GEOG357 Lab 3: Supervised classification

Learn the supervised process and take control in the driver seat.
We'll do it for the 2011 and a newer cloud-free 2023 PG scene
Submit assignment as one .pdf (or .txt ?)

0. Start

In Catalyst, Open your copy of the Landsat 5 TM image: **pg14sept2011.pix**
Change display to bands 543 and enhance (root)
Expand the image display so it doesn't need panning – but you don't need to maximise …

Check your pix file for bands/channels:

Change to the 'files' tab and click the + next to rasters; this will show your channels – you should have bands 1-7 and then 5 unsupervised CHANNELS (not bands) from last week and maybe a sieve, plus one or two bitmaps (which we won't use)
Switch back to 'maps' tab.

1. Supervised Classification session: Prince George 2011 image

Unsupervised classifications are quick, supervised take longer due to the need to identify cover types and create training sites. I'm picturing 10-15 classes for the image: e.g. Water, wetlands, Coniferous, Deciduous, Mixed forest (?), Fields, Cutblocks, Industrial, residential etc..

A. Steps to start supervised classification:

- Analysis -> Image Classification -> Supervised
- Select your 2011 file **.pix ,** click on 'open'
- Click on New Session .. then 'add layer' button (if you need to)
- You will need 3 empty 8-bit channels - unless you already have empty channels
- Display channels should be 5,4,3 (RGB)
- Input channels should be 2,3,4,5,7; tick in Input channel column *
  I chose these five for 2 reasons: so you realise you are not limited to 3, and because the r value between bands 2-3, and 5-7 is 0.90, so there is some contrast between each pair. Check if you don't believe me – right-click on filename -> scatterplot
- Set the training channel to an **empty** channel number (tick in that column)
- Set the output channel to another **empty** channel (tick in that column)
- Don't tick training or output in any channel with data already or you will lose it.
- OK

'Training Site Editing' box appears ..

2. Training Areas: *In the Training Site Editing box*

- Add a new class, (class->new) click on its name and change it to water,
- In the main focus window (directly below View), pick the 3rd icon in the 2nd row,– it looks like a zig-zag ‘new shapes’, and from the dropdown, pick -> raster seeding.

The raster seeding window controls its functionality with two tolerance settings
These are currently set to 10 and 1X, upping either increases the deviation that is allowed in including pixels adjacent to wherever you might click. For now leave these as is and don’t change the 1X anyway). Zoom into the image with water and click on the water, I suggest the confluence (T'enneh) of the two rivers, but not too close with the zoom.

Throughout seeding, the 'raster seeding panel' MUST be visible onscreen (or you can't seed)
Ensure that the training areas (channel) is highlighted in the table of contents - seeding can be picky. If you lose the raster seeding window, get it back from the zig-zag option e.g. it disappears if you use the rectangle area zoom in (left of + - buttons)

Select 2-3 seed points for water - you will want to cover clear water (Nechako R) and silty water (Fraser); also pick a lake e.g. Tabor ... you could change training site colours, but DON'T PICK BLUE for water or you can't see your sites - the defaults are usually OK.
The goal is to find a sample of water pixels, not to 'fill' the river:

When done each class, 'save', add the next class, and repeat seeding process by class. You may need to modify the tolerance values for other classes, e.g. reduce 10 to 5 or 7 for ‘residential’

Oooops- what happens when ... you added too much, picked a seed point that spread too much, added something to the wrong class etc.
There are two ways of dealing with this, the easiest is to remove/clear that seed and start over...

- In the training site editing window highlight the class with the mistake
- Select Edit – Clear Selected

You can also use this to remove the entire class e.g. class not intended

- Select Edit – Delete Selected

Usually pick 2-3 seed points for each class training 'set' - and view for different DNs for the same cover type, e.g. due to different illumination (more important in the mountains). For fields and cutblocks, try to pick sites that represent the range of possible values (screen colours). There is overlap between these two classes and also variations due to cutblock regeneration.

Very important: when done the sites for each class, hit ‘save’ then go to class -> new and rename this to a class, then start selecting again – make sure the class you intend, is highlighted.

Add / consider these classes for your classification

- Water
- Wetlands (NW portion of the image by Swampy Lake)
- residential (zoom in to find houses/gardens) – I changed 10 to 5 or it took too much
- industrial / concrete (maybe change tolerance back to 7 ?)
- cutblocks – regenerating
- Recent cutblocks (e.g. from pine beetle)
- agricultural fields
- Parks ? very tricky as they all look different, but often the same as fields, maybe don't do it !
- deciduous forest
- coniferous forest – maybe 5 tolerance (include shadowed eskers slopes and sunnier slopes)
- mixed forest – this class can be tricky – only if you think you can see 'medium' greens
- bare ground – roads, runway, and chip piles
- Gravel (river edge) – may look similar to 'bare ground'

Any others you can identify ? – remember, you can merge classes afterwards

When you've done them, in the **training sites window**, check **tools -> signature separabilities** for suitability of classes and training areas - ideally each matrix value is > 1.8
If some pairs are very low (below 1.5) you can either delete one or merge the pair if it's logical Class-merge (and I think the method is obvious)

- but this is only a lab exercise, so no need for excessive rigour

Q1a. Save the signature separability report, as a .txt file
1b. comment on low separability values – and if 'easily' fixed, try it e.g. merge the two, or change training samples – otherwise comment on why they have happened

Tools-> classification preview
shows how it will look .. go back if it looks doomed and make edits as needed e.g. new classes

You have these options - try them all as they are previews and show immiediately: Click off the classification layer now and then to view the 543 composite underneath to check how they compare with the land cover types

- Maximum likelihood
- Maximum likelihood with null class (null = black) – what is missing, do we need more ?
- Parallelepiped (mine looked awful !)
- Parallelepiped with MLC tiebreaker (mine looked OK !)
- Minimum distance

Remember that urban classification is challenging due to the mosaic of many small features – and that some features look similar e.g. agricultural and sports fields, and some cutblocks.

3: Run the Classification

Pick the best one based on observation of areas you know – I was watching for the eskers area – did one or more avoid confusing the esker shadows with water ?
As per lecture notes, one might expect the best to be the MLC option .. but then maybe not ?

You should preview how this compares with the RGB bands, by toggling the classification on / off and check areas you know e.g. has it cleanly separated water, which class contains UNBC, how about the residential versus industrial areas in the bowl ?
When you are done, select Save and Close in the Training Site Editing Window.

right-click on the **classification metalayer** in the 'maps' list and select **Run Classification**.
Select your preferred classification. Make sure the show report button is ticked. Note the 'use bitmap as mask' option – we won't use it here to restrict a classification to only part of the image.

Q2a. Save this report as a .txt file
... use copy/paste to only retain down to the line with Kappa value - exclude training site info
2b. comment briefly on the least and most confused classes with possible explanation why

I**n the report window** which classes are most confused (in the confusion matrix) ?
- why are these classes confused and not others? (maybe none!). Note that the accuracy shown is inflated as this is only comparing with training areas, which should mostly be correct.
Did any classes get 0 pixels – if so, why – maybe you need to add another training area

The **confusion matrix** gives a rough guide on the classification based on the training sites
Save the report as a .txt file so you can refer to it if need be after the report window is closed.

Reference data listed in the columns of the matrix represent the number of correctly classified training samples. Errors of omission are represented by the non-diagonal **column** elements, and errors of commission are represented by non diagonal **row** elements.
The Accuracy Statistics report lists statistical measures of overall accuracy and per class.

Run a second classification type into the empty 8-bit channel.
Right-click on the **classification metalayer** and select configuration **–** change the output channel to the empty channel, so it doesn't overwrite the previous one.
Then run the classification again using your second preferred method.

Q3a again save the report text file (down to the Kappa value)
b. briefly comment on how this compares with your first classification – do either (or both) help distinguish the dark slope shadows versus water ?

4. SIEVE or FMO

SIEVE parameters : I suggest sieve to 1 acre minimum (4 pixels) for an urban area
input = ## (classification channel) .. note – not the training channel
Polygon size threshold = 4 (for 1 acre minimum), or 12 for a hectare...
Connectedness - can be 4 or 8 (try both and compare) –
exclude values list = ## (where ## is the class number for water - preserve small lakes)
output port should be viewer - PCT
select log tab and run ...

View the result, compare with the unsieved classification (tick the sieved layer on and off)
If happy with the result, Re-run the sieve putting output to your pix file – use BROWSE button - remember everything goes into the 2011 pix file, don't put it somewhere else … Run

After this is done, view your raster layers (files tab) listing to check you have a sieve layer – It may not labelled other than 'contents not specified' .. if so, go to the files tab, double-click on the name and change to something like 'sieve to 1 acre'. [you can also rename under the 'maps' tab but this only changes the name onscreen- the channel remains unnamed if you close it]

Compare with your unsupervised classification:
Layer->add->pseudocolor- and select your 'best' unsupervised channel from last week –> finish
It's not easy to compare as the class colours don't match, but you might find more 'order' in today's result as you had more control over the classes. You are the boss !

- you could also or instead try tool FMO if you preferred its look from last week.

FOR REFERENCE ONLY: Training site steps: the non-seeding (polygon) method
Click in your chosen area

- Zoom in maybe to +4 or even +8
- Select 'Trace and Close '
- Draw your training area (holding the left mouse button down- don't 'click') in a polygon
- make sure you draw a closed polygon; make it fairly large within the feature
- Select 'Fill'
- Click inside your polygon: it should fill
- You could add another training site, if you feel this one has not captured a good cross-section of the class (keep it simple at this point)
- Type in the class name (water) in the description column
- Save
- New

Why ever use this – maybe if there are large homogenous area cover types – not in cities !

5. Landsat 8 OLI image data: pg20july2023.pix

We should see how the 16-bit data from Landsat 8 OLI behaves in comparison – and enough with that cloudy 2013 image data – the clouds and shadows screw up classification.
.. and also an earlier date 20 July is closer to the solstice with lesser topographic shadows.

Copy to your folder from **L:\GEOG357\pg20july2023.pix**
You can also copy the Pan file: **pg20july2023pan.pix**
open the multilayer file in Catalyst and change display to 654 (SWIR-NIR-Red) – enhance
you can also open/display the Pan file if you wish

Compare with 2011 by clicking the top layer on and off:

- Overall 2023 is greener as its July and vegetation is greener versus September
- Pine beetle cutblocks have also greened up
- The pine beetle block opposite the NSC has been cleared in 2015
- Boundary Road was built west of the airport to bypass PG from Quesnel to Jasper

Repeat the classification process for the 2023 image. When you start the process, add three more layers for training and classification – they should be 16-bit unsigned. Technically they only need to be 8-bit as we won't need more than 256 classes, but the dumb software inserts new 8-bit layers above the bands in the file database, so Band 1 becomes channel 4, and all the way down.

Input: first check the scatterplots for bands 2 v 3 and 6 v 7 … quite high r values, so we can go with just bands 4-5-6 for input; use the new empty channels for training and the classification – they should be 8 and 9

Seeding: Note that the **two tolerance settings in the raster seeding step will need to be <u>much</u> bigger**. Sadly, the help does not help – experiment – we've found tolerance at 50 and 4x for water, 50 and 16x for the other classes seem to work. If the raster seeding window disappears, you may need to reset these values.
- and you may find the same algorithm from 2011 seems to work best, or maybe not ?

Try a second classification to see if it's better – remember to use 'configure' / change the output channel

Q4a. Do one classification and save the report as yourname2023report.txt
b. Briefly describe any differences compared to the 8-bit data values for 2011

Q5. Where do you see the most confusion – which classes and could this have been avoided ?
5b. Does the higher sun = less shadows and 16 bit data seem to help with the eskers shadows?

SIEVE or filter - algorithm FMO as you did last week as there are likely many isolated pixels
Comparing 2011 and 2023 might be interesting but you'd need to match the class colours

Footnote: selecting an area of interest and running a detailed unsupervised and supervised classification could the basis for a suitable project in the last 3 weeks of term, though not the only type – see future labs as well ..

Repeat of Lab 3 assignment questions

Q1a. Save the (2011) signature separability report, as a .txt file
b. comment on low separability values – and if 'easily' fixed, try it e.g. merge the two, or change training samples – otherwise comment on why they have happened

Q2a. Save the first classification report as a .txt file
submit only down to the line with Kappa value - exclude training site info
b. comment briefly on the least and most confused classes with possible explanation why

Q3a save the 2nd classification report text file (down to the Kappa value)
b. briefly comment on how this compares with your first classification – do either (or both) help distinguish the dark slope shadows versus water ?

Q4a. Landsat 8 OLI 2023 classification (only down to Kappa value)
b. Briefly describe any major differences compared to the 8-bit data 2011 and September-July

Q5. Where do you see the most confusion – which classes and could this have been avoided ?
b. Does the higher sun = less shadows and 16 bit data seem to help with the shadows or forest?

Please submit as one .pdf (not .doc) … include the 4 text tables for Q1-4 within … try to format so they don't take extra space – ideally each matrix would be retained as one table, not split into two. Possibly the whole submission could be plain text …. ?

Don’t do this section – for reference only and possible use in projects.

Appendix. Full Accuracy Assessment (time consuming and picky-tricky)

How could you more fully assess your classification - you'd need some independently derived 'ground truth' plots to test against: (these would usually be generated from ground work or other knowledge). The purpose of the accuracy assessment is to create a report which indicates the accuracy of classification results compared to the raw image data. It compares what is assumed to be correct with an image classification based on pixel groupings.

Select Analysis dropdown -> Image classification -> Supervised-> Post Classification Analysis-> select "Accuracy Assessment".
Select your file (yet again)
In the new accuracy assessment window, click on ‘select classified image’
And choose the **Sieved** results of the Supervised Classification
Select "Load Reference Image" and load channels 5, 4, 3. … OK
This image will be used during the random sample selection.

Generating Random Samples

Select "Generate Random Sample" and increase the sample number to 30.
Leave the "Sample Control Options" section to the default answer, "yes". Press "Accept".
In the new Accuracy Assessment Panel, the idea is to highlight the first point in the sample list and observe where it is located on the image (the cursor will automatically be placed at this location). You should likely zoom in to both find the cursor and identify the class type (as best you can). Compare this pixel to the list and select the class you believe it should belong.

Ideally the pixel is in a relatively homogenous (not mixed) area .. but if this is not the case, you can ignore/reject that pixel. Once the random sample you are working with has been highlighted, you can simply select the class it belongs to (click on that class) and the information automatically transfers to the random sample list. This works the same as the "Transfer" button.
Samples could also be taken from vector segments e.g. verified vegetation polygons.
The user is given the option of selecting the channel and the class attribute associated with it.

Produce Accuracy Report
Once all 30 samples have been assigned, select "Accuracy Report" then 'generate report'
The results of the accuracy assessment are shown in the form of a Sample report Listing, Error (confusion) matrix and as accuracy statistics. Don’t be discouraged by low values – in a project/thesis you would take many more samples to get higher stats – at least 30 per class.
The Sample Report Listing can be used to determine which samples were classified correctly.
You would save this report as a .txt file in your folder for later reference.