Lab Assignment #7 – Supervised Classification of Multi-Spectral Imagery

1. Logistics

Date assigned:	Monday, December 09, 2024
Date due:	Monday, January 20, 2025 (via Moodle)
Points:	100 points

Please submit all homework as a single PDF file via Moodle.

In this lab, we will work with optical remote sensing data to look at various land-cover classification algorithms and their use for change detection

2. Initial Setup and Data Sources

Landsat
Landsat data are provided free of charge from the USGS. The data are free, but you will eventually need to register/create an account to obtain data for the second half of this lab. See here: https://landsat.usgs.gov/landsat-data-access. We suggest you create an EarthExplorer account at https://earthexplorer.usgs.gov/. The data for the first half of the lab is available on Moodle.

Information about the Landsat 8 OLI (Operational Land Imager) and the TIRS (Thermal Infrared Sensor)
sensor can be obtained from the lecture and other documents online. A general introduction to the
Landsat 8 sensor is available at https://landsat.usgs.gov/landsat-8 (with additional links on that webpage). Landsat-8 data is provided as a suite of GeoTIFFs for a given tile. Each GeoTIFF/band represents a different wavelength (e.g., ‘visible red’, ‘visible green’, ‘near infrared’, etc.). For reference for this lab, the bands for Landsat-8 are:

Landsat 8 Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS)	Bands	Wavelength (micrometers)	Resolution (meters)
	Band 1 - Ultra Blue (coastal/aerosol)	0.435 - 0.451	30
	Band 2 - Blue	0.452 - 0.512	30
	Band 3 - Green	0.533 - 0.590	30
	Band 4 - Red	0.636 - 0.673	30
	Band 5 - Near Infrared (NIR)	0.851 - 0.879	30
	Band 6 - Shortwave Infrared (SWIR) 1	1.566 - 1.651	30
	Band 7 - Shortwave Infrared (SWIR) 2	2.107 - 2.294	30
	Band 8 - Panchromatic	0.503 - 0.676	15
	Band 9 - Cirrus	1.363 - 1.384	30
	Band 10 - Thermal Infrared (TIRS) 1	10.60 - 11.19	100 * (30)
	Band 11 - Thermal Infrared (TIRS) 2	11.50 - 12.51	100 * (30)

* TIRS bands are acquired at 100 meter resolution, but are resampled to 30 meter in delivered data product.

Caution is needed when using older Landsat products, however, as band designations change between generations (https://landsat.usgs.gov/what-are-band-designations-landsat-satellites).

Uncompress the Landsat Archive (LC08_L1TP_193023_20170602_20170615_01_T1.tar.gz). *Note that this file is compress twice.* You can use 7zip or a similar package to uncompress the gzip file (.gz), then you uncompress the tar file (again with 7zip or a similar tool). Each file is a Landsat band.

You can view the Landsat 8 file by opening it in QGIS.

As before in Lab 2, we will atmospherically correct the Landsat Data. For more specific instructions, please refer back to Lab 2, or simply use the corrected images from Lab 2 for the first part of the lab.

3. Introduction to Supervised Classification – Potsdam/Berlin

Using the Landsat-8 scene from the Berlin region that we worked with in Labs 2 and 4, we will use the Semi-Automatic Classification (SCP) Plugin in QGIS to divide the images into four main classes:

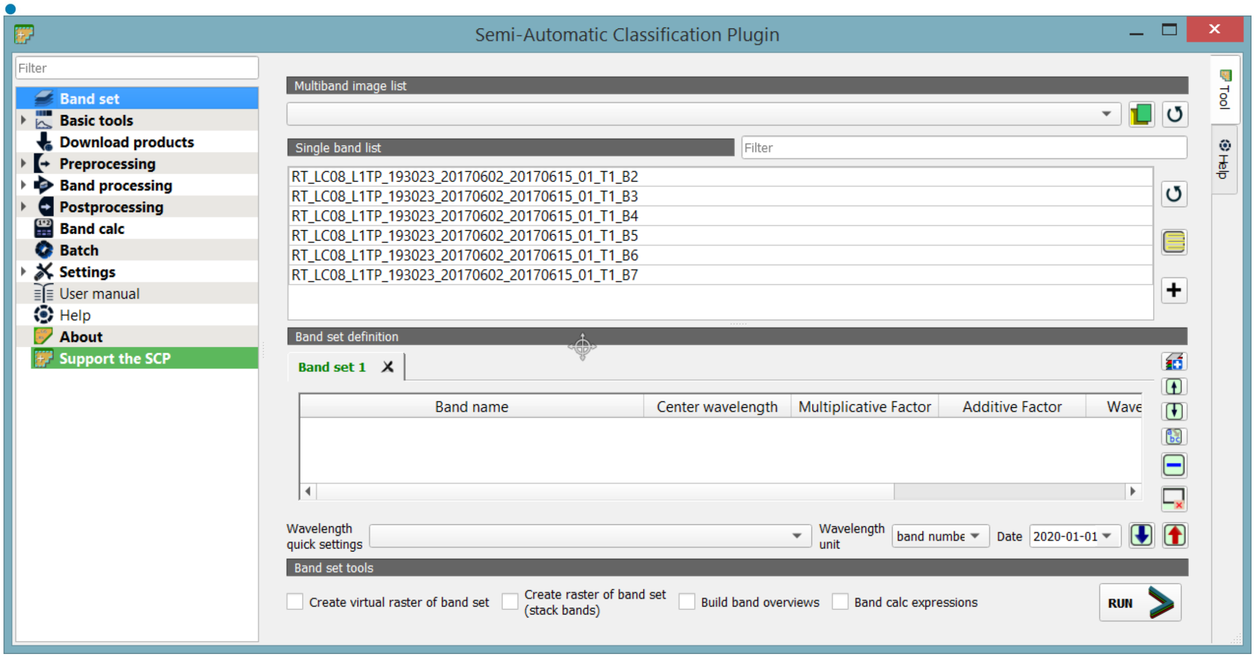
(1) Urban / Built Up

(2) Water

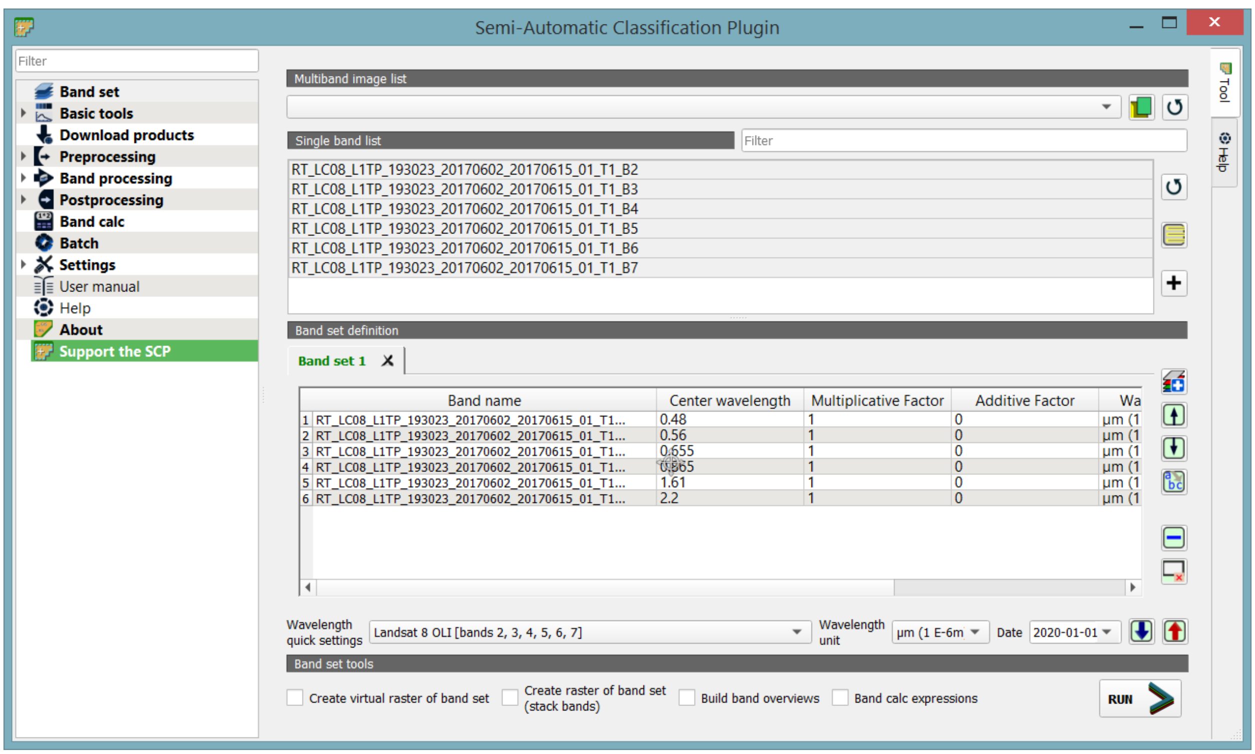
(3) Forest

(4) Cropland

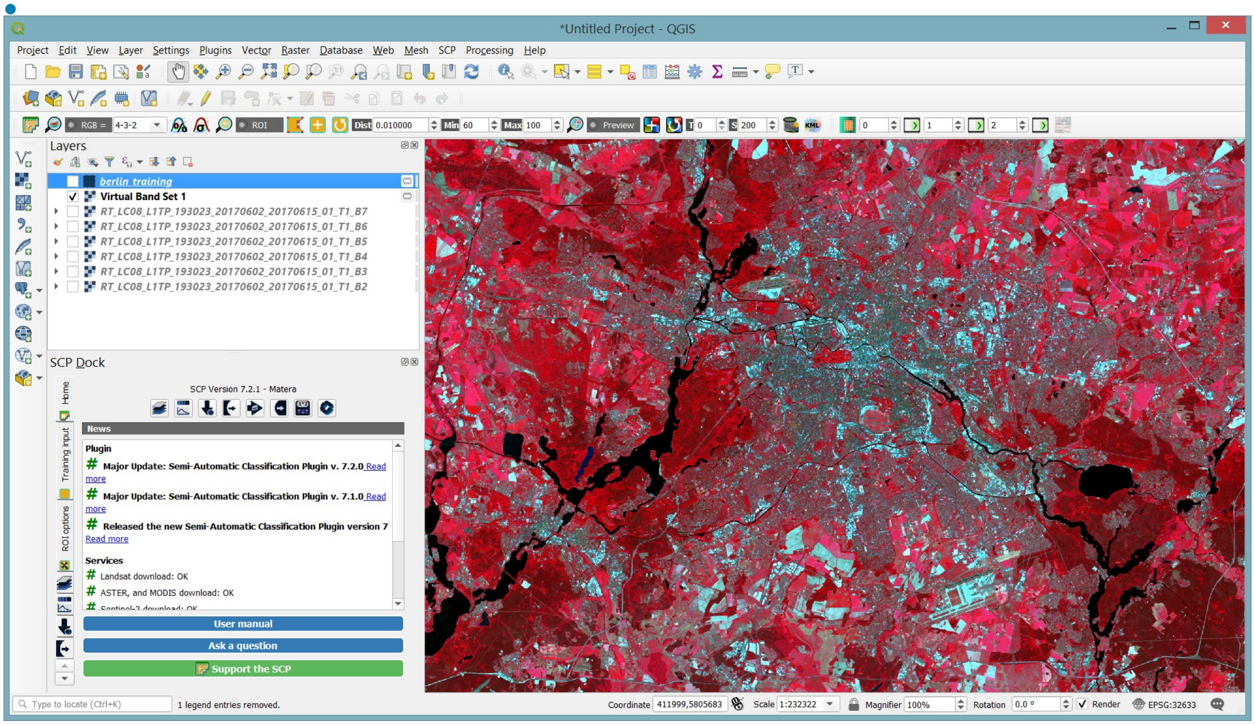
Using the **SCP Input menu** in the **SCP Dock**, load in Landsat-8 bands 2-7. You can load all rasters at the same time, to create a band set.



You can use the **Band Set** menu to ensure that bands are designated in the correct order.

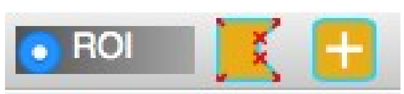


Once the bands are correctly ordered in the band set, we can visualize the data either in true color (3-2-1) or in false color (4-3-2). False color may be more useful for differentiating between land cover types.

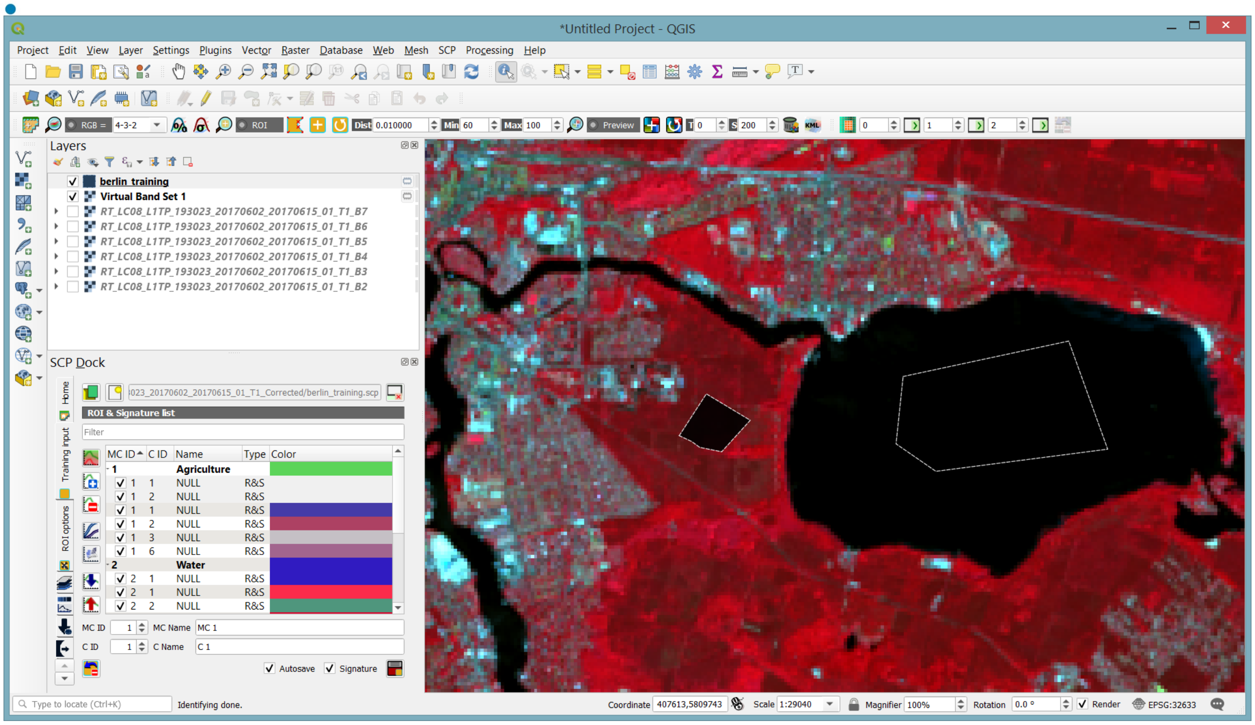


To teach the program which spectral signatures correspond to which land covers, we need to create training input. To create training data, we will designate regions of interest (ROI) of known land cover types to teach the program what the spectral signature of each is. To create the new training input, select **Create a new training input** from the ***SCP Input*** dock. This will prompt you to save the new training file, which you can call something like *'berlin_training.scp'*.

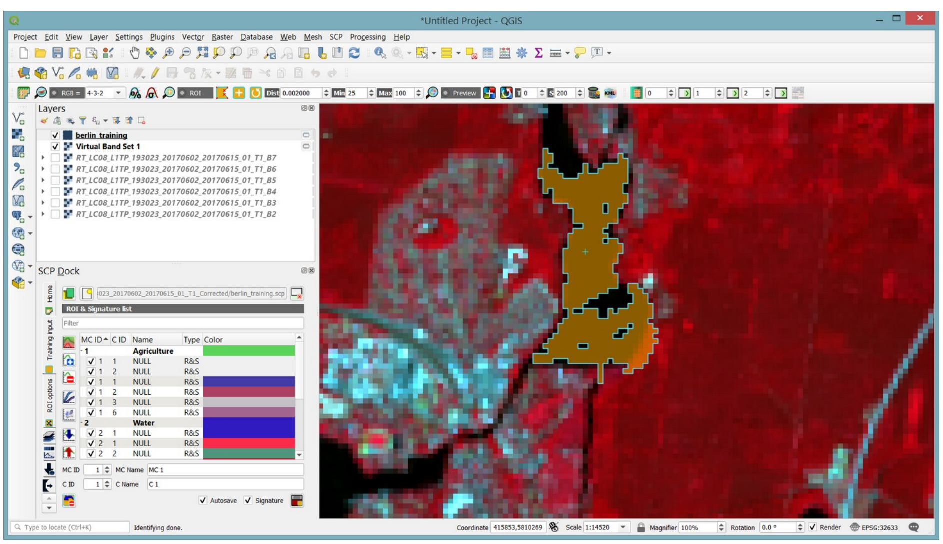
Once you have created the training input file, you can start designating ROIs. To manually create a polygon ROI, select the ***'Create a ROI polygon'*** button. You can then define polygon vertices by left clicking in the ROI. Close polygons by right-clicking.



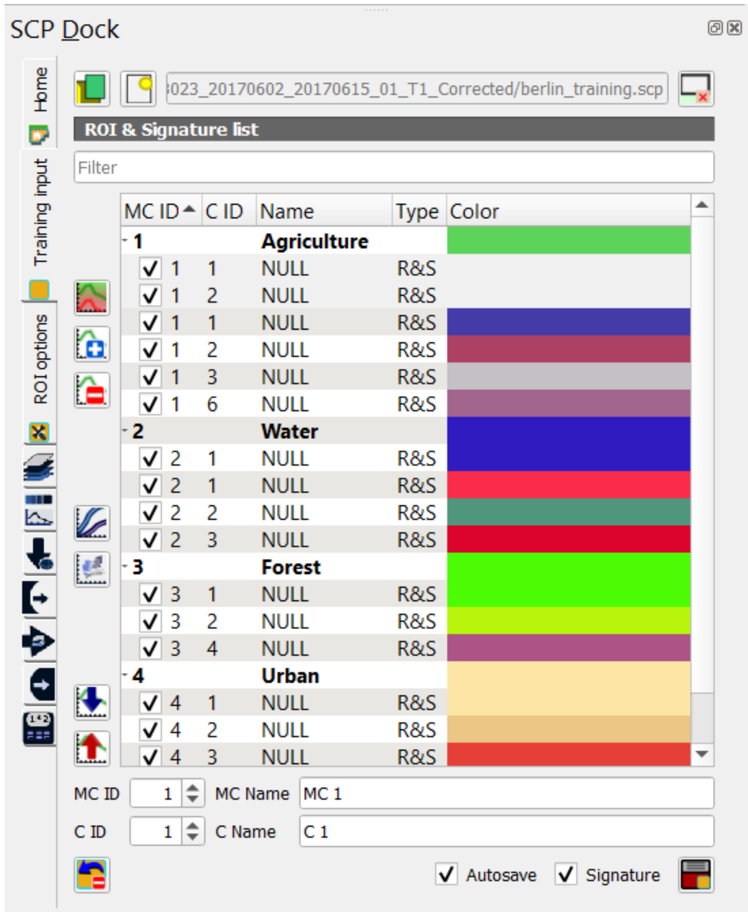
In the Classification dock, you can enter a description for the **Macroclass** (e.g., 'water') and the **Class identification** (e.g., 'Lake'). When you create the next ROI, the class ID will immediately increase. However, if you want to create a new macroclass (e.g., 'Built-Up', 'Forest', etc.) you will need to manually change the macroclass ID.



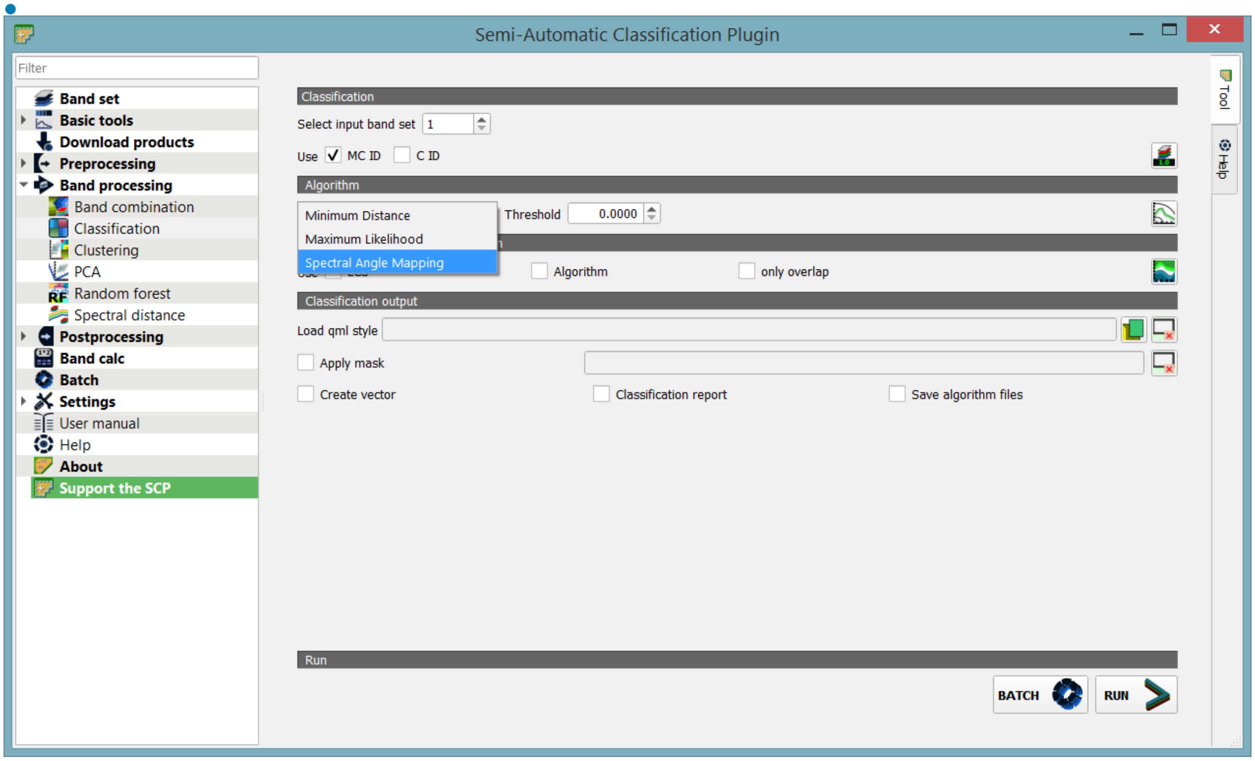
The other way we can create ROIs is the with the ***'Active ROI Pointer'*** option. This will automatically generate an ROI of like-spectral-signature within a certain distance of where you click. The automatically generated ROI will exclude regions that are not similar to where you have clicked.



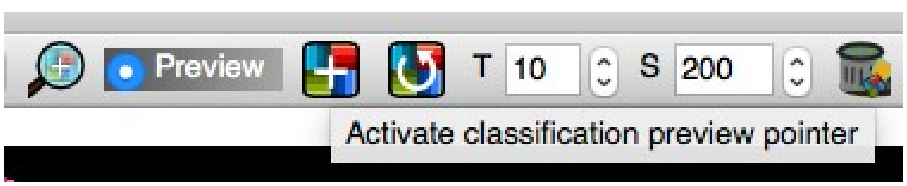
Using either the manual ROI or the automatic ROI generation, create ROIs for **(1) water, (2) urban, (3) forest,** and **(4) cropland** classes. In the case of cropland especially, you may want more than one ROI. As you can see in the false-color image, there are fields that likely correspond both to bare fields and growing fields. Using the **Macroclass** option, we can create two different cropland classes. In the example below, I have also created two classes under the "Built-Up" macroclass.



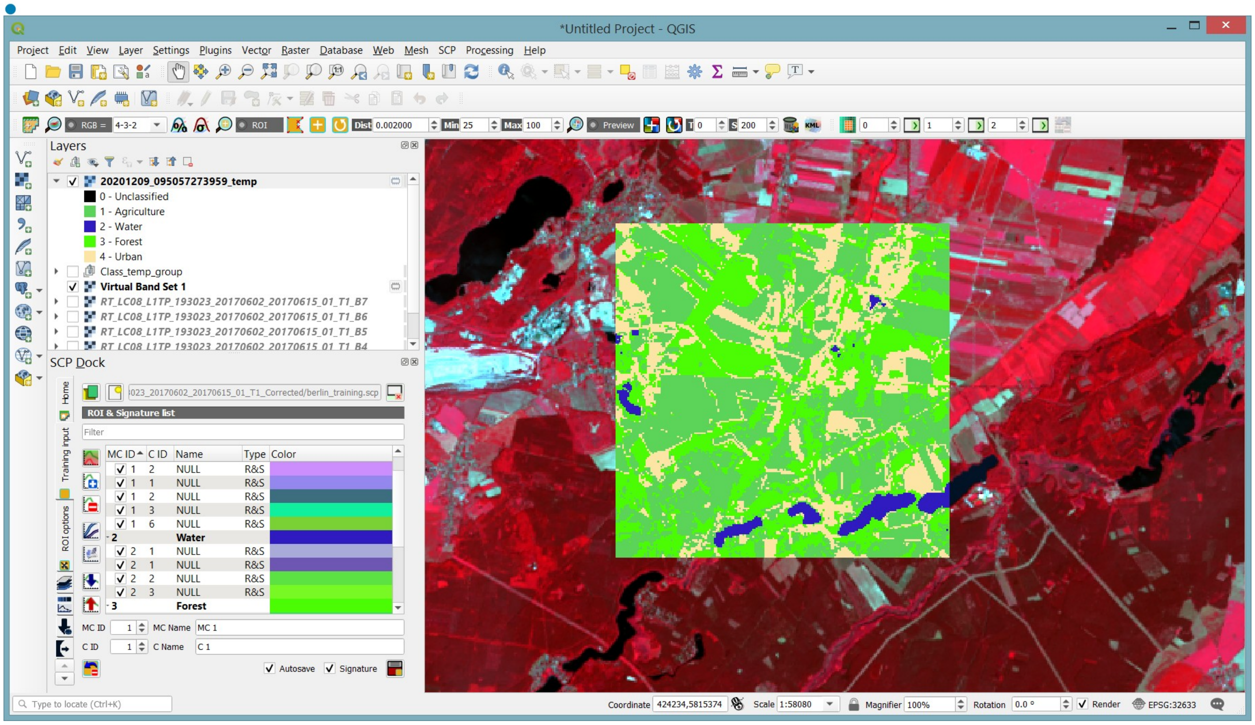
Once you’ve created all of the ROIs, we can preview the classification. But before we do this, we need to select the classification algorithm to use. You can change this with the **Classification Algorithm** menu in the SCP dock. For now, designate the classification algorithm as *‘Spectral Angle Mapping.’*



It is worthwhile to preview the classification before classifying the entire image. This will give us an immediate idea if the ROIs do a good job of describing the land cover classes or not. We can use the **Preview toolbar** to classify a subset of the image.

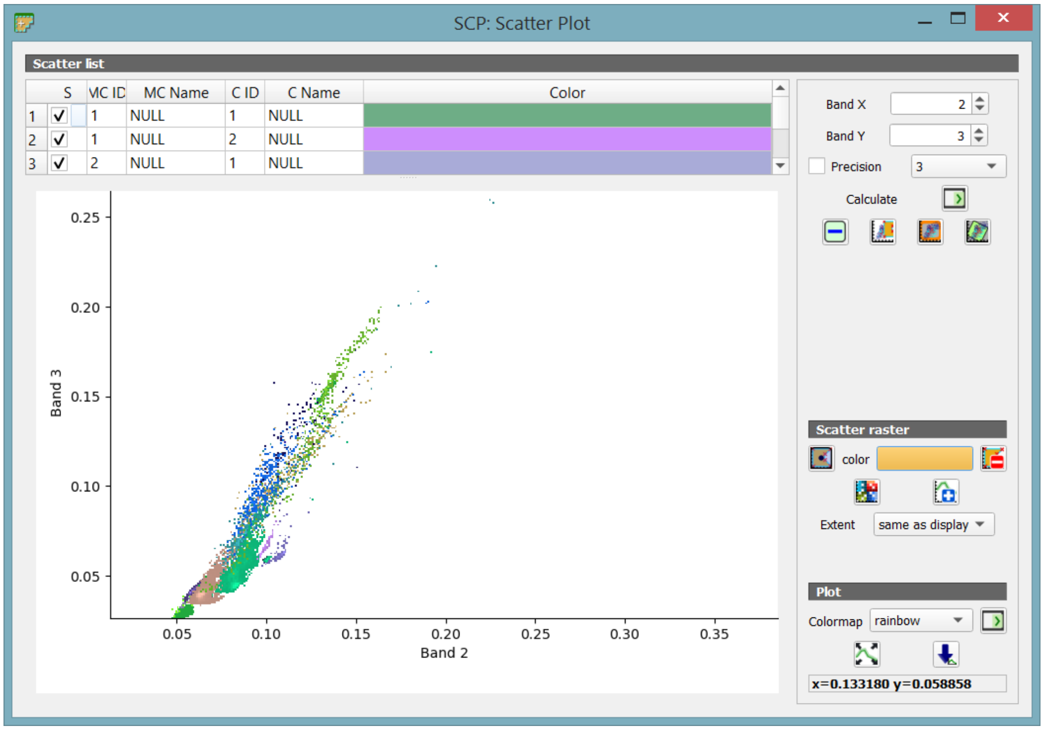


Select the ***'Activate classification preview pointer'*** and click a place of interest on the map. Within a few seconds, a preview of the classification will be generated in QGIS. Does the classification seem realistic?

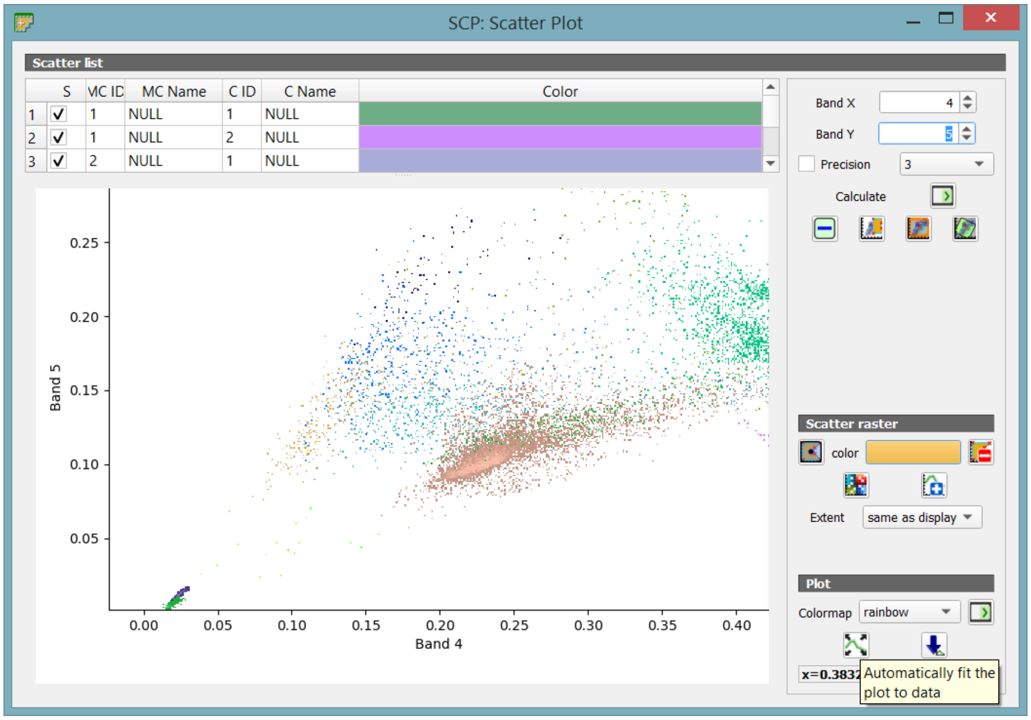


Another way we can judge whether or not we have distinct classes in our training input is to look at a Spectral Scatter Plot. This will generate a scatter plot of the different classes' spectral values for different bands. You can access this by clicking the button in the **ROI Signature dock** called ***'Add highlighted items to scatter plot.'*** As you can see in my example below, there is not a very meaningful difference between what I have designated 'urban' and 'industry'. In this case, I can merge the two ROIs by highlighting them and using the option to ***'Merge highlighted spectral signatures obtaining the average signature'.***

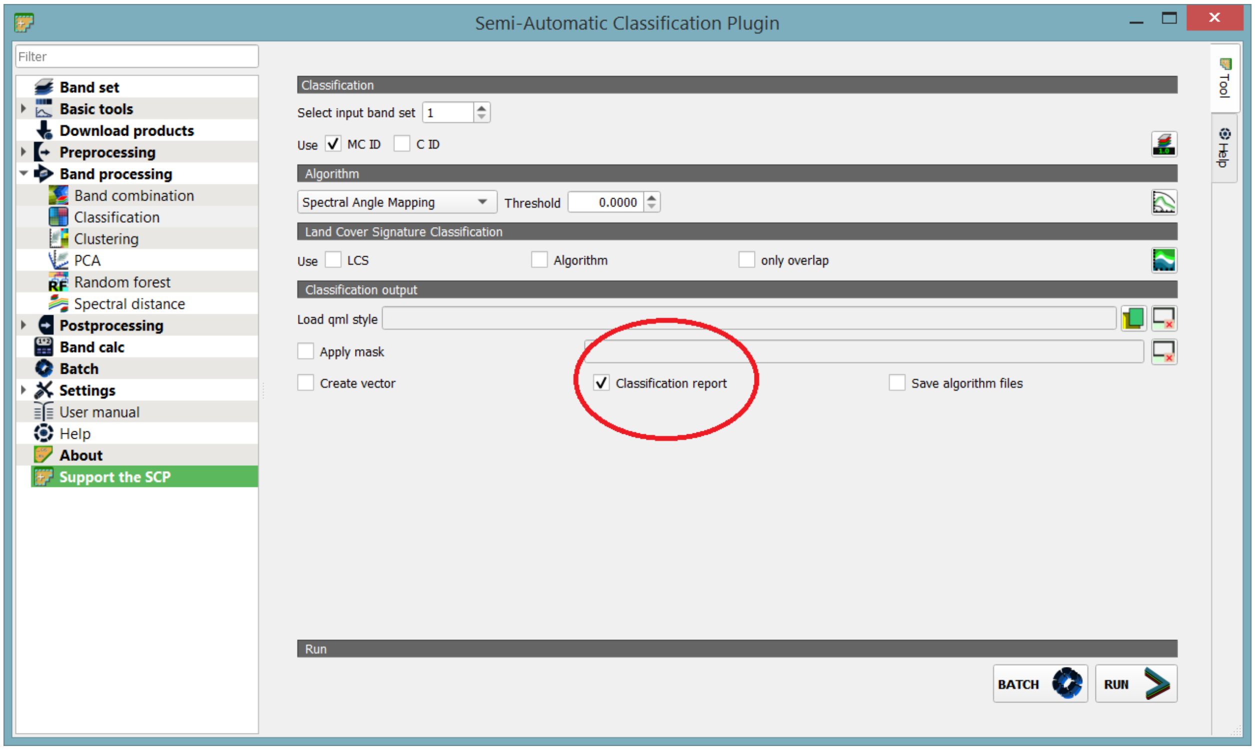
Scatter plot between B2 and B3



Scatter plot of B4 and B5



When you are happy with the classified output, you can save it using the **Classification Output** tab of the SCP dock. Click the *Classification Report* radio button to get an CSV output that describes the percent and total area in each class.



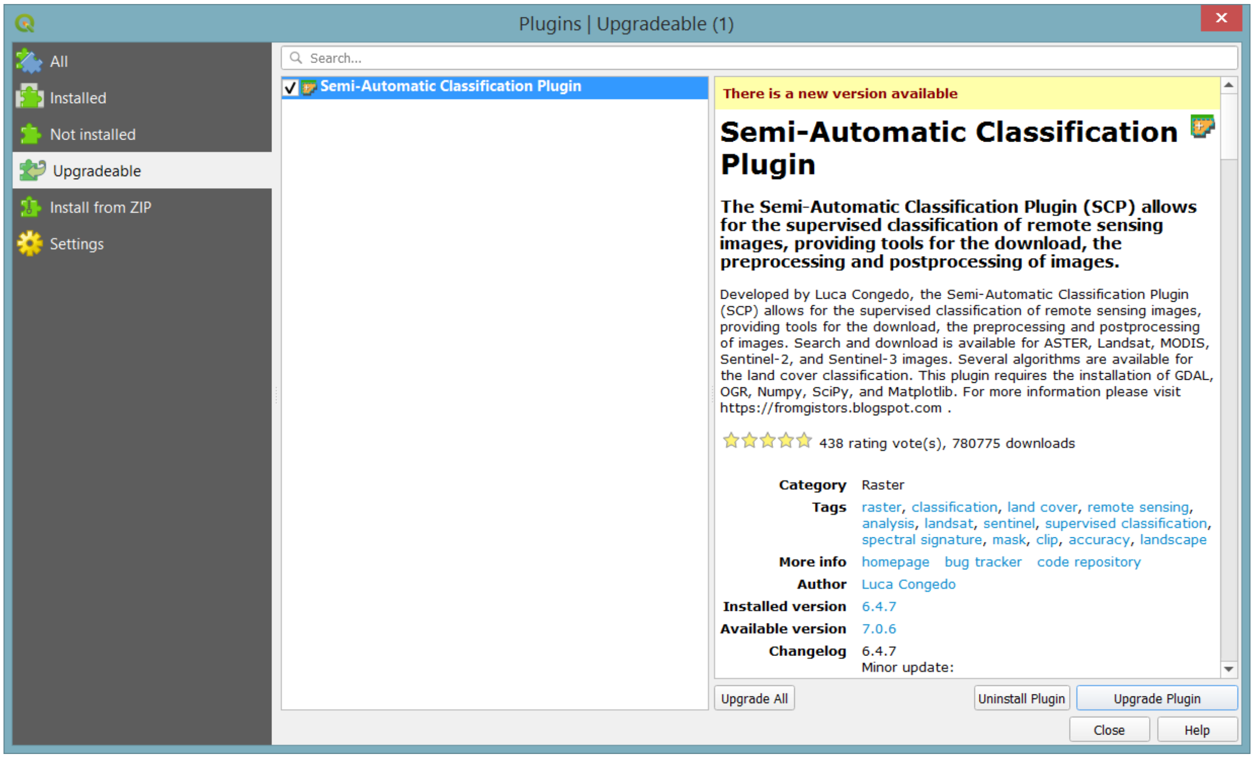
<u>Question 1:</u> Create 3-4 additional ROIs for each class to create a more robust classification of Berlin using Spectral Angle Mapping. Does this improve your classification? Include sample maps and the classification report as a table (20 points).

<u>Question 2:</u> Repeat your classification with the other two classification algorithms (Minimum Distance and Maximum Likelihood). Describe the difference between each output. Include a sample map of a small area (zoomed in to highlight differences) showing all three classification results (20 points).

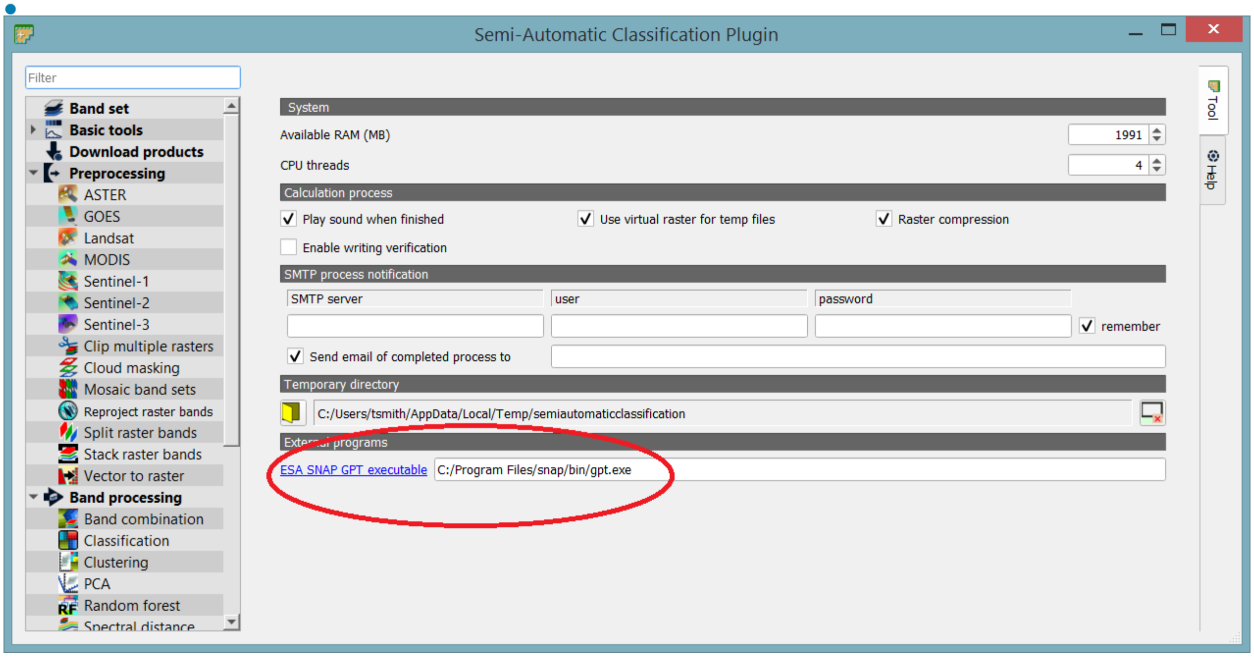
<u>Question 3:</u> How does the supervised classification compare to the unsupervised classification? (10 points)

4. Random Forest Classification

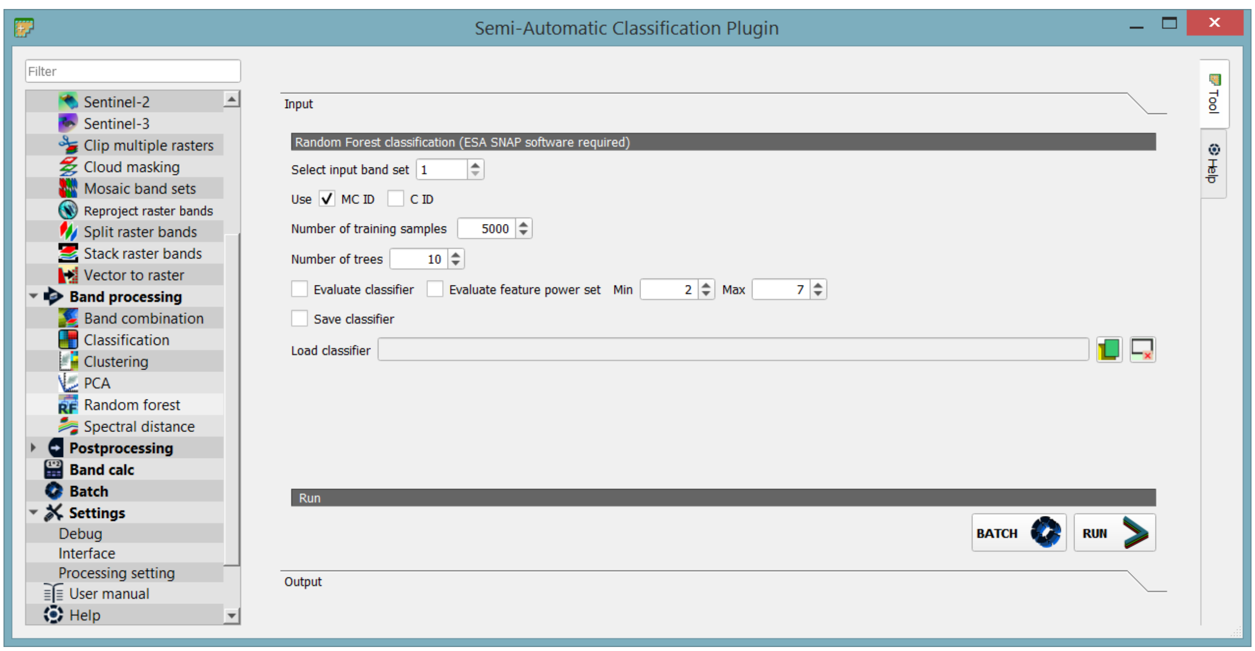
To use the Random Forest classifier, you need to first make sure you have the most up-to-date SCP Plugin:



You will also need to point QGIS/SCP to the ‘gpt.exe’ executable of SNAP to run the tool:



Now we can take a look at the RF classifier:



<u>Important Note: This tool does not seem to work on VRT data – you need to load in the actual TIF files to make this work!</u>

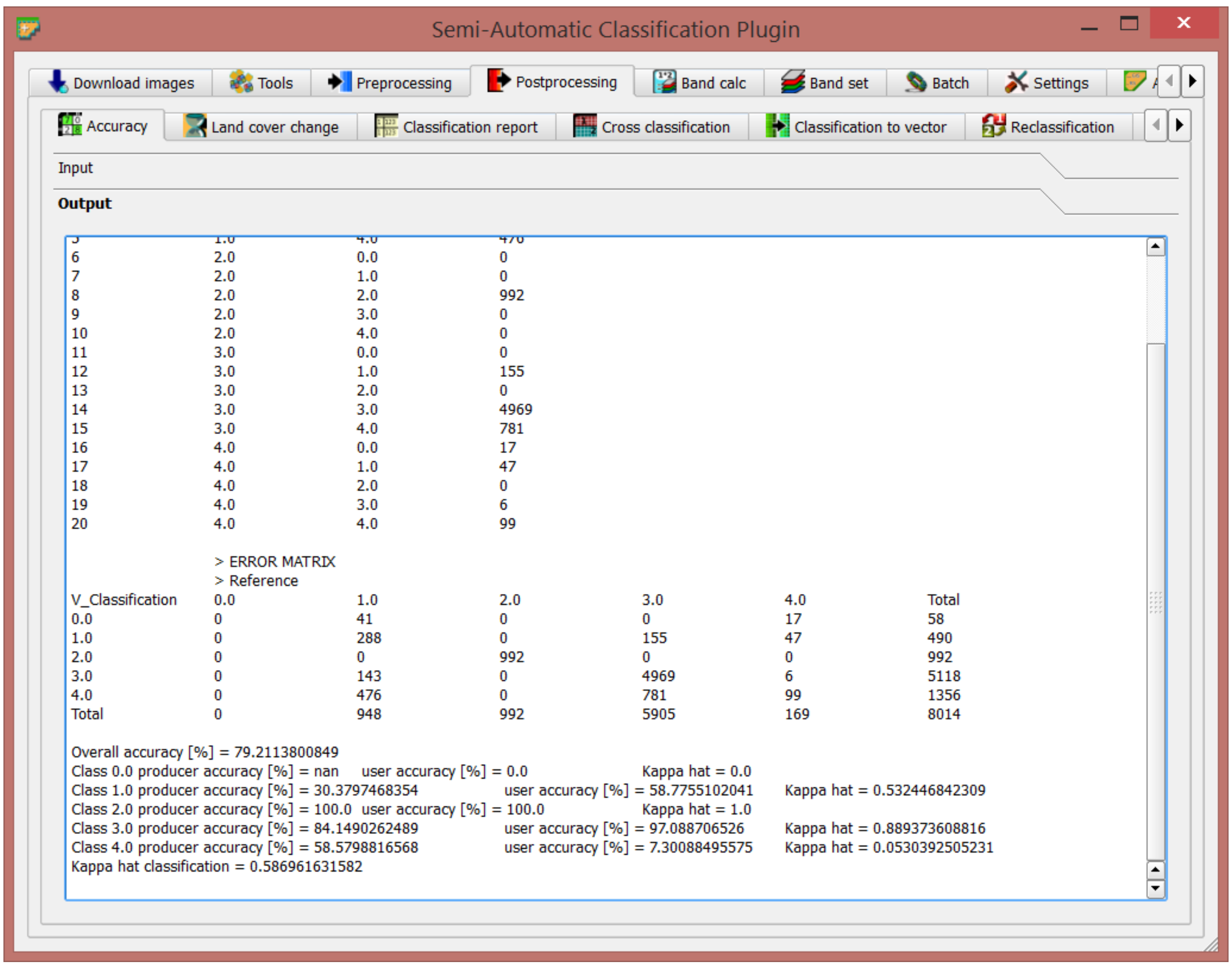
Try running the tool on the default settings (above) with the same training input as you used previously. How do things compare?

> **<u>Question 4:</u>** Provide a map of your Random Forest classification (with scale, legend, etc). How does this compare to the results from the other algorithms we've used in this lab? (10 points)

Troubleshooting hint: If this is taking a very long time to run, you can try decreasing the number of 'trees' in the RF algorithm. You could also try subsetting your input image to a smaller spatial area – this will take some time to run if you do it over the whole scene!

5. Classification Accuracy

While we can visually see which areas seem to be well and poorly classified, it is often useful to do a more rigorous statistical analysis of the classification results. The first step is to assess how well your training data compares to the final classification. We use here the **Accuracy Assessment** tool. This will return a raster showing the classifications within each of your training zones as well as a table telling you about the accuracy of each of your classes. As inputs for this tool, choose the classification raster and the training dataset you used to make the raster. **Note that this will work for any of your classifications (any algorithm), as it just compares your result to the input data!**



For each of your landcover types you get a total error and a User's/Producer's error. More simply, User accuracy refers to 'inclusion' (e.g., how many wrong areas get included in the classification) and Producer to ‘exclusion’ (e.g., how many areas are missed). These statistics can help determine which land cover types in your classification are most useful. In the picture above, water is the only class (Class 2) that is really well classified in my map.

Repeat this step with your Random Forest classification. How do the two compare?

> **<u>Question 5:</u>** Please include the Accuracy Report for all of your classifications (Spectral Angle, Max Likelihood, Min Distance, and Random Forest). Which one works best? It is fine to include these as screenshots (20 points).

Next, use the ‘Classification report’ tool to get bulk statistics (i.e., how much area within each raster is covered by each land cover type.

> **<u>Question 6:</u>** Please include the Classification Report for all of your classifications. Where are there differences between the results? It is fine to include these as screenshots (20 points).