GEOG357: Remote sensing Lab 2
Unsupervised classification

General Goal:

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

Open your copy of **pg14sept2011.pix** from last week, and display bands 5-4-3 in RGB

1- Bitmaps

Traditional analogue mapping from panchromatic aerial photography uses interpretation and manual digitising. We'll 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 L5 TM Band 4 (L8 OLI Band 5). Extract water only - imagine you don't have a GIS water layer for PG

Click around in the water and note the highest general DN for water in Band 4 (Near IR) – numbers displayed in the green gun below the image - be sure not to catch any shoreline / mixed pixels. Note (and avoid) the surface algae masses in Tabor Lake.

You should also review the histogram for Band 4 to get an idea of the DN range for water. Look for the 'saddle' low in the histogram and its approx.. DN value

zoom in a bit, maybe around the two rivers confluence ("T'enneh").
We will use the DNs to try to identify ONLY water via the **Threshold** algorithm:
Tools -> Algorithm Librarian -> click Find - and type **THR** (then enter)
It will find the algorithm and then click Open
To reduce clutter, you might want to move the algorithm librarian window or close it.
Select your 2011 file and Tick channel 4 .... the bottom panel should be checked on 'viewer'
click on the Input params tab, enter 0 for minimum and a conservative value for water as maximum - go low at first below your estimated value, e.g. 17 so it's easier to 'up' it (there are no DNs below 17)
- click the Log tab and then Run (the screen shows the progress or errors)
Some (but not all) water should fill with a default colour – red at first

Now up your maximum (Input params) to a value close to your estimated value and run again
Is all the water now filled ? right-click on the bitmap, select representation editor and it gives you # pixels that are water (1) and not (0)

If some water areas are not coloured, highlight the layer on the left (in maps tab) 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, but notice the bitmap starts to include shadows on the north side of the eskers north of the Nechako River in the Pidherny area.

IMPORTANT POINT: when you read the DN values at bottom, they apply to whichever display layer you have active – it may not be the top one. Just as in GIS software, highlight the layer you are trying to query to be sure. If the active layer is the image, you see the 3 band DNs; if the bitmap above it is the layer, you will see 1 (water) or 0 (land)

Eventually, you should find a value that hits just water and most of it – if you go too high, it starts to fill in non-water, but there is no perfect DN – either some water gets omitted or some non-water gets included. Record what this value is (it will be an answer you will use later)

Once you have your best number value, Go back to the first tab (files) in THR

check the box below viewer, hit browse and navigate to your copy of **pg14sept2011.pix**

Run again, and this time it will add a bitmap (1 or 0) = water or land to your file.

It is NOT overwriting your file, just adding a bitmap to it

NEVER accept the useless default of ‘untitled.pix’ – it must go in your 2011 .pix file

Check by switching from maps to files in the TOC (table of contents), and this should list a bitmap as well as the rasters; switch back to the maps tab.

The .pix file contains all layers – it’s not like an Esri shapefile, more like a geodatabase (gdb)

Compare the bitmap with the city rivers GIS layer: - load from L:GEOG357\shapefiles\rivers.shp

Turn the bitmap on and off so you can see the vectors; they are not identical as the vectors were produced from higher resolution air photos, and may be more up to date than 2011.

Much easier than manually digitising all the rivers and lakes.. !? - but maybe less accurate

Note that this could be a complete water layer if there were no shadows, e.g. in flattish regions

2. Classification Introduction

Classification involves identifying 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 image, so we will choose relatively uncorrelated bands to run the classification e.g. 5-4-3 and NOT 3-2-1

BAND CORRELATION
You should notice the following for Landsat TM data:

1. VIS (visual) bands show cultural features well and in detail
2. NIR shows the land/water boundary sharply. Water appears black
3. SWIR shows dryness or inverse of moisture (high DN = low moisture)
4. TIR shows basic variations in temperature (though we won't use this)

Ensure you are zoomed out to show the full extent – symbol to left of zoom options or right-click on file name and select Overview of Layer

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), and the higher the r vale (corr coeff.)

Green v Red   r = 0.95

Green v NIR   ?

NIR v SWIR1   r = 0.71

Blue v TIR ?

NIR v TIR ?

SWIR1 v SWIR2 (5 v 7) ?

you should see a high 'r' between visible bands, and less with IR (except 5 v 7)

r values correspond to the image data displayed – values **will vary if you are zoomed in**

Q1a: what is the r value between the red and near-IR bands ? (note it down – you need it later)

Review also the histograms: layer → histograms

This displays the 3 bands in the composite and represent the VIS, NIR and SWIR. Click on each histogram to expand each one at a time: higher SD = higher contrast

The standard deviation value is an indication of data spread (lower for visible), and the bimodal graphs for IR reflect the very low DNs for water. Higher sd = ~ more information.

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 empty layers - 8-bit channels 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 2011 .pix file
- Select New session
- Click on add layer and add five (5) 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 (and tick in input column)
- The colour ticks in the first 3 columns are simply what is displayed
- Select the first empty channel (8) as output
- NOTE: the designated output channel will be overwritten, so if you specify a band number (1-7) you will LOSE the band data - **always** double check output channel number

- it should be an empty channel. Otherwise you can wipe out your imagedata – Don’t let this be you! (but it is recoverable)

- click OK

In the Classify window, select these options:

- Algorithm: **K-Means**
- Max class: 6
- Max Iteration: 6
- 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 new DNs are 1-6 (one number for each class)

This report could be saved if needed for future work (but is 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 the 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 iterations

Right-click the **Classification MetaLayer** and select **run classification**

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

Change **iterations also to 10** - view the result, now most clusters have pixels. This shows the value of using at least 10 iterations. **review the report stats**:

Once more, Right-click the **Classification MetaLayer** and select **run classification**

This time, try the default 16 classes/clusters and 16 iterations - and run classification again. This time, save the report (as a .txt file). This should seem better – click classification Metalayer off and on to compare with the image.

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: expand ‘output’ checkmark and double-click on the legend colours to edit them. [or right-click on classification metacolour – edit PCT]

You can also change colours in a classification -> right-click on the layer name->Edit PCT
Note that some shadows get grouped with water same as in the Bitmap– the north facing slopes of the esker ridges north of the Nechako River; next week we will see how this can be corrected.

Q1b: Which is the band (3, 4 or 5) that best differentiates water and forest — it might be a bit obvious, and why ? (note down your answer) – review the report if you still have it, or query some water and forest locations for their DN values.

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

See which cluster numbers correspond to:

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

Simplified interpretation:

- Dark-green: coniferous.
- Light-green: deciduous.
- Pink: arable / agricultural.
- Yellow-green: Playing fields / golf courses
- Dark-purple: industrial.
- Light-purple: residential.
- Black: deep or sediment-free water.
- Blue: shallow or sediment-laden water.
- White: bright: Canfor chip piles (brightest feature on this image)

Fuzzy k-means

Right-click the **Classification MetaLayer** and select Configuration
Change the output channel to 9 (input = 543 as before)
Make sure to click on the Fuzzy K-means button this time - Go for the default - 16 / 16 in the clusters and iterations - how does this compare with the previous K-means?

Note that it runs slightly slower because the clustering process includes the spread of pixels in each cluster and probability functions, not just which seed point / mean each pixel is closest to.

Does it do any better in mapping the classes above? Hard to tell with different colour tables. You can reload the K-means result for comparison using Layer-Add-> pseudocolor -> channel 8, and then click the top display layer on and off to se the one below.

ISODATA classifier

Right-click the **Classification MetaLayer** and select Configuration
Change the output channel to 10 (input = 543 unchanged)

Now select **Isodata** as the method. make all options the default 16 minimum clusters, maximum, and desired clusters and 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. Save the report

If you need o display previous 3 classifications, select **layer->add->pseudocolour**,

and select the previous classification channel – the software should have labelled them suitably. Click the fuzzy / isodata layers off /on to compare with the underlying colour composite … it may be hard to objectively tell, but for some the isodata slightly retains homogenous areas within the same class – not very scientific, but hey it's only a lab. Note especially how well each classification sorts water from forest, and coniferous forest from other; also details in the bowl.

Select which you think is the best classification of the ones you've tried for the next step - i.e. 'maps' given class(es) to your satisfaction.... this is quite subjective - for a project or job, you'd spend more time on this. Check for example which one also separates the residential areas in the bowl versus the light industrial areas (more purple in the 543 composite).

Classify once more: Analysis -> Image Classification -> Unsupervised

but this time we'll add bands 2 and 7 as extra input as some people advocate using all bands,

tick channel 11 for output and pick 543 as display, but tick 2,3,4,5,7 for input - you are not limited to 3 input bands, only visual display. Use your best algorithm (K-means, Fuzzy, or Isodata). Has it done any better than with only 3 input bands? May be hard to tell – I didn't think so. Some researchers also think too many inputs can 'confuse' the algorithm.

We can explore classification accuracy and filter or ‘SIEVE’, but we’ll leave those to next week

5. Prince George 2013 Landsat 8 OLI image

Copy the file **pg20july2023.pix** from L:\GEOG357 into your geog357 folder

The Landsat 8 Operational Land Imager (OLI) has 16-bit data … DNs can range from 0-65,535 (but rarely do) Since 2015, both NASA Landsat and ESA Sentinel programs have 16 bit data. It can behave differently to the 8-bit data from Landsat 5 in several ways as we'll see.

We could add the newer image to this 'project' but it's less confusing to start a new project – you can save the other one as a project e.g. pg2011 (.gpr)
File-> New project and open the two 2023 image files

Display the 2023 image, using the optimal SWIR, Near IR, Red combination – it has the extra band 1 (coastal blue), so the optimal combination is 654 instead of 543 for Landsat 5.

1. Bitmap – create the water bitmap same as you did for the 2011 image, of course you'll need a very different threshold value – *Q2a: record this value*

2. Classification Intro: try the same scatterplot combinations as with the 2011 image– are the values any different ? *Q2b: What is the r value for Red versus Near-IR ?*

Check the 654 histograms for 2023 and review the range of DNs for the near-IR band
How do you interpret the double minor modal peaks for low DNs in the histogram ?

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 new **16-bit** channel for classification. Note: really we only need an 8-bit layer as we are unlikely to generate >256 classes/clusters. But if we do that the software, inserts it ABOVE the band in the file, so Band 1 becomes channel2, Band2 becomes channel 3 etc..
This is very annoying, and I'm not sure why they have not fixed this.

Has the 16-bit image data helped better distinguish between water and the esker shadows ? and if so, it might have been the 16 bit data, or the higher sun in July- I think it may be more the latter reviewing a later season file from September 2023.

You can visually compare the 2011 and 2023 classifications though their equivalent clusters may be different colours; you could spend 10 minutes recoloring one to match the other and also 'merging' two clusters where they seem to be the same feature – coniferous perhaps ? Of course the other variable here is the changes between 2011-2013, which we'll examine further in the course – not necessarily in PG. You can save any colour edits in your 'save project option'' as well as when you do the edits (during the edit PCT option).

Lab assignment 1 (5%) – along with next week's lab on supervised classification – store your answers from Q1-2 above and include them with the questions posed next week.

Next week's lab: Supervised Classification