

GEOG432: Remote sensing Lab 3

Unsupervised classification

Goal:

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

Login

Start Geomatica2015 using the Linux local launch Geomatica icon
This starts the PCI Geomatica software, and opens a focus display window

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

Copy **pg14sept2011.pix** from `/home/labs/geog432/data2016` into your geog432 folder

in Focus -> open **your** copy of the 2011 Landsat 5 TM image
switch the display from RGB 123 to 543 (right-click -> RGB mapper)

note the mountain pine beetle areas (brown), and within them logged (2005-2011)

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.

> **add a grayscale layer and display Band 4.**

Layer-> add-> grayscale-> and pick band 4 (NIR)

> Click around in the water to find the typical range of values
- note down the maximum value that you consider to be water

We will now create a new 'BITMAP' layer ... 1-bit which holds values of 1 or 0 that will correspond to water and not water.

> switch the map 'tab' to files

> Right-click on the filename and layer->add->bitmap (it will be #2) - this is now created

Now add this to the display as below (it will be blank initially)

> switch tab back to maps

> **layer-> add -> bitmap** (pick the new bitmap just created)

Now to pick all DN's you think are water .. and start with a low estimate, for example if you found 25 to be maximum, start with 10 and work up; we use a simple scripting language 'EASI modelling'

> Select **tools -> EASI modeling**

In the EASI window, first select your file on the drop down (it's annoying it doesn't automatically do this); type these lines below in the EASI box

```
If (%4 < 10) then
```

```
%%2=1
```

```
endif
```

** This says if the DN is less than 20 then turn that pixel 'ON' in the bitmap (%%2)

**Note that % is the code for a channel (band) and %% is code for a bitmap

Now click '**Run**' ...

You should see some pixels turn blue, but not all water (because 10 is too low), change the **10 to a higher number** (e.g. 15) and click RUN again .. repeat this process until you get all the water. You will find it will also start to pick up other dark features, perhaps shadows. No worries, it's just an experiment.

If you mess up, you can do this ...

- a. save the EASI file;
- b. clear the bitmap display by typing %%2=0 (run) and then
- c. load the saved EASI file.

Don't spend too long on this, it's just a demo on the language and how it might be used
Note that if you perfectly capture a feature you want, you can save it as a new channel.
- Proceed now to real multispectral classification ...

Turn off (or remove from display) the bitmap and band 4 displays

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 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 water in detail
2. NIR shows the land/water boundary very sharply. Water appears black
3. TIR shows variations in temperature (mostly).
4. MIR shows moisture inversely proportional to DN's (high DN = low moisture)

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

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

2 v 3

2 v 4

4 v 5

1 v 6

4 v 6

5 v 7

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

3. Unsupervised Classification

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 need to create some empty channels to contain classification layers:

Switch from **'maps' tab to 'files'**

Right-click the filename and **new -> raster layers ..** and **add** three (3) - 8 bit layers

Expand the check box next to 'rasters' to ensure you now have 3 empty channels

Switch tab back to **maps**

You are now ready to classify:

- **Analysis -> Image Classification -> Unsupervised**
- Select the file to use (your copy)
- Select New session
- We need to specify the display, input bands and output band
- Select TM bands 5,4,3 (R, G, B) and as input channels (tick in column)
- Select the first empty channel (8) as output and Accept
- NOTE: the designated output channel will be overwritten, 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: 5
- 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 giving the 5 clusters, # of pixels in each, and average DN's for bands 3,4,5

.. image displays in 'PC' (pseudo-colour), the DN's are 1-5 (one number for each class)

Can you identify the clusters as classes approximately?

tick the PC (classification layer) off and on and alternately view the classification and 543 composite.

View the report also - it gives the number of pixels in each class, and the mean DN for the input channels selected

It should be very 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, it's 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

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

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

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.

Note some cluster classes may be 'hybrid' or mixed, e.g. grassy areas could be regenerating cut areas, city parks, etc..

Jot down which cluster numbers correspond to:

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

You will likely find the last two classes covering the chip piles at Canfor - these have the highest DN's in the TM bands and no similar features to be confused with

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 ?
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 10 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 classification

In order to display both classifications, **layer->add->pseudocolour**, select the previous classification channel

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.

Try it once more but with all TM bands (except 6) as input using your best algorithm (K-means or Isodata); pick 543 as display, but tick 1,2,3,4,5,7 for input - you are not limited to 3 input bands, only in display.

4. Merging clusters

If two clusters seem the same land cover class, we can merge them, with EASI modeler
For example if we found clusters 3 and 4 in channel 8 to be the same, we could type:

If %8 =3 then

%8=4

Endif

This would change all pixels with DN 3 in the classification to DN 4 and 'join' / merge them with those already with 4 (it could also be done in reverse, merging DN4 with 3).
Try the merge one way or the other – this is only a lab; it won't matter if its incorrect.

5. Filtering / sieving the classification

The classification will have isolated pixels, which are mostly 'noise' (tuesday's lecture).
These can be reduced using SIEVE
First record what is the cluster number(s) for water .. this is to retain small lakes

SIEVE filter

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 **classification** channel)

Polygon size threshold= 11 (11 pixels = 1 hectare),

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 threshold.

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.

NOTE: NEVER accept the useless default filename '**Untitled.pix**'. You must have all your data layers in the same file. I will patrol user accounts, and you should never have a file named Untitled.pix in your folders ...

6. McBride Landsat 8 scene

Follow the same process above for the mountain scene - **mcbride2014.pix**
- note that channel 8 is a 'dummy' layer to preserve the band sequence, as Band 8 (PAN) cannot be included in the PIX file as it has twice the spatial resolution

Copy the file to your folder

Open it and display bands 6-5-4

Add 3 new 8-bit raster layers

Check the scatterplots here for the same band combinations as in section 2 – are these any different ?

Run **K-means** (16-16)

How does it compare with the PG urban scene in cluster clarity ?

Change the water cluster class to blue to make it easier to view

You may be able to recognize at least these classes:

- Water
- Bare Ice
- Snow-covered ice
- Bare rock
- Agricultural
- Deciduous vegetation - alpine meadows
- Deciduous - avalanche slopes
- Coniferous trees

You may see that some clusters could be merged or that some types are misclassified;

Run again using the **Isodata** algorithm

Experiment also using all Bands 1-7 as input

See whether one of the k-means or isodata channels appears to give the better result by checking the water or other features you can see, and run SIEVE on the best result

Flip between the 654 composite and the sieved classification to check how useful this classification might be ... if you like the result, we can vectorise it in next week's lab ..

When you are done with the PCI Linux Server, **logout** properly - don't just hit the x as this leaves it running and affects system performance for other users

Time to spare - check the earthexplorer website to download georeferenced JPG images for the environmental change exercise - see that link on the syllabus page

Next week's lab: Supervised Classification