University of Oregon - GEOG 485/585 – Remote Sensing 1 – Fall 2021

Lab Assignment #4: Supervised and unsupervised classification

Objectives: In this lab, you produce a land cover map of the McKenzie River Valley in July 2021 following the Holiday Farm Fire using both a supervised and unsupervised classification in QGIS.

Logistics:

Date assigned:	Week 4
Date due:	Before the beginning of Week 5 labs
Points:	100 points
Deliverables:	Post a PDF document on Canvas with all answers and necessary graphics using the answer sheet given to you. Your responses are expected to be in complete, grammatically correct sentences based on knowledge gained from lecture, reading, and lab exercises. Remember that some of the questions have multiple parts so make sure you answer all parts of the questions.

Preface:

The lab instructions will be available on Canvas (canvas.uoregon.edu) and the class network drive (details below). Data used in labs are not posted on canvas and should be accessed through the SSIL network (you can also remotely connect to this network from off campus, following these instructions).

The class data is available at the following path:

`R:\GEOG485_12740_05\Class_Data\Lab4\`

You should store your work in a separate folder for each lab on your local computer. **Note:** I recommend anything you save on your local computer should also be backed up to a USB stick or Google Drive account at the end of each session. Your folder should have the following path (except has your own SSIL username):

`R:\GEOG485_12740_05\Student_Data\Your_Username`

For this lab, create a "`Lab4`" folder in which to store your work inside your user folder. Your lab write-up and any additional materials should be turned in online to the course website on canvas.uoregon.edu. Lab questions are due by the start of the next lab, but it is suggested you do these questions first thing.

Lab overview:

The goal of image classification is to assign a particular range of pixel values in spectral bands to a “class” or category of interest. In this lab, you will first implement a supervised classification using your knowledge of the area, image interpretation skills, and understanding of the spectral classes to train the classifier. There are a number of supervised classification methods, many of which have a statistical basis that requires specification of certain parameters in advance. We will use the *maximum likelihood* method. As discussed in lecture, this method assumes that the statistics for each class in each band are normally distributed and calculates the probability that a given pixel belongs to a specific class. Each pixel is assigned to the class that has the highest probability (that is, the maximum likelihood).

1. Import data

- Import all seven Landsat bands from `R:\GEOG485_12740_05\Class_Data\Lab4\data\` into your QGIS workspace. Note that I have clipped these rasters because take a long time to classify the entire Landsat image.

2. Supervised classification

2.1. Install the Semi-Automatic Classification (SCP) Plugin

We need to download an extension called the *Semi-Automatic Classification* plugin, a free open source plugin for QGIS that allows for the supervised classification of remote sensing images.

- Click **Plugins** from the top menu → **Manage and Install Plugins** and search for **Semi-Automatic Classification Plugin**. Click **Install Plugin** then **Close**
- After installing the plugin, you should see a window **SCP Dock** appear on the bottom left corner of your QGIS window
- There should also be a new **SCP toolbar** near the top of the QGIS window where you can adjust the RGB color composite and create ROIs

2.2. Loading images into the Semi-Automatic Classification Plugin

- In the **SCP toolbar**, click on the left-most SCP button and select the **Band set** tab (you can also click on the button in the SCP dock)
- Click the **Refresh list** button on the right next to **Single band list** to load the Landsat 8 bands
- Select all the Landsat 8 bands by **left-clicking** while holding **Ctrl**. Make sure they are all selected
- Click **Add band to band set** button on the right
- In the **Wavelength quick settings** at the bottom, click on the drop down list and select **Landsat 8 OLI [bands 1, 2, 3, 4, 5, 6, 7]** to automatically set the center wavelengths for the selected OLI bands

- That’s all you have to do to get the SCP plugin to recognize your selected raster image, you can now close the dialog box

- To check that the bands were properly loaded, select “4-3-2” from the drop down list next to where it says **RGB =** in the **SCP toolbar**.
- You’ll notice that you now have a layer called “*Virtual Band Set 1”* in your layers panel
- You can edit the contrast enhancement by clicking the **Local standard deviation stretch of band set** button to the right of the **RGB =** button.
- Try changing the color composite by typing the band numbers into the **SCP toolbar**. You do not need any commas or spaces, just the band numbers that correspond to the Red, Green, and Blue display channels. Load a *SWIR-NIR-Red (6-5-4)* composite. Press enter after typing

2.3. Creating training sites

The key step in supervised classification is identification of training data, in the form of Regions of Interest (ROIs). <u>We will identify **at least six** land cover classes or ROIs</u>. Of course, non-land features may be classes as well. The more specific your ROIs are the better the classification will be. You also need to strike a balance between covering enough pixels to spectrally represent each land cover type and covering too large of an area that you lose the specificity. It is important to remember that there are subtle distinctions in brightness between classes which we may think of as being the same – these may necessitate the creation of multiple classes (e.g. burnt forest, clear-cut forest, healthy forest).

- To produce ROIs, go to the **SCP Dock** (i.e. the lower left panel) and click on the **Training input** tab. This should open the **ROI & Signature list** panel
- Click on the **Create a new SCP training input** button
- Save this (currently empty) file to your Lab 4 student folder with a sensible name such as “*ROIs_training.scp* ”

Before we start producing our ROIs we need to sensibly label our training data.

- At the bottom of the **SCP Dock**, set **MC Name** → “*land_cover*” and **C Name** → “*snow*”
- While we’re here, you can also untick the **Signature** button, we can compute that later
- Navigate back to the **SCP toolbar** (i.e. at the top) and click the **Create a ROI polygon** button. Your cursor should now look like a plus sign.
- Start by drawing a polygon around some pixels that you can identify as snow by left-clicking. Right-click to finish the polygon.
- Now holding **Ctrl** on the keyboard, repeat for a few more areas of snow. If you don’t hold **Ctrl** you will lose the previous polygon you drew
- Once you have digitized three or four snow polygons that you are satisfied with, click on the **Save Temporary ROI to training input** button on the bottom right of the SCP dock. This will add your ROI to the list.

Now repeat this step but for areas that represent healthy forest cover.

- First change **C Name** to "healthy_*forest*"
- If you make a mistake you can left-click on the row, then right-click and select **Delete items**.

Repeat these steps until you have produced ROIs for **six different classes** (C ID). Note that you do not need to change the C ID (it will automatically increase by one every time you create another ROI).

2.4. Extract spectral signatures

Once you have created multiple ROIs for six unique classes, we can calculate the spectral signature of each ROI.

- Highlight each class ROI by Ctrl + left-clicking on them in the list (they should be highlighted in blue)
- Now click on the **Calculate signatures for highlighted items** button
- Be patient as it may take a few minutes to compute
- If you get a warning message saying: "`Warning [12]: The following signature will be excluded if using Maximum Likelihood`", your ROI is too small (or too homogeneous) for the Maximum Likelihood algorithm. You should go back and create slightly larger ROIs.

2.5. Training site evaluation

We will now calculate spectral distances between classes to get a sense of what and analyze them.

- Now highlight each class ROI again by Ctrl + left-clicking on them in the list (they should be highlighted in blue) and click the **Add highlighted signatures to spectral signature plot** button.
- A new **Spectral Signature Plot** window with the spectral signature plot should appear. In the plot we can see the line of each signature and the spectral range (minimum and maximum) of each band. The larger is the semi-transparent area of a signature, the higher is the standard deviation, and therefore the heterogeneity of pixels that composed that signature.

- Let's look at the details of the six classes. On the left side of the **Spectral Signature Plot** window, click on the **Signature details** tab.
- This table shows the *mean* (although it is labeled **Values**) and **Standard deviation** for each class for each of the seven Landsat bands.

- Calculate spectral distances between your classes in the **Spectral Signature Plot** window by clicking on the **Calculate spectral distance** button. On the left side of the window, click on the **Spectral Distances** tab. This will display four statistics for each combination of your classes which help us assess if different classes that are too similar could cause classification errors. We're interested in the *Jeffries-Matusita* *distance* (calculates the separability of a pair of probability distributions), where 0 indicates identical signatures and 2 indicates completely different spectral signatures. Any value greater than 1.8 is considered good separability.

****
Question 1 (20 points):

a) List the Jeffries-Matusita distances for each pair of classes.
b) Provide an example of a pair of classes that have good separability and a pair of classes which have poor separability?
c) Why does one pair of classes have better separability than the other (HINT: use the **Spectral details** and **Plot** tabs to provide evidence for your answer)?
d) If you repeated training site selection are there any classes you would want to add, remove or split up?

****

3. Supervised classification using maximum likelihood method

Now that we have our training ROIs, it is time to perform a supervised classification. We will be using the "**maximum likelihood**" method.

- Open up the SCP window by clicking on the left-most button on the **SCP toolbar**.
- On the left side of the window expand **Band processing** and then click on **Classification**.
- Keep the input band set at 1 and make sure that the "*C ID*" is checked because we want to use the class statistics to train the classification.
- Under **Algorithm**, click on the drop down menu and select "*Maximum Likelihood*".
- **Threshold** represents different things depending on the classification method. For Maximum Likelihood, this sets the probability threshold. Let's start with setting the probability threshold to "0" (meaning all pixels will be classified). Using a threshold between 0 and 100% means that the algorithm will not classify pixels with a lower probability value than this threshold value, preventing too many unlikely pixels from being put into that class.
- Click **RUN**. A window will pop up asking you to save the classification output. Save as "*supervised_classification.tif* " and click **Save**.

If everything runs smoothly, you should see a new raster layer with your supervised classification.

- Change the colors of the classes to something sensible by going into the **Layer Styling** panel and opening the **Symbology** window.

****

Question 2 (20 points):

a) Which land cover classes did you choose to classify in this image?
b) What colors did you use to display your classes?
c) Provide a PNG image of your supervised classification with your answer.

Note that there is no right answer to this question, I am just looking for a clear, logical description of your thought process.

****

4. Post-processing of the classified map

Since we used a pixel-based classification, it is likely that our classified map has a lot of speckle (individual spurious pixels scattered like salt and pepper throughout the image). We can resolve this using a bit of post-processing.

- Open the SCP window by clicking on the left-most button on the **SCP toolbar**.
- On the left side of the window, expand **Postprocessing**, then click on **Classification sieve**. This allows for the replacement of isolated pixel values with the value of the latest neighbor patch, and it's useful for removing small patches from a classification.
- In **Select the classification**, open your supervised classification map (you might have to click on **Refresh list** to see your classification).
- The **Size threshold** is the size of the patch to be replaced (in number of pixels). Try out different values (between 2 and 5 seems to work best), and pick whatever looks right.
- The **Pixel connection** refers to whether diagonal pixels are considered connected (pixel connection "4") or not connected (pixel connection "8") in a 3x3 window. Keep the default pixel connection "4".
- Click **RUN** A window will pop up asking you to save the Classification sieve output.
- Compare the post-processed map to the original map. You should notice that there are far fewer errant pixels on the new map.

5. Analyzing your results

There are several different ways to analyze your results. The first is, of course, to compare the classified map to the original image and visually inspect its quality. A second way is to look at the class statistics by using another post-processing tool.

- On the left side of the SCP window, expand **Postprocessing** and click on **Classification report**
- In **Select the classification**, open your *post-processed classification map* (you might have to click on **Refresh list** to see your post-processed classification).
- Click **RUN.** A window will pop up asking you to save the report as a .csv file.
- The output should appear, showing the total number of pixels in each class, percentage, and area in square meters.

****
Question 3 (20 points):

a) Provide a table showing the results of your classification, adding a column with the name of each class.
b) To what extent do these area/percentages corroborate what you expected before you ran the classification?

****

6. Unsupervised classification using K-means clustering

Now we will perform an unsupervised classification on the image.

- Open up the SCP window by clicking on the left-most button on the **SCP toolbar**.
- On the left side of the window expand **Band processing** and then click on **Clustering**.
- Keep the input band set at 1 and make sure that the "*K-means*" is checked.
- Check "*Distance threshold"* and set it to 0.0001, set the "*Number of* classes" to **6** and *"Max number of iterations*" to **10**.
- Click **RUN**. A window will pop up asking you to save the classification output. Save as "*unsupervised_classification.tif* " and click **Save**.

- Now retrieve the report for this classification by going to **Postprocessing → Classification report**
- In **Select the classification**, open your *unsupervised_classification* map (you might have to click on **Refresh list** to see it).
- Click **RUN.** A window will pop up asking you to save the report as a .csv file.
- The output should appear, showing the total number of pixels in each class, percentage, and area in square meters.

****
Question 4 (20 points):

a) Are you able to interpret any of the classes detected by the unsupervised classification? If so which ones? Use the classification map and report as guide.
b) Provide a table showing the results of your classification, adding a column with a rough interpretation of each class. If you cannot interpret the class, leave it as N/A
c) Provide a PNG image of your unsupervised classification with your answer.

****

Overall, classifying this image has proved a challenging task. The supervised classification is very sensitive to the training data and it is not completely clear what unsupervised classification represents. Use Question 5 below to reflect on this assignment and think about how you could improve this classification task.

****

Question 5 (20 points): Describe how the following might be used to improve your classification?

a) An object-oriented approach
b) Ten more Landsat images over the same area between 2011 and 2021

****

This concludes Lab 4! Remember to type up all answers on your word document, convert files to PDF and upload it to Canvas by the deadline.