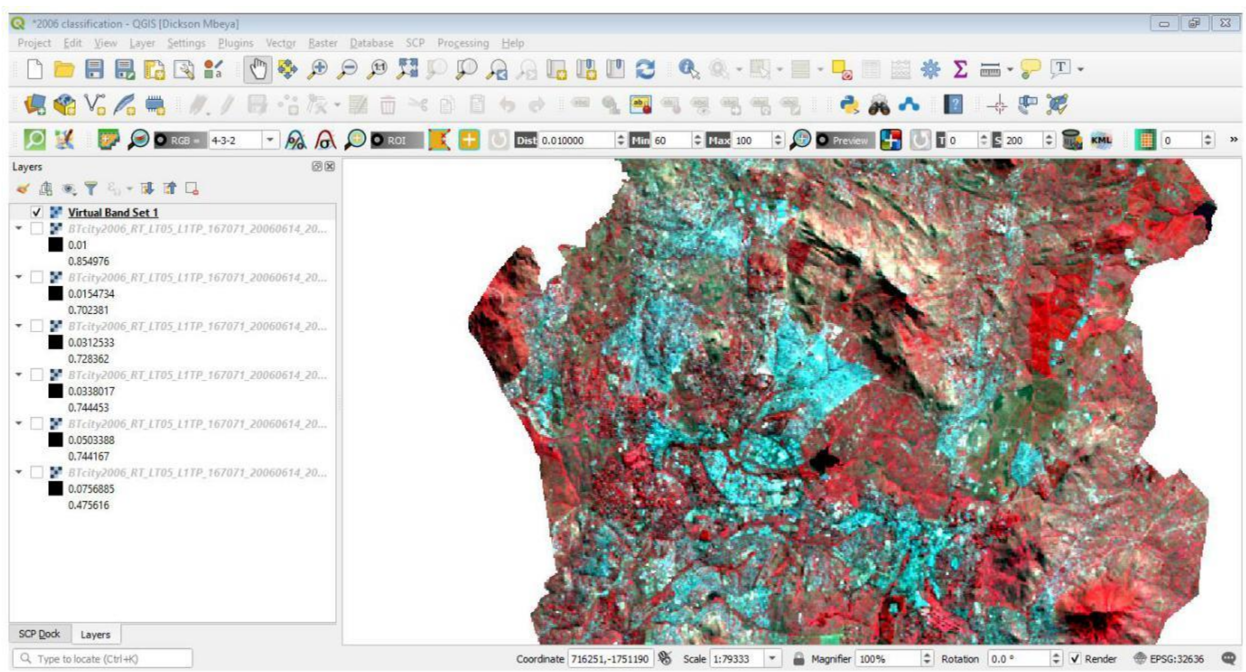
Objectives:

- To learn how to install QGIS Semi-Automatic Plugin and how to create a multispectral band set
- Identify land cover classes
- Generate regions of interest (ROIs)
- Conduct a supervised classification on a Landsat 5 image

Section 1: Explore your data and install SCP plugin

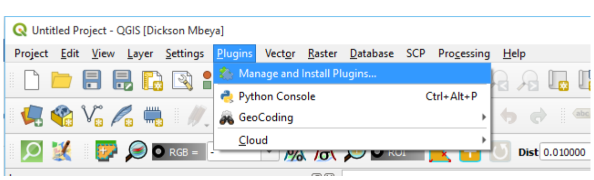
You have been provided with a data folder in which inside there is a folder named **BT_CITY_2006 Images.**

Open the **BT_CITY_2006** folder. Inside that folder open the QGIS Project with title: Supervised Classification Lab 2006 Image. The project containing preprocessed images. The project should show something like this.

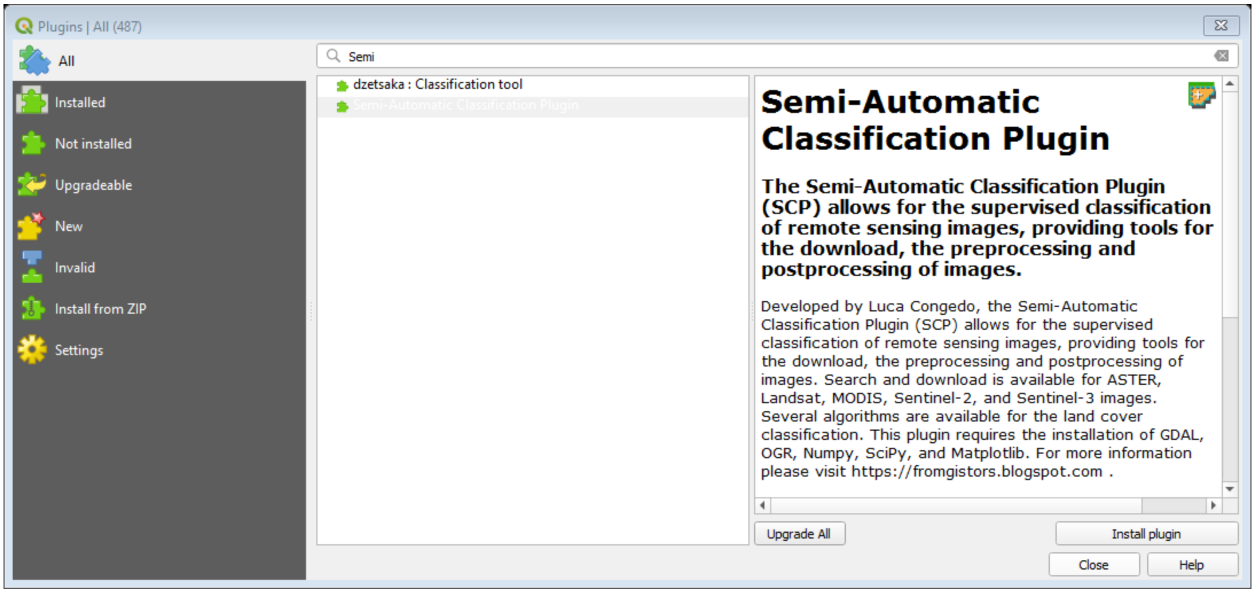


You will classify land cover for this image. If you didn't create a band set in the previous exercise, create in by following the procedures outline in the previous exercise.

1. From the main menu, select `Plugins > Manage and Install Plugins…`



2. In the search filed of the Plugins manager, search for Semi-automatic classification plugin and click the `Install plugin` button:

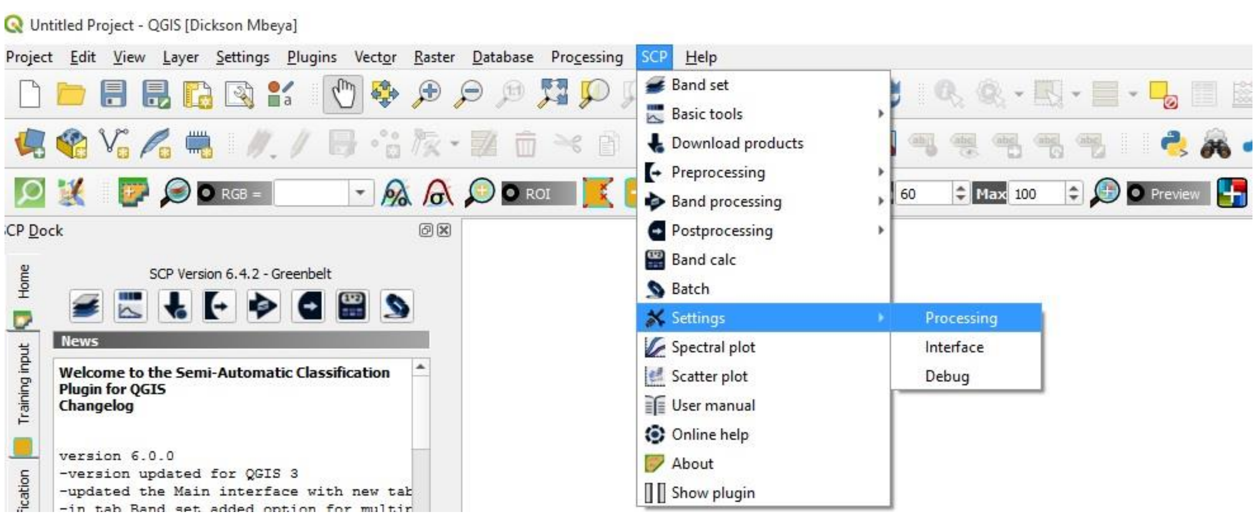


The SCP should be automatically activated; however, be sure that the Semi-Automatic Classification Plugin is checked in the menu Installed

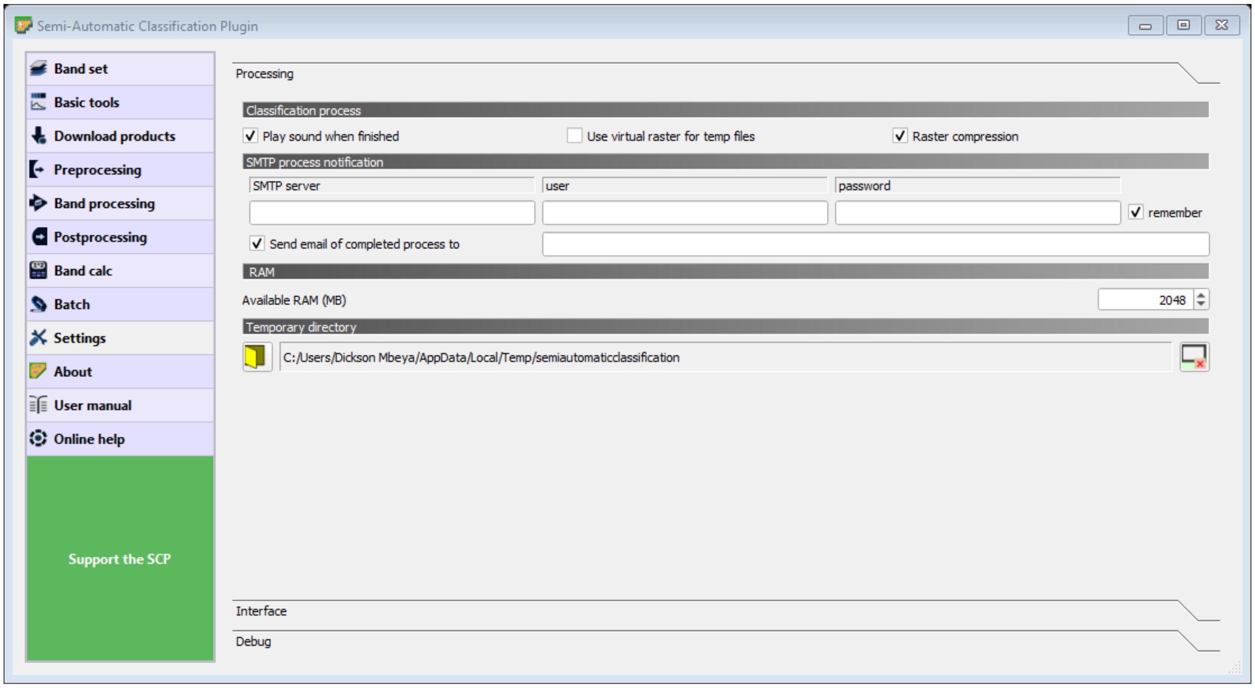
Section 2: Configuration of the SCP plugin

After installation of the plugin, a dock and a toolbar should be added to QGIS. SCP menu is also available in the Menu Bar of QGIS.
The configuration of available RAM is recommended in order to reduce the processing time

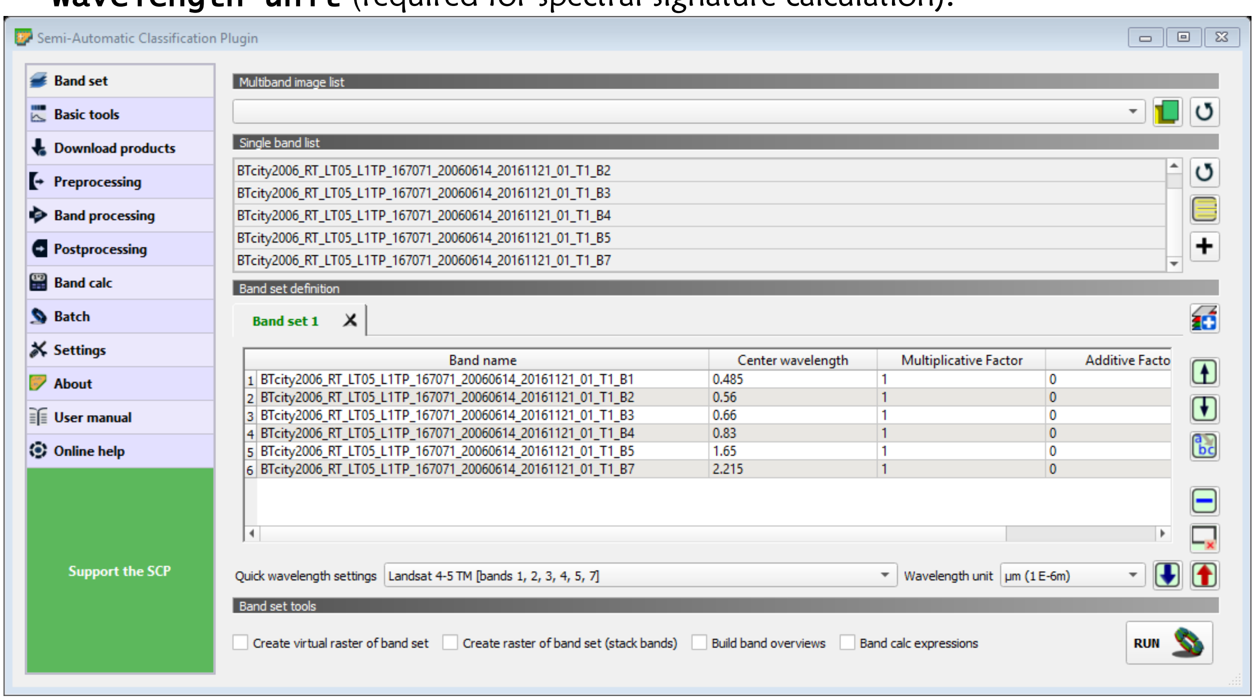
1. From the SCP menu, select Settings `>Processing`

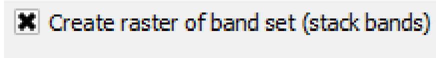


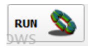
2. In the settings, set the available RAM (MB) to a value that should be half of the system RAM. For instance, if your system has 2GB RAM, set the value to 1024MB.



3. You need to define and create a permanent **Band set** which is the input image for SCP. Open the `tab` **`Band set`** by clicking the button in the SCP menu
4. Add the bands to the Band set
5. In the table Band set definition, order the band names in an ascending order by clicking the  button.
6. Select the **`Landsat 4-5 TM`** from the list **`Quick wavelength settings`** in order to set automatically the **`Center wavelength`** of each band and the **`Wavelength unit`** (required for spectral signature calculation).

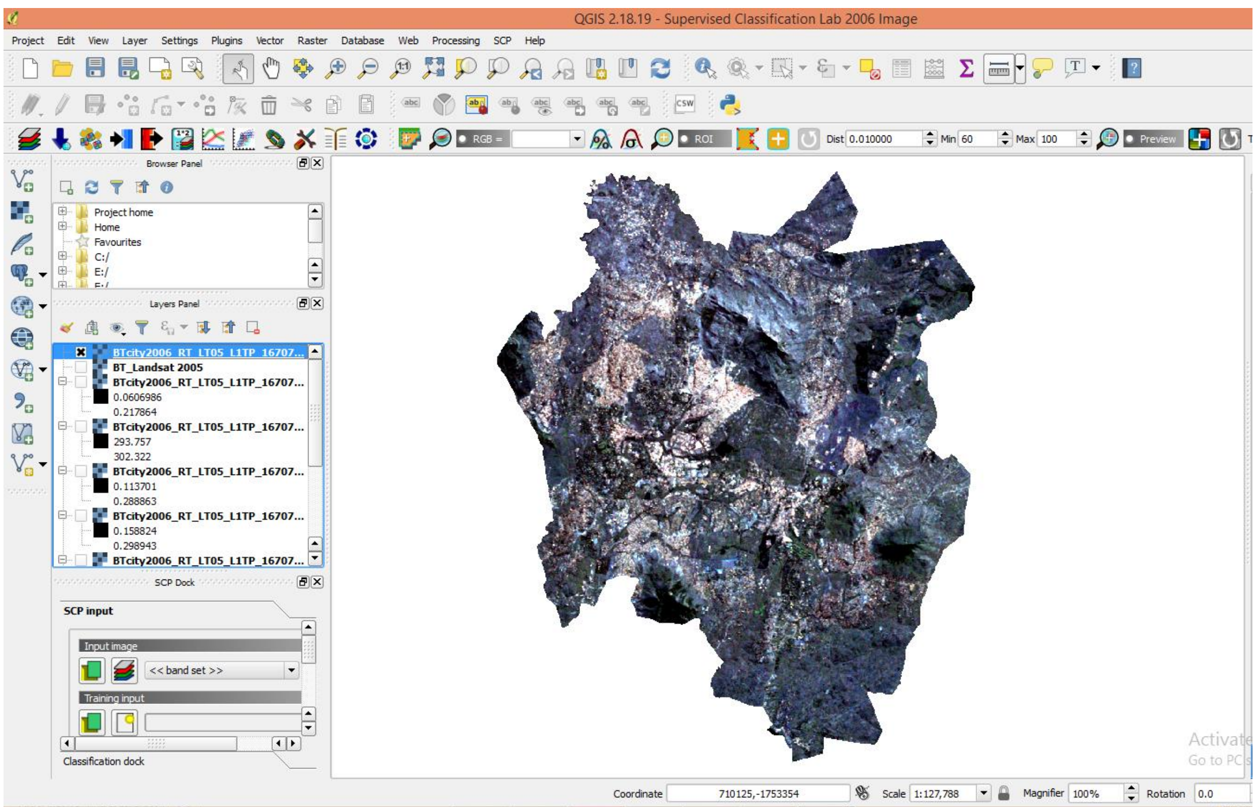


7. Enable the **`Create raster of the band set (stack bands)`** option under Band set tool .

8. Click on Run button 

9. When you here a sound pop, then your processing is done. You will see a new layer on top of the layers list. That the layer you need to do the land cover mapping and classification. You need to have a map displayed as below.

10. Hide all the layers in the QGIS Layers panel except the virtual raster to speed up visualization.
11. Close the Band set.



Section 2: Creating the Training Input File

The training input file (.scp) is a file required for storing ROIs (Training areas) and spectral signatures. This file is created according to the characteristics of the active band set.

1. In the SCP dock, select the tab `Training input` and click the button  to create the Training input. Name the Training input `2006training.scp`. The path of the file is displayed and a vector is added to QGIS Layers panel with the name `2006training`.

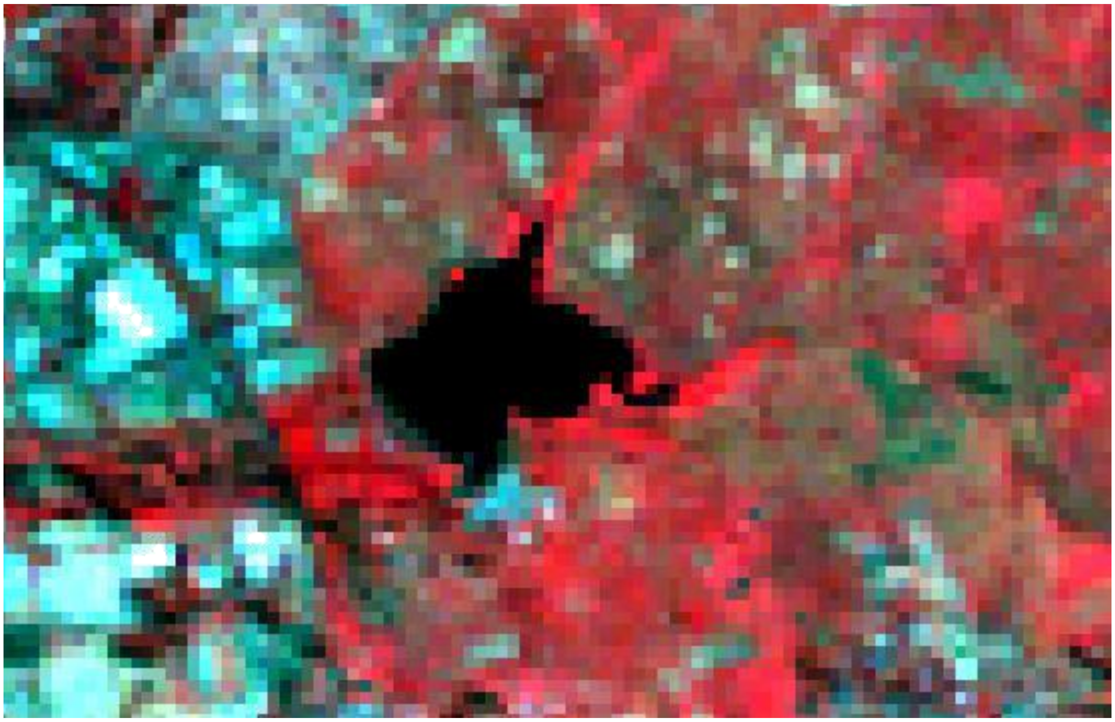
You will create ROIs defining the Classes and Macroclasses. Each ROI is defined by a Class ID (C ID) and each ROI is assigned to a land class through a Macroclass ID (MC ID). Separating spectral signatures of different materials helps in achieving good lasification results. You will create several ROIs for each macroclass by setting the same MC ID, but assigning a different C ID to every ROI. The macro classes to be used in this classification are as shown in the table below:

Macroclass Name	Macroclass ID
Water	1
Built up	2
Vegetation	3
Bare land/ bare soil/ low vegetation	4

ROIs can be created by manually drawing a polygon or with an automatic region growing algorithm. The following general rules must be followed when creating training sites:

- Select as many training sites per class as possible.
- Select training sites throughout the entire image, not just one area.
- Training site selection must be of spectrally homogenous areas (as much as possible).
- Training sites must be as large as possible.

2. Zoom in the map over the dark area in the image which is the Mudi dam, a water body.

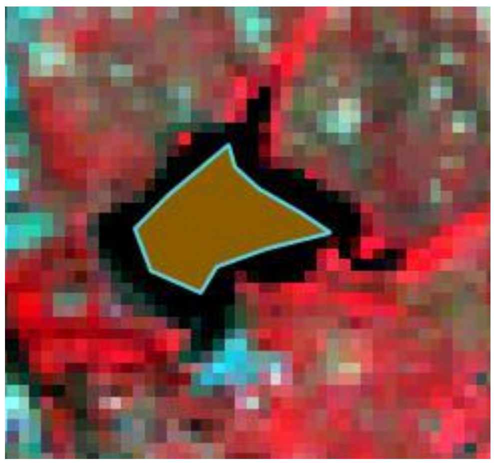


3. Click on the button  in the working toolbar to manually create a ROI inside the dark area.

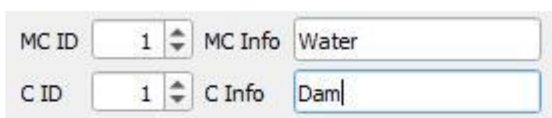
After clicking the button in the working toolbar you should notice that the cursor in the map displays a value changing over the image. This is the NDVI value of the pixel beneath the cursor. NDVI is displayed because the function Display is checked in **Training input.**

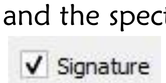
The NDVI value can be useful for identifying spectrally pure pixels, in fact vegetation has higher NDVI values than soil. Very low values of NDVI (0.1 and below) correspond to barren areas of rock, sand or snow. Moderate values represent shrub and grassland (0.2 to 0.3). High values indicate thick vegetation (0.6 to 0.8)

4. Left click on the map to define the ROI vertices and right click to define the last vertex closing the polygon. An orange semi-transparent polygon is displayed over the image which is a temporary polygon:



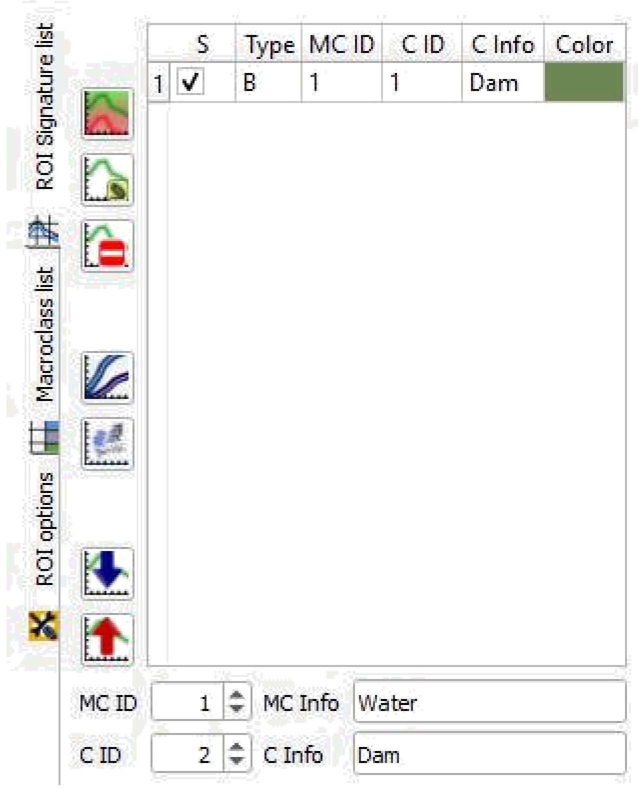
5. Open the `Training input` to define the Classes and Macroclasses.
6. In the ROI signature list, set `MC ID = 1` and `MC Info = Water`. Also set `C ID = 1` and `C Info = Dam`:



7. Click  to save the ROI in the `Training input`. After a few seconds, the ROI is listed in the `ROI Signature list` and the spectral signature is calculated because the Signature option was checked: 

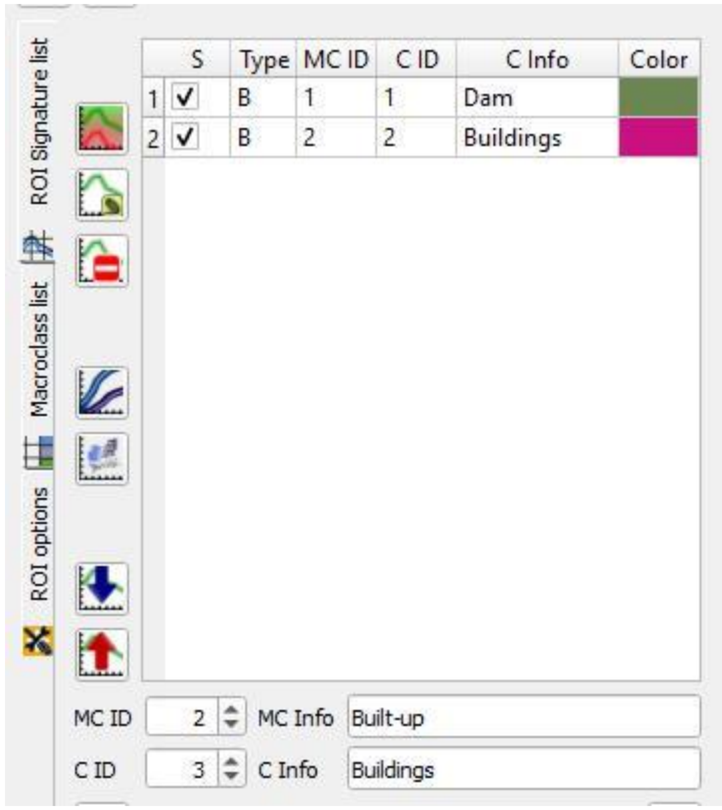
.

8. The `C ID` in `ROI Signature list` is automatically increased by 1. Saved ROI is displayed as a dark polygon in the map and the temporary ROI is removed. You can also notice that in the ROI Signature list, that the `Type` is `B`, implying that the ROI spectral signature was calculated and saved in the Training input.

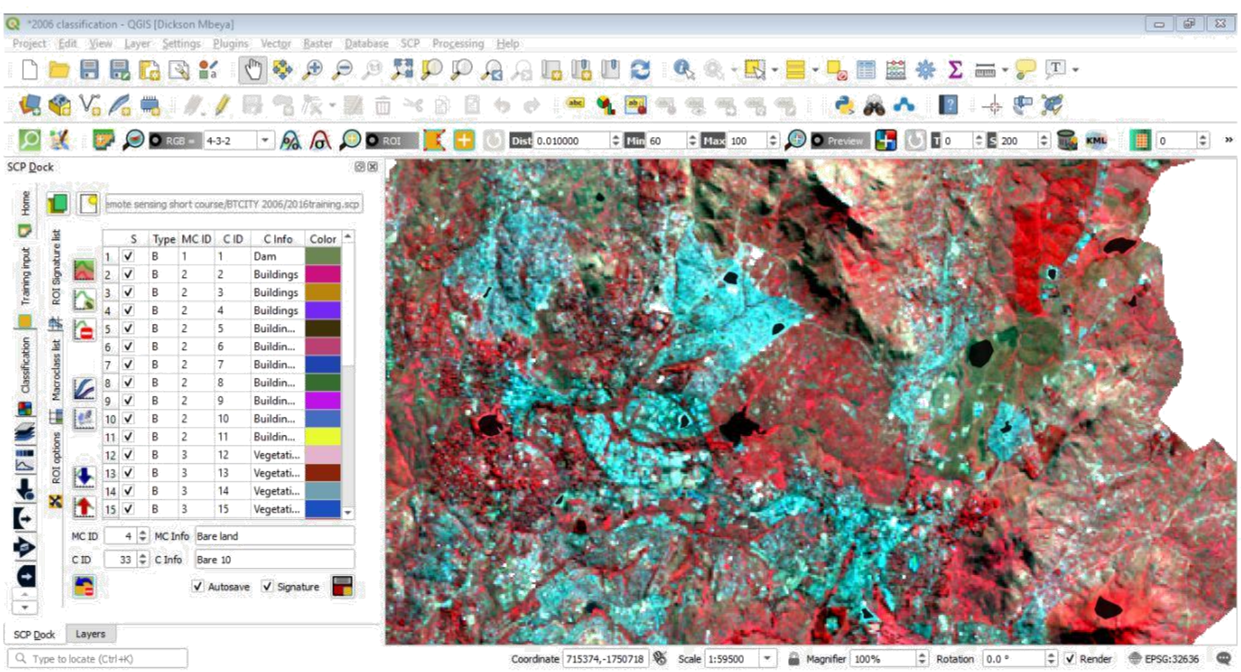


9. Zoom in to a built- up area in the map and create a second ROI for built-up class by following the same procedures 3 to 8.

10. In the **ROI Signature list**, Set **MC ID = 2** and **MC Info = Built-up**; also set **C ID = 2** and **C Info = Buildings**
11. Save the ROI in the Training input



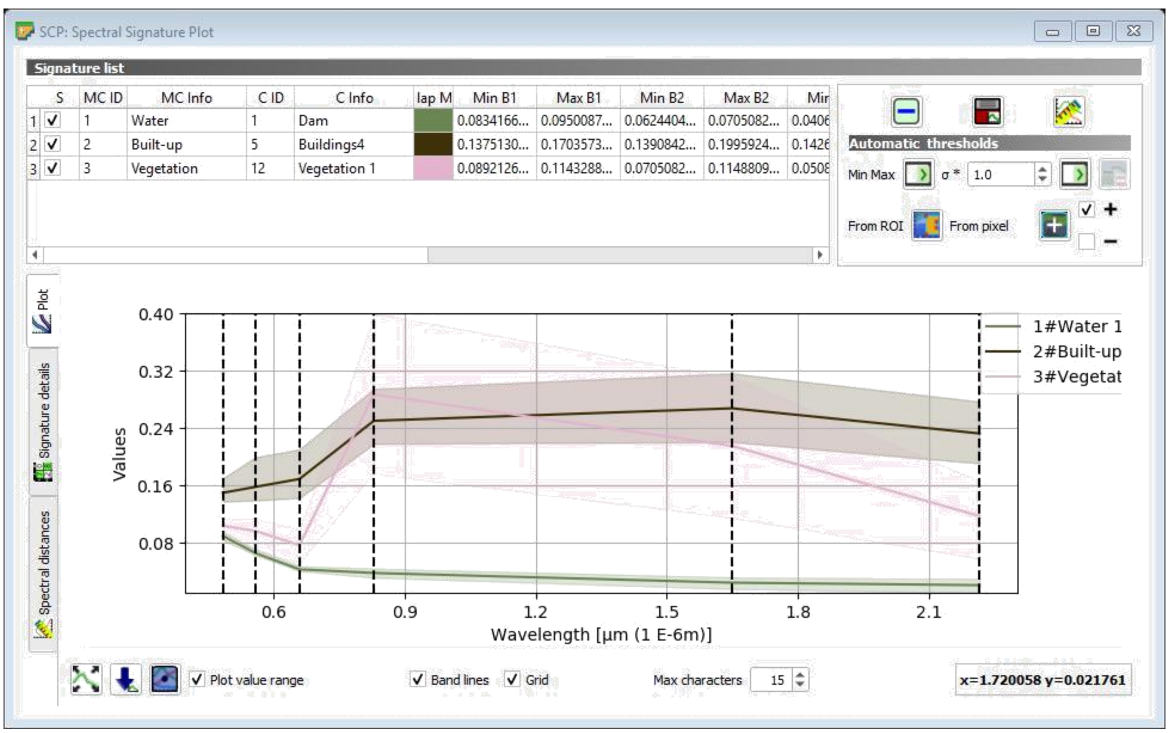
12. Create as many ROIs as possible for each Macroclass. Always remember to assign the right MC ID, MC Info, C ID and C Info. Change frequently the Color Composite in order to clearly identify materials at the ground. Use the mouse wheel on the list **RGD=** of the working toolbar for changing the color composites rapidly. You can also apply image stretching by using the buttons  and  for better displaying the input image.



Section 3; Assessing the Spectral Signatures

Spectral signatures re used by classification algorithms for labelling image pixels. Different materials may have similar spectral signatures such as built-up and bare soil. If spectral signatures used for classification are too similar, pixels could be misclassified because the algorithm is unable to discriminate correctly those signatures. Thus, it is useful to assess the spectral distance of signatures to find similar signatures that must be removed. You can assess spectral signature by displaying a signature plot.

1. Highlight three different spectral signatures with click in the table Plot Signature list and click the button  . The Spectral Signature Plot is displayed in a new window:



In the plot, you can see the line of each signature and the spectral range (minimum and maximum) of each band (the semi-transparent colored like the signature line).

2. Move and zoom inside the plot to see if signatures are similar (very close).
3. Double click on the color in the plot signature list to change the line color in the plot.

You can also calculate the spectral distances of signatures. Spectral distance is useful to evaluate spectral distance or (or separability) between training signatures or pixels in order to assess if different classes that are too similar could cause classification errors. In SCP, the following algorithms are used to assess similarity of spectral signatures:

Jeffries-Matusita Distance:

It calculates the separability of a pair of probability distributions. It can be particularly meaningful for evaluating the results of Maximum Likelihood classifications. The Jeffries-Matusita Distance is asymptotic to 2 when signatures are completely different and tends to 0 when signatures are identical

Spectral angle:

This is the most appropriate for assessing the Spectral Angle Mapping algorithm classifications. Spectral angle goes from 0 when signatures are identical to 90 when signatures are completely different

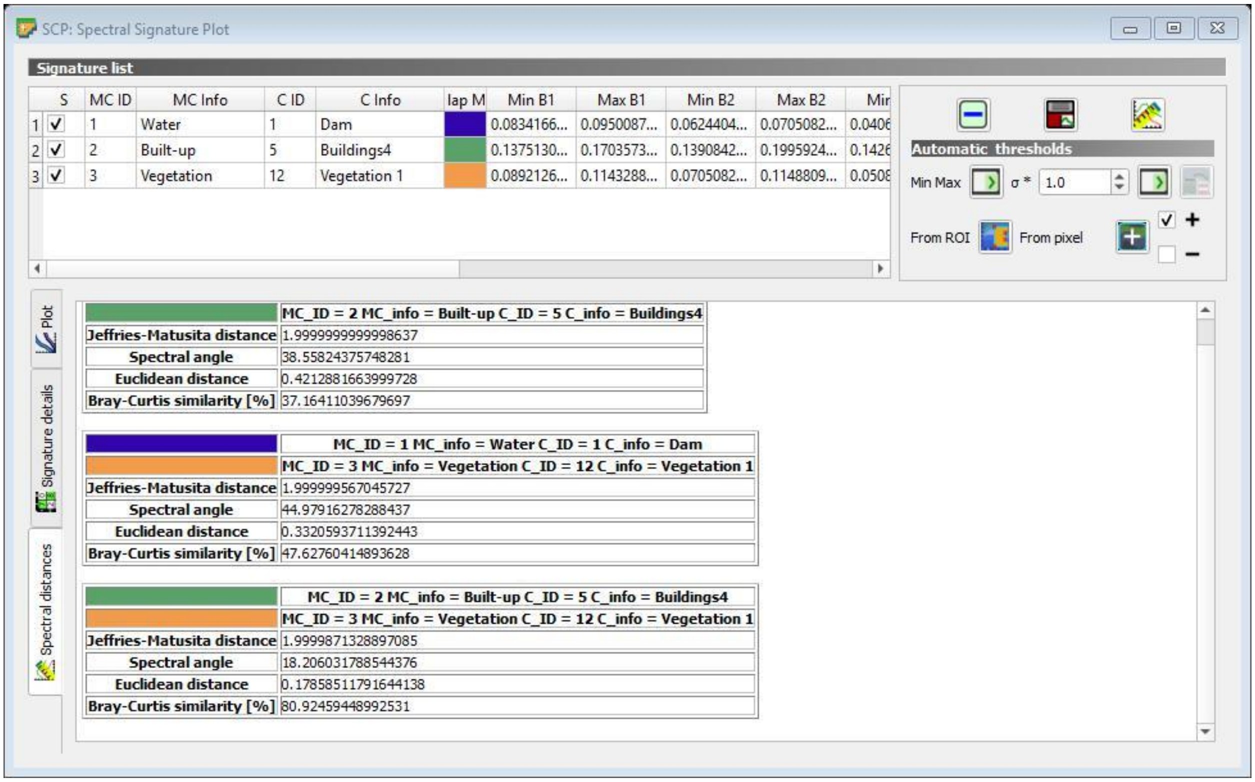
Euclidean distance:

This is useful for evaluating the result of Minimum Distance classifications. The Euclidean Distance is 0 when signatures are identical and tends to increase according to the spectral distance of signatures.

Bray-Curtis Similarity:

It is a statistic used for assessing the relationship between two samples. It is useful in assessing the similarity of spectral signatures. The Bray-Curtis similarity is calculated as a percentage and ranges from 0 when signatures are completely different to 100 when spectral signatures are identical.

4. Highlight two or more spectral signatures with click in the table to plot signature list and click the button . Distances will be calculated for each pair of signatures. When you open the tab Spectral distances, you can notice that similarity between signatures vary according to considered algorithm.



The similarity of signatures is affected by the similarity of materials (in relation to the number of spectral bands available); also, the way we create ROIs influences the signatures. Spectral signature values, standard deviation and other details such as the number of ROI pixels are displayed in the Signature details.

Section 4: Create a classification preview

The classification process is based on the collected ROIs and the spectral signatures. It is useful to create a classification preview in order to assess the results influenced by spectral signatures before the final classification.

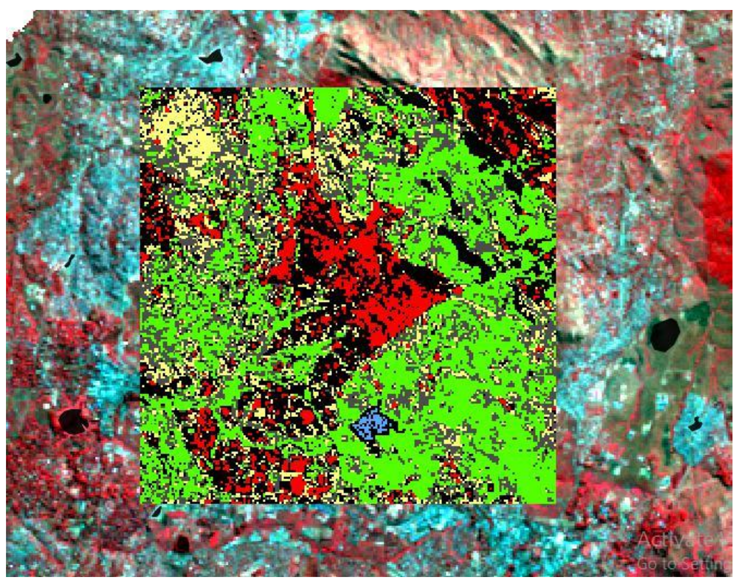
1. Before running the classification, set the color of land cover classes that will be displayed in the classification raster. In the `ROI Signature list`, double click the

   color in the column **'**`Color`**'** of each ROI to choose a representative color of each class.

2. Set the color for macroclasses by assigning blue color to Water, red to Built-up, green for Vegetation and yellow for Bare land:

	MC ID	MC Info	Color
1	1	Water	
2	2	Built-up	
3	3	Vegetation	
4	4	Bare land	

3. Open the **`Classification`** tab to set the use of classes or macroclasses. Check the Use ✓ MC ID and in **`Algorithm`**, select the **`Maximum Likelihood`**.
4. In the Classification preview set **`Size= 300`**. Click the button  and then left click a point of the image in the map. The classification process should be rapid and the result is a classified square centered in the clicked point:



The button  is used to refresh/reload the preview.

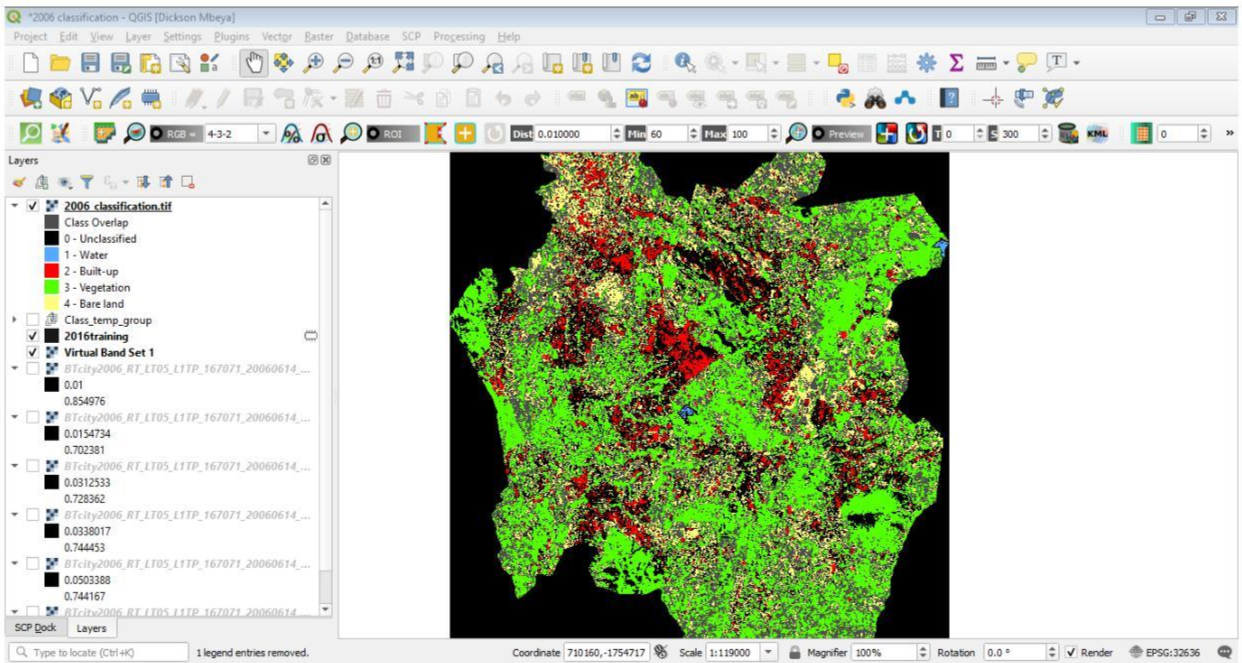
Section 5: Creating the Classification Output

If you are satisfied with the results of the classification previews in that pixels are assigned to the correct class defined in the ROI Signature list, you can proceed to perform the actual land cover classification of the whole image. If you are not yet satisfied with the classification previews, you can create new ROIs in the image.

In Classification, check 

. In the classification output, enable the Classification report options. Click the button 

and save output in a new folder **BTCITY 2006** classified. When processing is done, the classified layer should be added to QGIS:



You should also have the classification report (.csv file) saved in the folder you have saved the classification output. The classification report shows the area statistics of each land cover class. Which land cover class has largest area in the 2006 image?

Congratulations! You have successfully classified the 2006 land cover for Blantyre city. However, you will notice that there are a lot of errors in the classification output because the number or ROIs is insufficient. Creating more ROIs can improve the classification.

Section 6: Assignment

- Create the final map for the 2006 classification adding to your map all necessary map elements and export it as pdf

Section 7: Land cover class area for the 2006 classification

Fill in the table below the areas covered by the four land cover classes.

LAND COVER CLASS	AREA 2006	
Water		
Built-up		
Vegetation		
Bare land		

Use the data you have filled in the table above to answer the following questions:

1. **Which land cover class has the biggest area?**

2. **Which land cover class has the smallest area?**

3. **What do you think could be factors that contributed to the differences in land cover?**

4. **What can be the impact of such land cover differences?**