

GEOG432: Remote sensing Lab 2

Unsupervised classification

General Goal:

This lab involves identifying land cover types by using algorithms to identify pixels with similar multispectral Digital Numbers (DN) and spectral signatures

0. Start – a bit about Linux

Login to the linux environment – you already have in order to view this webpage
Linux enables both a command line interface ('CLINT') and a Graphic User Interface (GUI) ... sometimes referred to as Windows Icon Mouse Pointer (WIMP). Usually the command line works faster, but is less familiar to new users.

For this course, you should create a new folder ... use the GUI to do this:
Find the 'files' (folder) icon at the bottom or use Places -> home folder
right-click in the new window, and select new folder
name it **geog432** (no spaces)

Notes on file and folder names

1. NO spaces EVER in folder (or file) names
2. Keep them concise ... GIS software has a letter limit when loading files from folders
3. Generally use only lower case letters – think like a computer !

In general you can likely do all labs using only the GUI; However here's a quick tutorial on a few key commands:

Command line example introduction

find the (black) terminal icon at the bottom or use applications -> accessories -> terminal (a window appears with prompt at your home folder)

Type these commands (enter after each command):

- ls – lists all files and subfolders
- cd geog432 – this changes the operating folder to your newly created geog432
- ls -l – this lists the details of all files in your folder none so far !
- du -sh – this totals your disk usage
- du -sh * – this details how much disk space you are using in each subfolder

one command you may need is to end a non-responsive task – similar to windows task manager:

`ps -u username` (where username is your login) ... this will list all active processes.

Now start GIMP (under graphics) ...

repeat the command directly above – you can use your up/down arrow to avoid retyping

`ps -u username`

Now you will see the GIMP process listed, and its process number

Let's pretend it has frozen (this may happen if you try to open a large orthophoto)

you can end the process by typing

`kill xxxx` (where xxxx is the process number)

if this fails to work, try

`kill -9 xxxx` (the -9 option = 'yes I really mean kill !')

this is a harsh way to end a process unless required – like stopping your car by turning off the ignition while moving; you may need to use this for example if PCI freezes

Two more variations on **cd** command (change directory)

`cd ..` takes you back to the previous folder

`cd` takes you back to your home folder

OK – minimise the terminal and home folder shells for future use – you just passed Linux101

In this lab, you will create new channels (layers) not just display existing bands, so you will need your own copy of the files to write to:

Use the folder manager to copy these two files from */home/labs/geog432/data2018* into your geog432 folder - the one you just created: (copy/paste - same as in windows)

pg14sept2011.pix

pg3sept2017.pix

Start Geomatica2017 using the Geomatica2017 local launch icon

This starts the PCI Geomatica software, and opens a focus display window

in **Focus** -> open your copy of **pg14sept2011.pix**

switch the display from RGB 123 to 543 (right-click -> RGB mapper)

1- Bitmaps

Traditional analogue mapping from panchromatic aerial photography uses interpretation and digitising. We'll first simulate a digital process from one band ... this only works if a feature type has a unique set of DNs. Water reflects almost no NIR and thus has very low DNs in Landsat TM Band 4 / L8 OLI Band 5. Extract water only - imagine you don't have a water layer for PG

Click around in the water and note the highest DN for Band 4 (Near IR) - displayed in the green gun. - be sure not to catch any shoreline / mixed pixels. We will use this DN to try to identify ONLY water via the **Threshold** algorithm:

Tools -> Algorithm Librarian -> click Find - and type **THR**
It will find the algorithm and then click Open
Check channel 4 the bottom panel should be checked on 'viewer'

Open the params tab, enter 0 for minimum and a conservative value for water as maximum - go low at first e.g. 5 below your estimated value, so its easier to 'up' it
- click the Log tab and then Run
Some (but not all water should fill with a default colour)

Now up your maximum to a value close to your estimated value and run again
Is all the water now filled ? click on a coloured area and non-coloured area - below it should indicate DN of 1 and 0 (water and land).
if some water areas are not coloured, highlight the layer in TOC (left) and query the water DNs (Band 4).. possibly some water pixels are a bit higher
- maybe you need to increase the maximum a bit more to cover these.
Note it uses a different colour each time, and you can turn off previous attempts

Eventually, you should find a value that hits just water and most of it - if so,
Go back to the first tab (files) in THR
check the box below viewer, hit browse and find your copy of **pg14sept2011.pix**
save -> it has now added a bitmap (1 or 0) = water or land to your file.
NEVER NEVER NEVER accept the useless default of untitled.pix
The .pix file contains all layers - its not like an Esri shapefile
We may use this in a later lab to create a water layer shapefile but it doesn't have to be perfect ...
Check by switching from maps to files tab, and this should list a bitmap as well as the rasters

A bitmap stores only 1 and 0 – in this case 1 for water, 0 for non-water

Now wasn't that easier than manually digitising the rivers and lakes.. !? (but maybe less accurate)

2. Classification Introduction

Classification involves finding a set of **unique** spectral signatures for a feature type in the image scene. We want our classifier to take advantage of maximum *information content* available in the imagery, so we will choose relatively uncorrelated bands to run the classification e.g. 5-4-3 NOT 3-2-1

BAND CORRELATION

You should notice the following for Landsat TM data:

1. VIS (visual) bands show cultural features and in detail
2. NIR shows the land/water boundary sharply. Water appears black
3. MIR shows moisture inversely proportional to DN's (high DN = low moisture)
4. TIR shows basic variations in temperature

*To view correlation between any selected band pair, Select **layer** -> **scatterplot***

Look at the correlation between the following bands. The more correlated the bands, the closer the plotted points fall along a straight line (one band is nearly a linear function of the other).

Green v Red

Green v NIR

NIR v MIR

Blue v TIR

NIR v TIR

MIR1 v MIR2

you should see a high 'r' between visible bands, and less with IR (except 5 v 7) – though it may not be especially clear in this dataset

3. Unsupervised Classification algorithms

You will now see the difference between a band (recorded by the sensor) and a channel – which can store a band, but also any other data generated by the user. We will add 3 more empty layers - 8-bit channels (it already has 2) which can store classification results with up to 256 classes.

You are ready to classify:

- **Analysis -> Image Classification -> Unsupervised**
- Select the file to use (your copy)
- Select New session
- Click on the layers button and add three (3) more 8-bit layers (channels)
- We need to specify the display, input bands and output CHANNEL
- Select TM bands 5,4,3 (R, G, B) as input channels (tick in input column)

- Select the first empty channel (8) as output and Accept
- NOTE: the designated output channel will be overwritten, so if you specify a band number (1-7) you will LOSE the band data - so **always** double check your output channel number - it should be an empty channel

In the Classify window, select these options:

- Algorithm: **K-Means**
- Max class: 6
- Max Iteration: 1
- Min Threshold: leave as is
- Max Sample Size: leave as is (with a bigger scene, we might specify a subset)
- Show Report button: 'on' (depressed)
- OK

.. this shows the report for the 6 clusters, # of pixels in each, and average DNs for bands 3,4,5
 .. image displays in 'PC' (pseudo-colour), the DNs are 1-5 (one number for each class)
 This report could be saved if needed for future work (not needed here)

Can you visually identify the clusters as classes approximately? (do a quick check)
 tick the PC (classification layer) off and on to view the classification and 543 composite.
 View the report also - it gives the number of pixels in each class, and the mean DNs
 It should be poor as there are too few classes (and even fewer iterations)

Right-click the Classification Layer and select run classification

Change the number of **classes to 10** - view the result, its better, but view the classification report
 - likely there are only 6-7 clusters containing most of the data

Right-click the Classification Layer again and select run classification

Change **iterations also to 10** - view the result, now most clusters have pixels

Run it one more time with **16** classes (default):

Right-click the Classification Layer again and select run classification

Change the number of **classes and iterations also to 16** - view the result, there may be an extra class or two with pixels reported

review the report stats: Which is the band (3, 4 or 5) that most differentiates water and forest –
 – it should be a bit obvious, and why ?

Can you match the cluster numbers with land cover types - click each colour to see which cluster it is, and try to visually match these against the list below.

Change the colours for easier reading if you wish- at the very least make the water cluster class blue, and coniferous forest dark green

Do this by right-clicking on classification metalayer-> Edit PCT

This changes the display - later we'll show easy way to merge clusters

Another way is expand 'output' checkmark and double-click on the legend colours

Note that some shadows get grouped with water – the north facing slopes of the esker ridges north of the Nechako River; next week we will see how this can be corrected.

Some cluster classes may be 'hybrid' or mixed, e.g. grassy areas could be regenerating cut areas, city parks, etc.. the bottom 3 clusters may intermix

Jot down which cluster numbers correspond to:

- Water
- Conifers
- Deciduous trees
- Industrial – urban
- Residential - urban
- Grass (e.g. soccer fields)
- Agricultural fields
- Cutblocks

You will likely find the last numbered clusters cover the Canfor chip piles - these have the highest DN's in the TM bands (like snow in the mountains!) and no other similar features

NOTE: this is not an easy image to classify due to the urban area .. but it's familiar ...

Fuzzy k-means

Select **Analysis -> Image Classification -> Unsupervised (.... new session, and pick an empty channel for the output – 9**

Go for 16 / 16 in the clusters and iterations - how does this compare with the previous ?

Does it do any better in mapping the classes above?

What is fuzzy k-means (see the help manual icon)

ISODATA classifier

Select **Analysis -> Image Classification -> Unsupervised (.... new session, and pick an empty channel for the output - 10**

One last time, select **Isodata** as the method. This time, you can give minimum clusters as 10, maximum as 15, and desired clusters as 12. (pick 16 iterations). Again view the result, and the classification report - almost all clusters should have a fair number of pixels (or the clusters are 'wasted'), and compare with the previous classifications

In order to display both or all 3 classifications, **layer->add->pseudocolour**, and select the previous classification channel – the software should have labelled them suitably

Select which you think is the best classification of the ones you've tried for the next step - i.e. 'maps' a given class to your satisfaction.... this is quite subjective - for a project or job, you'd spend more time on this.

Classify once more **Analysis -> Image Classification -> Unsupervised** but also add bands 4 and 7 as input, channel 11 for output and pick 543 as display, but tick 3,4,5,7 for input - you are not limited to 3 input bands, only in visual display. Use your best algorithm (K-means, Fuzzy, or Isodata). Has it done any better than it did with only 3 input bands? (may be hard to determine)

4. Merging clusters

If two clusters seem to be the same land cover class, we can merge them visually at first by giving them the same colour using the PCT editor – later we'll actually merge them. At this point, make sure you are comfortable changing screen colours.

5. Filtering / sieving the classification

The classification will have isolated pixels, which are mostly 'noise'

These can be reduced using the SIEVE algorithm

First record what is the cluster number(s) for water .. this is so you can retain small lakes

The classification can be cleaned using the SIEVE filter, called up from Focus via :

Tools -> Algorithm Librarian

click the 'Find' button and type in sieve and then 'Open'

The parameters will be similar to these:

input = ## (selected your best **classification** channel)

Polygon size threshold= 12 (11 pixels < 1 hectare), why – do the math, each pixel is 30x30m

Connectedness - can be 4 or 8

exclude values list = ## (where ## is the class number for water)

output port should be viewer -PCT (at first)

select log tab and run ...

View the result, compared with the unsieved classification - click the sieve on and off to compare .. try it again now with a 22 pixel (2 ha) threshold if you wish

Now save which one you consider better, - run it again, specify your filename under the 'files' tab in sieve, as output at the bottom - make sure you name your copy of **pg17sept2011.pix**. This is now saved as a new channel in your file - check which one it is, by viewing your list of channels in the main tab on the left of the image, under 'Files' and expand the rasters list.

NOTE: NEVER accept the default filename '**Untitled.pix**'. You must have all your data layers in the same file. and you should not have any file named Untitled.pix in your folders ...

6. Prince George 2013 Landsat 8 scene

Follow the same process in steps 1-5 above, but for this more recent Landsat 8 image – it has the extra band 1 (coastal blue), and 16 bit data (many more DN vales).

* Be careful to select the band combinations you want for 2013, as the bands and channels for 2011 will also appear as options

1. Bitmap – create the water bitmap, of course you'll need a very different threshold vale
2. Classification Intro: try the same scatterplot combinations – are the values any different ?
3. Classification algorithms – no need to repeat everything – just do the one which you felt worked best – no prizes on this, it's an intro lab, not a project or thesis. You'll only need to create one or maybe two new (8-bit) channels for classification. Has the 16-bit data helped distinguish between the water and shadow ?
4. No need to redo this unless you want to ...
5. Sieve – repeat SIEVE for the 2013 image classification – using either 12 or 24 pixels

Assignment questions are designed to enhance your understanding of RS data, display and classification, based on the first two labs. Send your answers either as an email attachment or in the message to: wheate@unbc.ca by Tuesday 25 September (6pm).

1. Which type(s) of land cover have the lowest and highest DN's for the NIR and MIR bands ?
i.e. NIR – lowest and highest DN's; MIR – lowest and highest DN's (this should likely be the same answer whether you use the 2011 or 2017 image).
2. What is the correlation value (r) between Red and NIR for a. the TM2011 and b. OLI2013
3. Why do we usually have to change/renew the enhancement after loading different bands?
4. Why is a 12 pixel threshold used in SIEVE to create minimum 1 hectare areas – show the math. Calculation Could it instead be 11 pixels ?
5. Why do we use 8-bit channels to store classifications even for Landsat 8 images, where the bands contain 16 bit data ?

Next week's lab: Supervised Classification