Unsupervised Classification

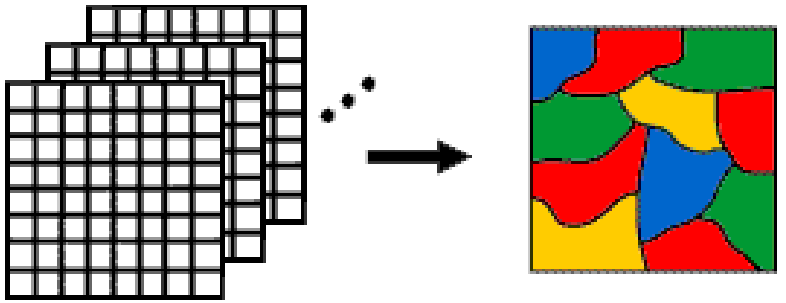
Classification = simplification, mapping

The early promise of satellite imagery: (1970s-80s)

A. Rapid map updating

B. Semi-automated mapping of ‘Land Cover’

- avoid manual digitizing … by classifying multispectral band data



We don’t need a million different pixels
They can be grouped into ‘n’ classes

Manual digitizing (yawn …)

Digitising is usually required for single band (Pan) imagery

e.g b/w photography

BC VRI (vegetation resource inventory)

BC TRIM data layers

7027 x 1:20,000 tiles

All our federal map sheets

NTS 1:50,000 example

All federal NTS map Sheets (13,370) were created from Air photos

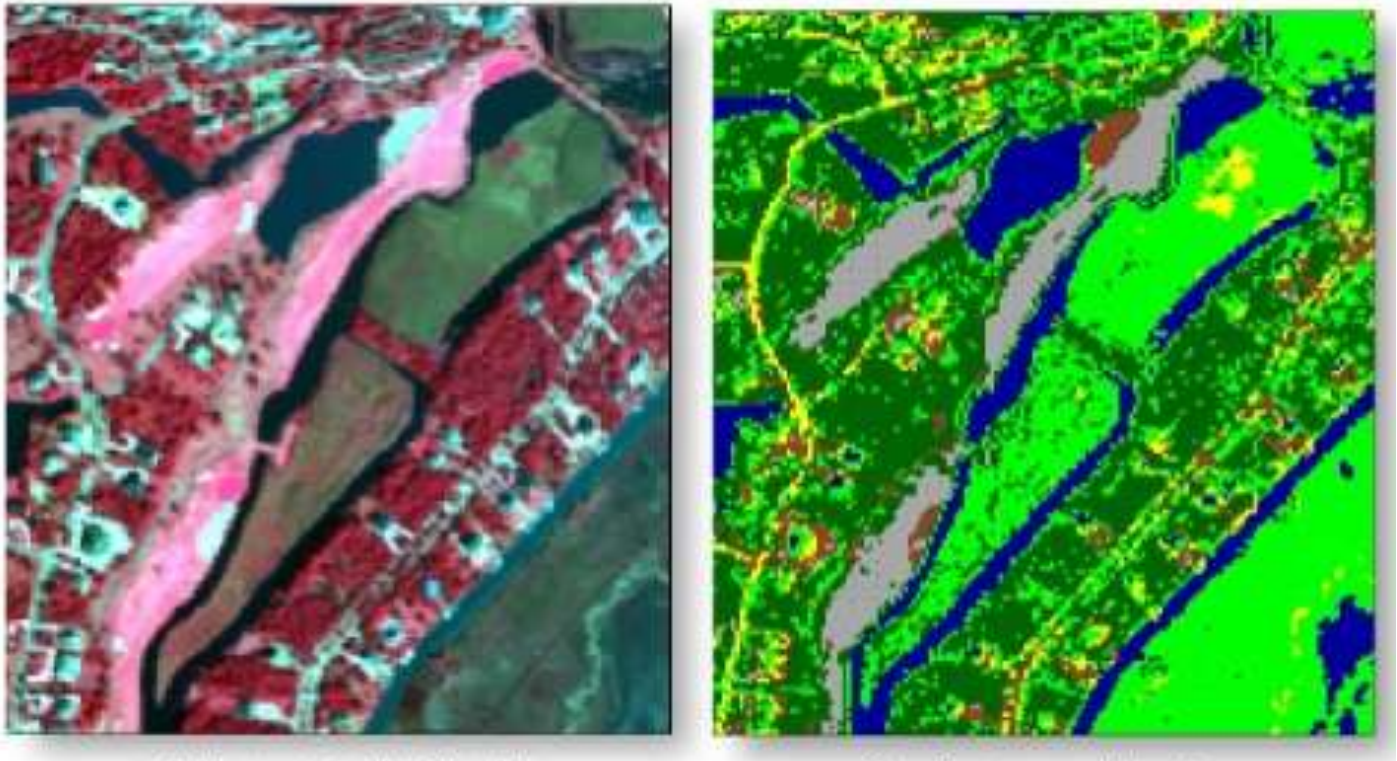
Human interpretation / classification relies on attributes such as:
Shape, pattern, texture, shadows, size, association, **tone, colour**

Algorithms mostly use Digital Numbers (DN) = digital version of tone (single layer) or colour – multispectral layers

Remote Sensing Classification

- Automated grouping of similar pixels using multispectral DNs
- Software was developed following 1972 -> (Landsat 1)
- Digital alternative to manual mapping of Land Cover

Information Classes Derived from an ISODATA Unsupervised Classification Using 10 Iterations and 10 Mean Vectors of an Area Near North Inlet, SC



a. Color composite of HyMap data

c. Classification map derived from

Land Use v Land Cover (LULC) e.g. parks

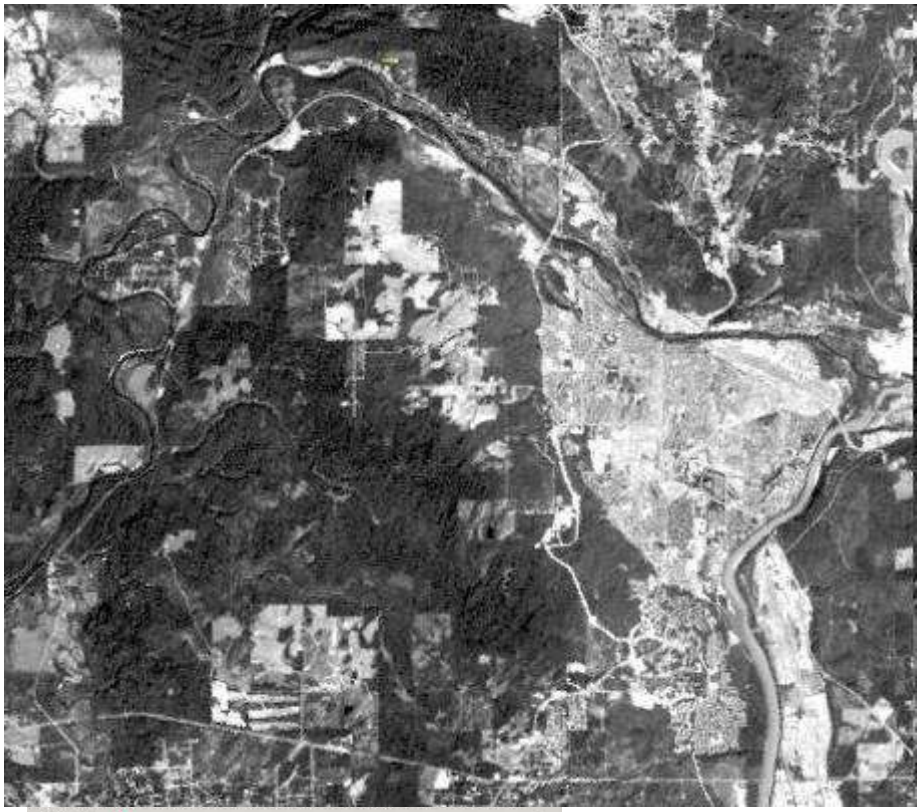
Sugarbowl-Grizzly Den

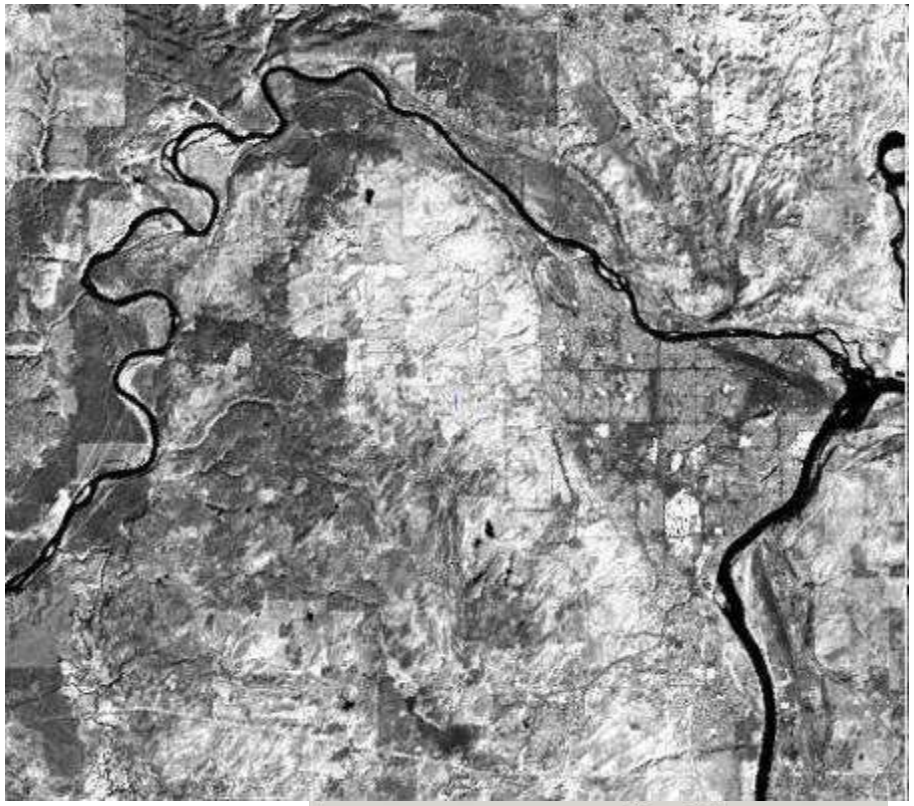
Bowron Lakes

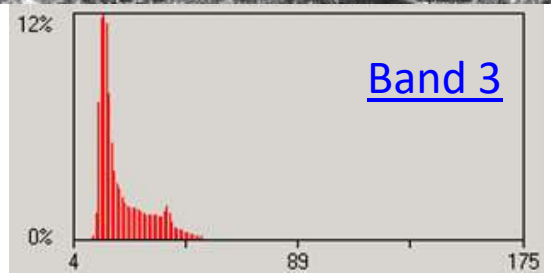
Mt. Egmont / Taranaki, NZ

Can we use just one band to classify ?

One image band could be treated as a monochrome air photo (as in interpretation)
Digital Numbers from one band alone are rarely enough – features are not unique

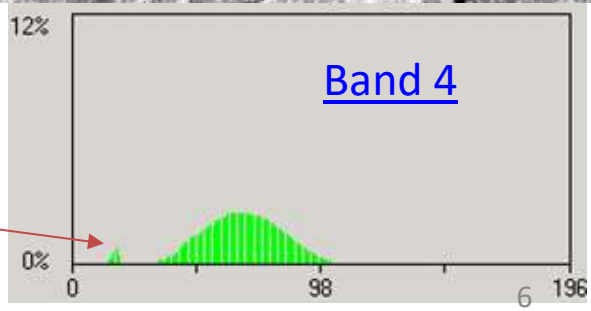


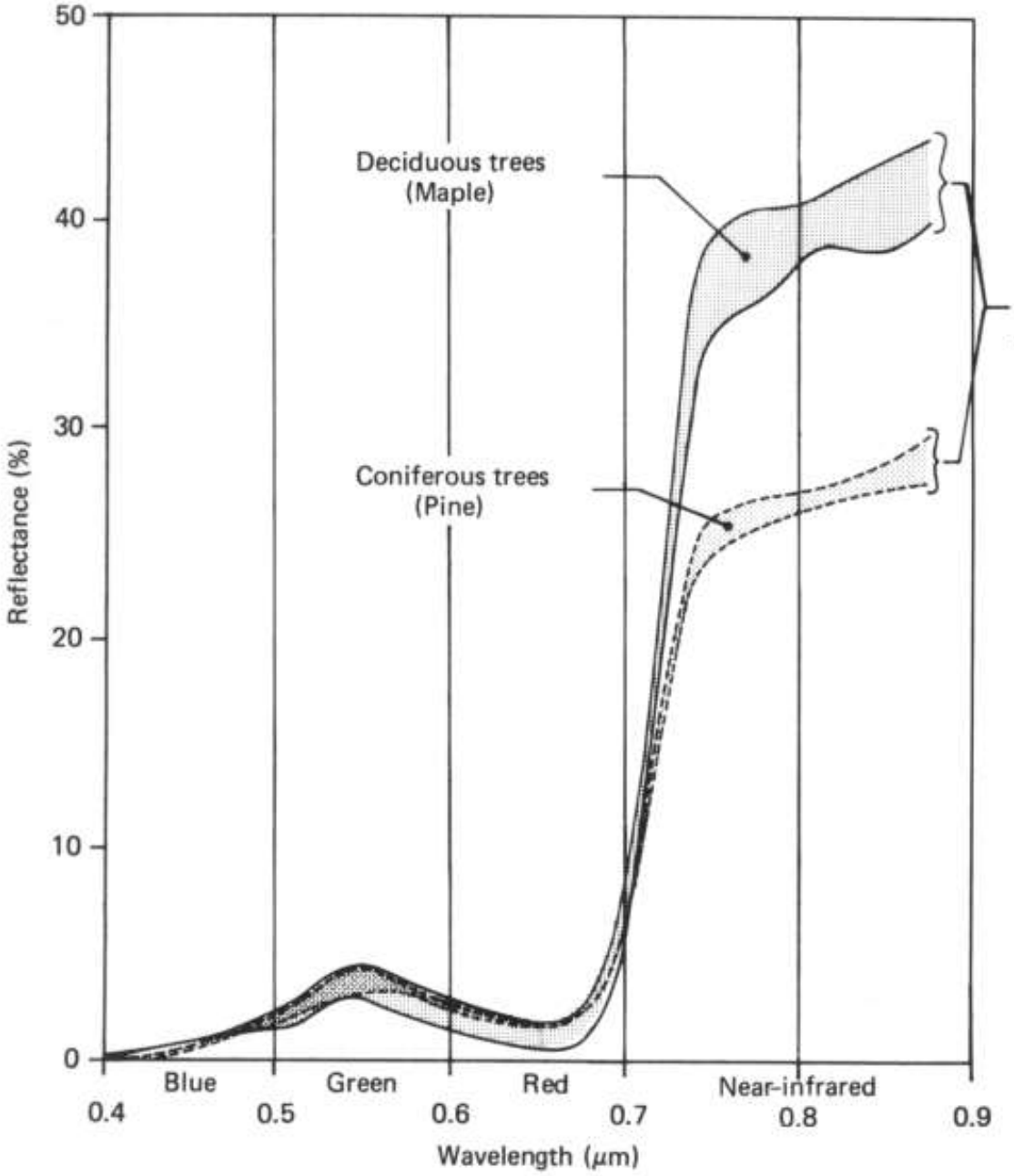




Landsat TM

Water is almost mappable in IR bands (low DNs)



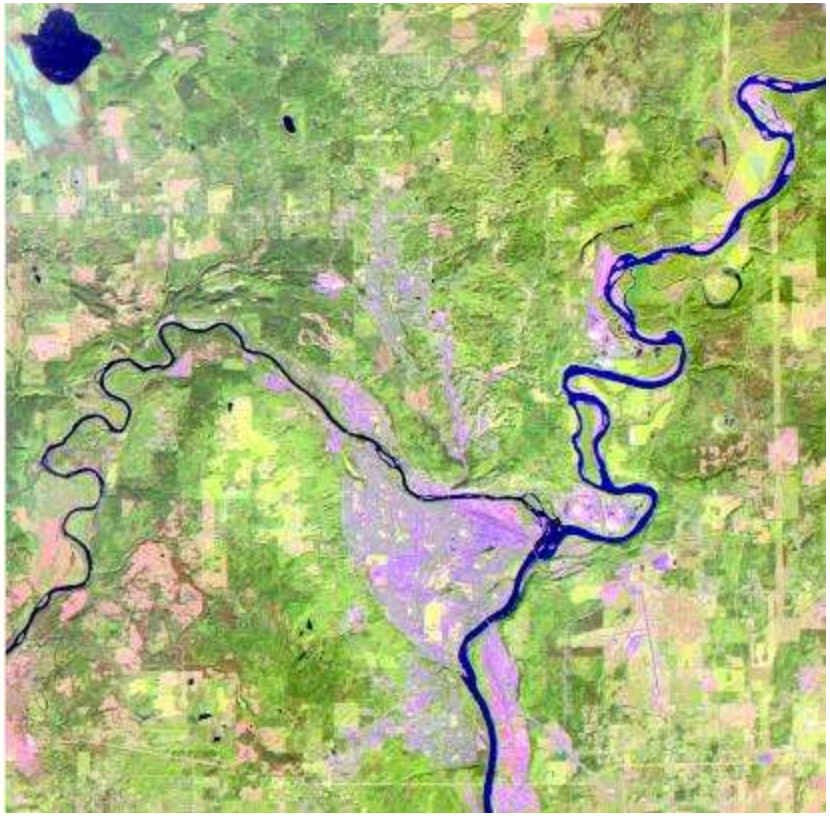


Are the vegetation types different enough in IR ?

Broadleaf/deciduous have greater surface area than coniferous needles – in spring/summer and therefore reflect more especially in near-IR

The visible bands are similar to each other

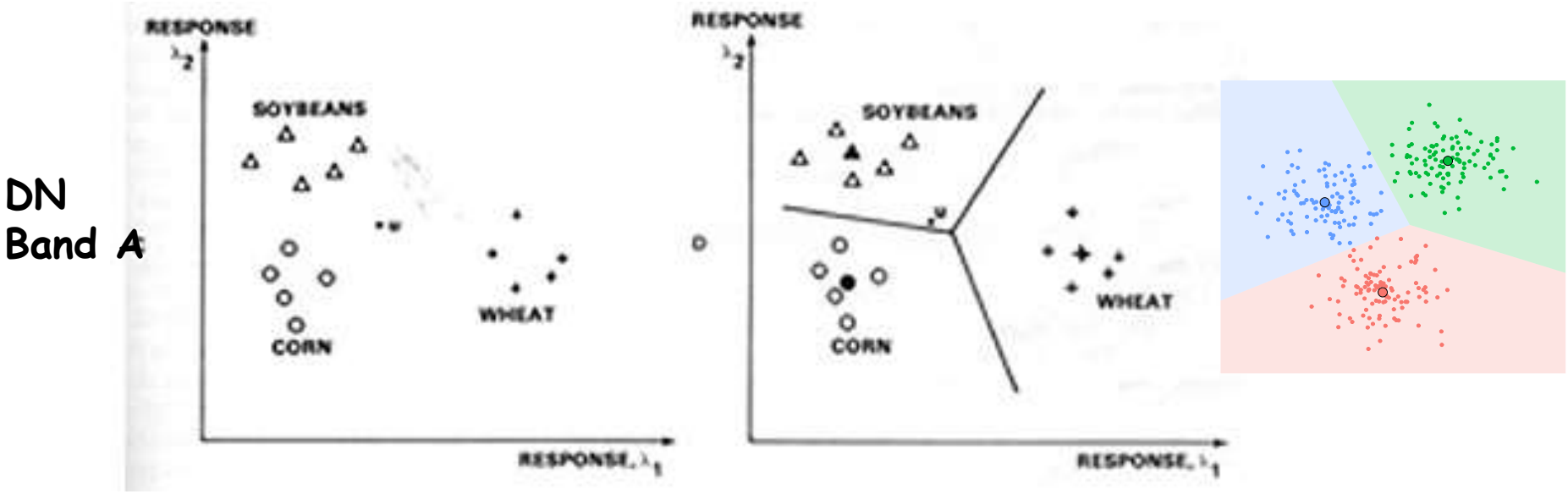
Band / channel selection
e.g. Thematic Mapper TM: 1-7 or Operational Land Imager OLI 8/9 (+TIRS)



Landsat TM has 7 bands: You would NOT select 3 visible bands to classify
The visible bands are similar – so the composite is low in contrast (left)

The role of multispectral sensing in classification

(fuzzy textbook figure)



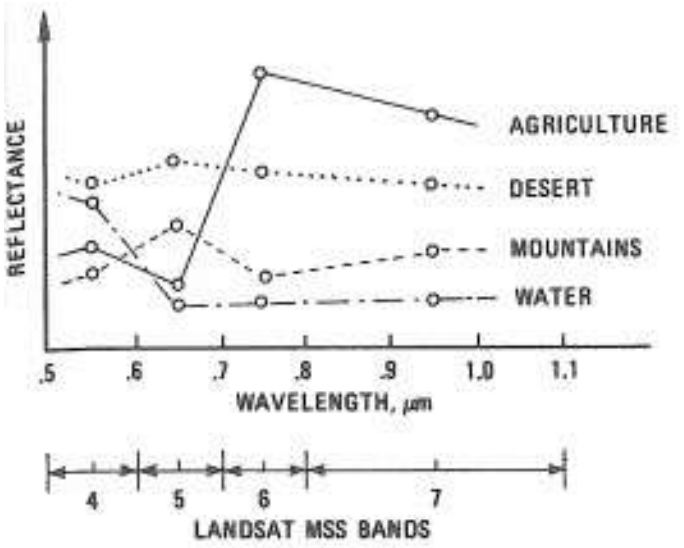
The value of using multiple bands

DNs in Band A are similar for corn and wheat

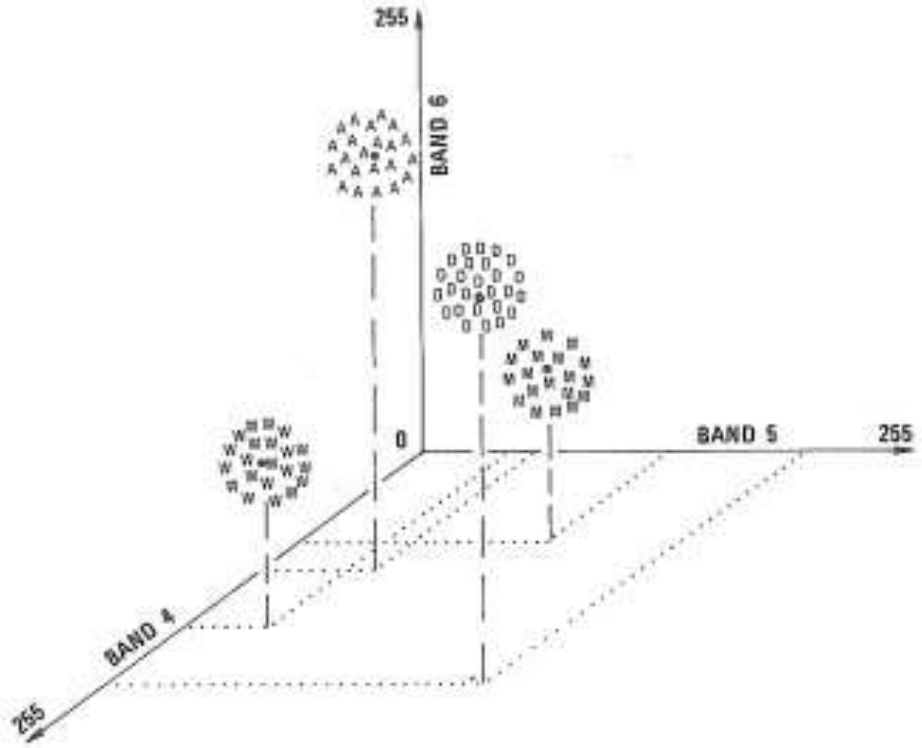
DNs in Band B are similar for corn and soybeans

… but if we use both Bands A and B, then all 3 crop types differ

4 land cover types - spectral reflectance **'3D' scatter plots**



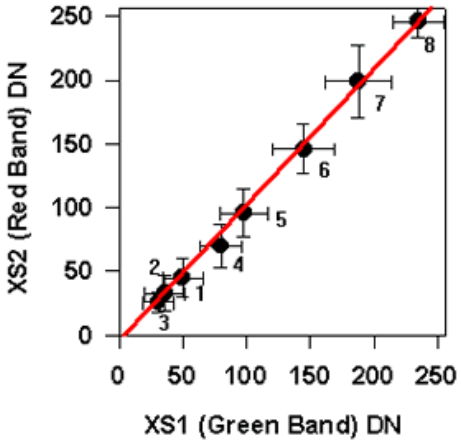
A. SPECTRAL REFLECTANCE CURVES FOR MAJOR TERRAIN TYPES.

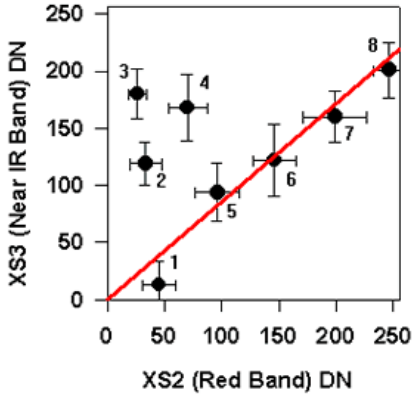


B. THREE–DIMENSIONAL CLUSTER DIAGRAM FOR CLASSIFICATION.

… Classification algorithms are '**per pixel**' classifiers

Band correlation coefficients and scatterplots





Example: PG Landsat 5 Thematic Mapper TM) data, 2011

Correlation: (r values between bands)

	TM1	2	3	4	5	6
TM1						
TM2	.97					
TM3	.96	.96				
TM4	.07	.16	.11			
TM5	.66	.72	.76	.64		
TM6	.77	.77	.81	.14	.80	
TM7	.83	.86	.90	.40	.93	.86

The Visible bands are highly correlated (similar) .. (r = .96 to .97)

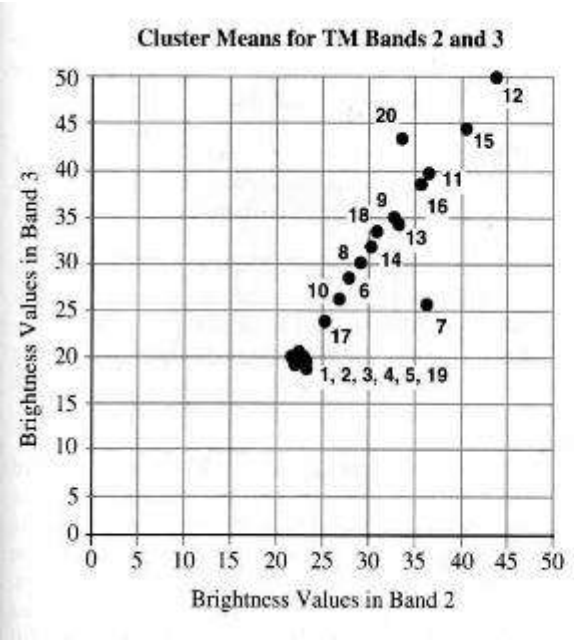
.. so also are bands 5 and 7 (r = .93)

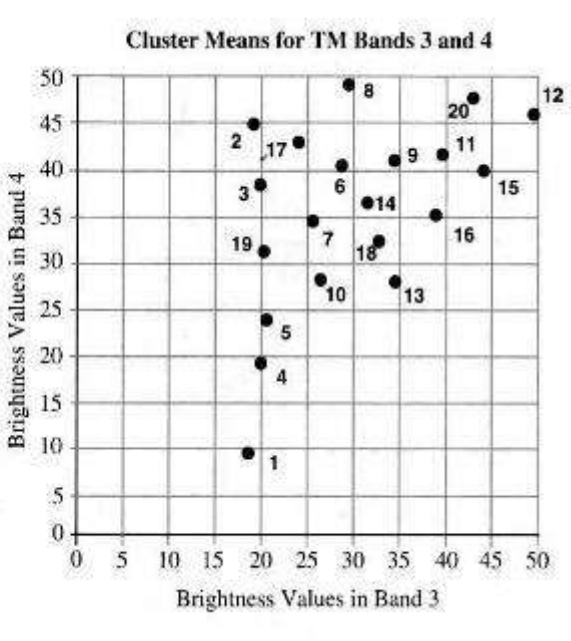
band 4 (near-IR) is not very correlated with Visible or SWIR (nor thermal)

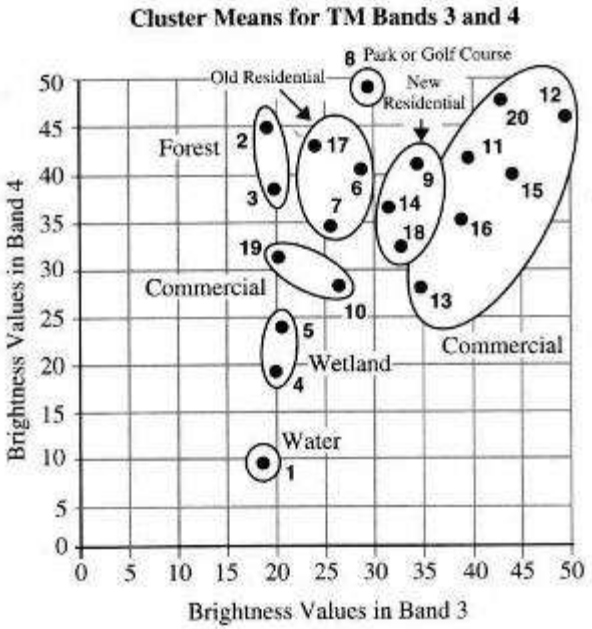
Note: these values will vary for different environments e.g. urban, desert, forested

Unsupervised classification = ‘clustering’

Visible bands (TM 2,3) versus Visible and Near-IR







Two bands are shown here for simplicity

Input bands selected: minimum 3+ bands;

Note: you can only display 3 bands, but you can input **more than** 3 (no limit)

… but the classifier can be constrained with too many inputs

Classification: Band / Channel Selection

How to choose which ones to use:

1. Low correlation e.g. Visible-NIR-SWIR

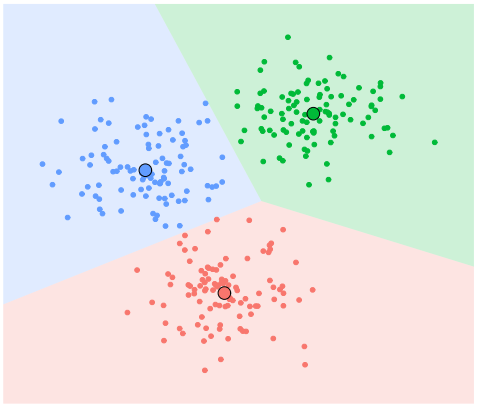
2. Past experience, visual examination, logical thinking

3. Channels that separate the features we want to identify
(based on DNs / spectral curves / histograms)

4. Or simply just use them all ?
- this can confuse the classifier and not find clusters

Unsupervised classification

Background



- user initially needs little 'a priori' knowledge of area

- the software clusters pixels by natural DN groupings
  based on similarity and contrast ~ 'natural breaks'

e.g. 1000 x 1000 pixel area = 1 million pixels,
many are alike and can be grouped

Steps

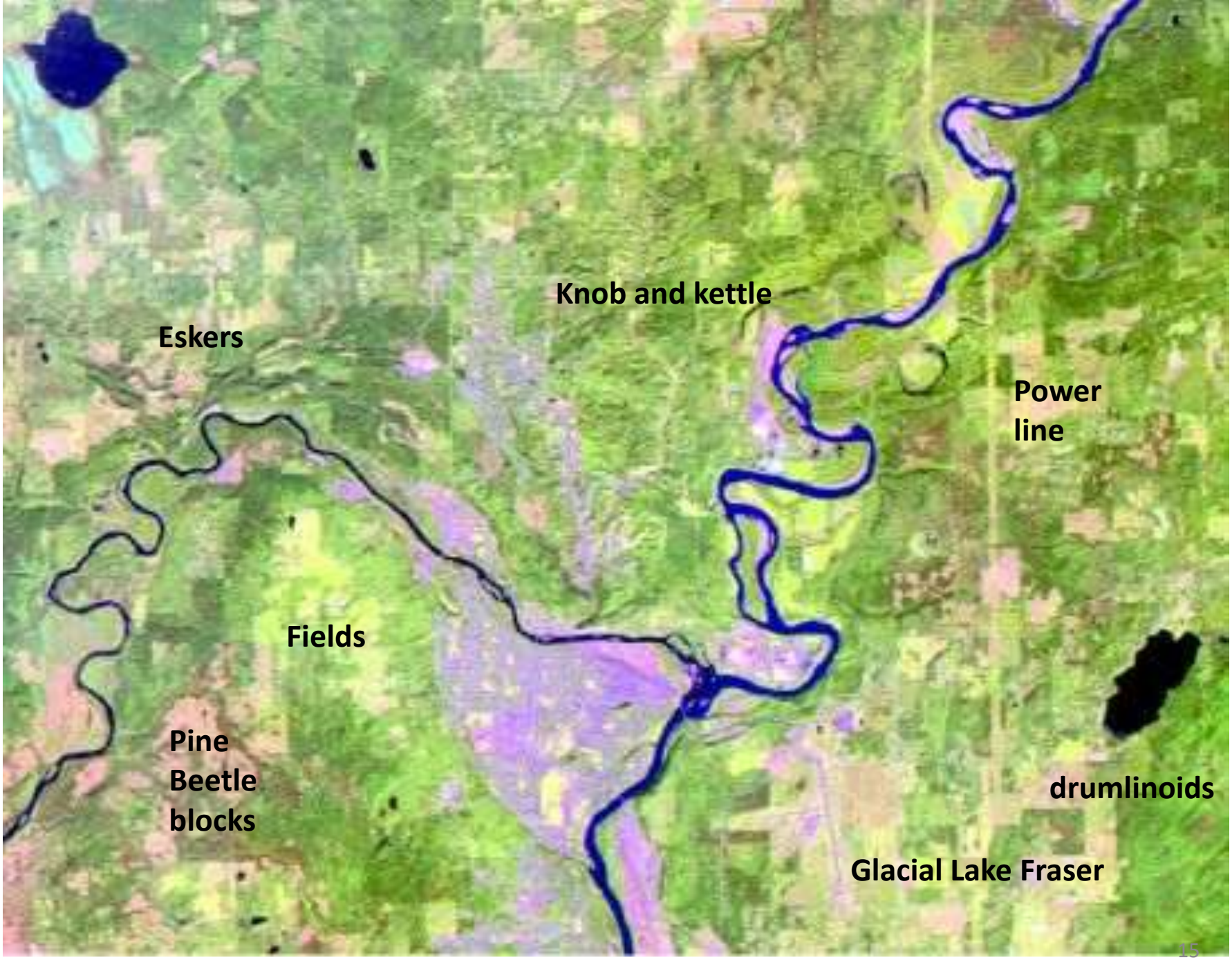
- determine how many clusters -> classes

- determine which input bands / channels to use

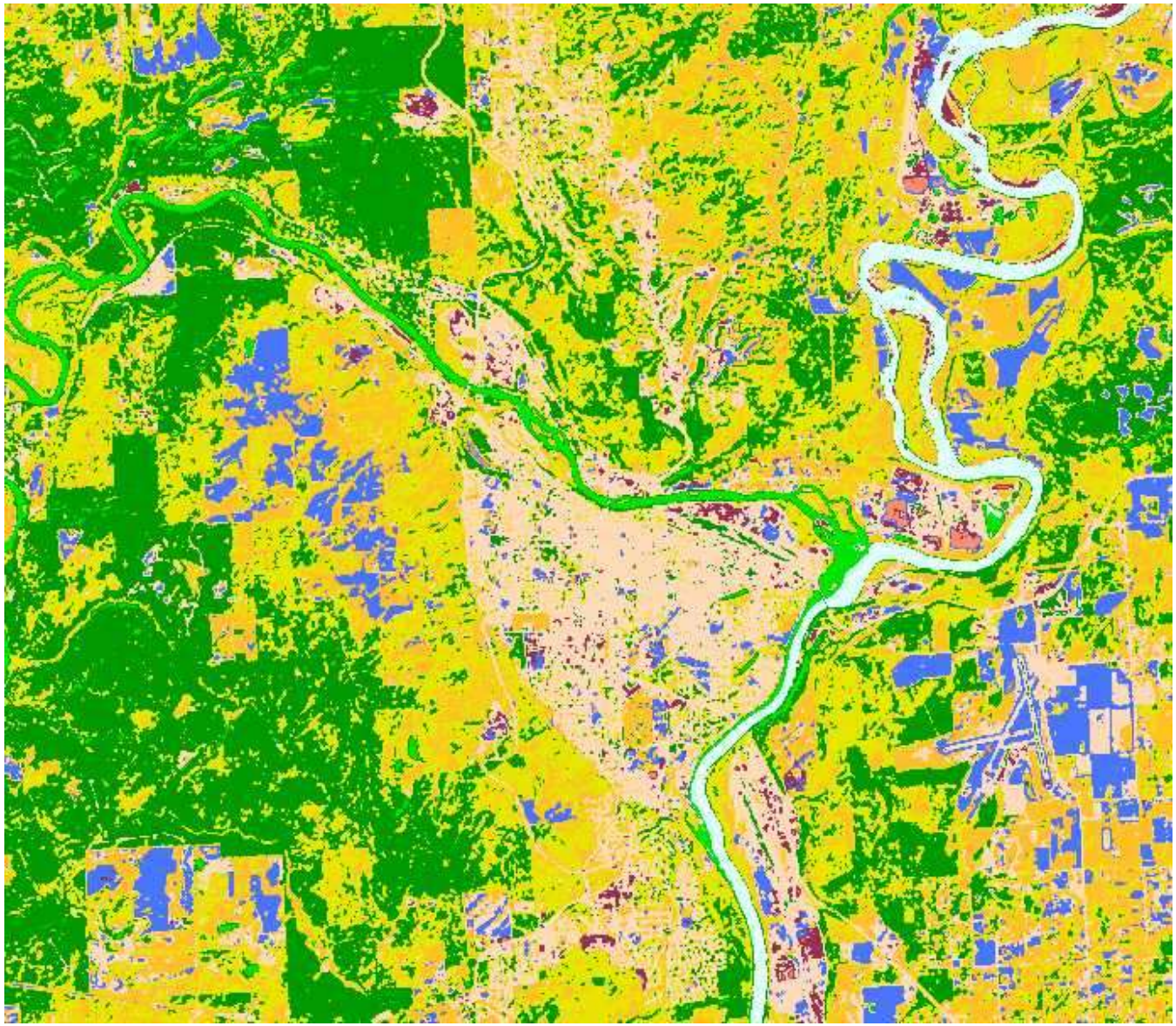
- run classifier : K-means or Isodata

- Rerun with more clusters if needed

- assign names to classes (merge classes if needed)


Knob and kettle
Eskers
Power line
Fields
Pine Beetle blocks
drumlinoids
Glacial Lake Fraser

Unsupervised result – 10 classes (clusters)

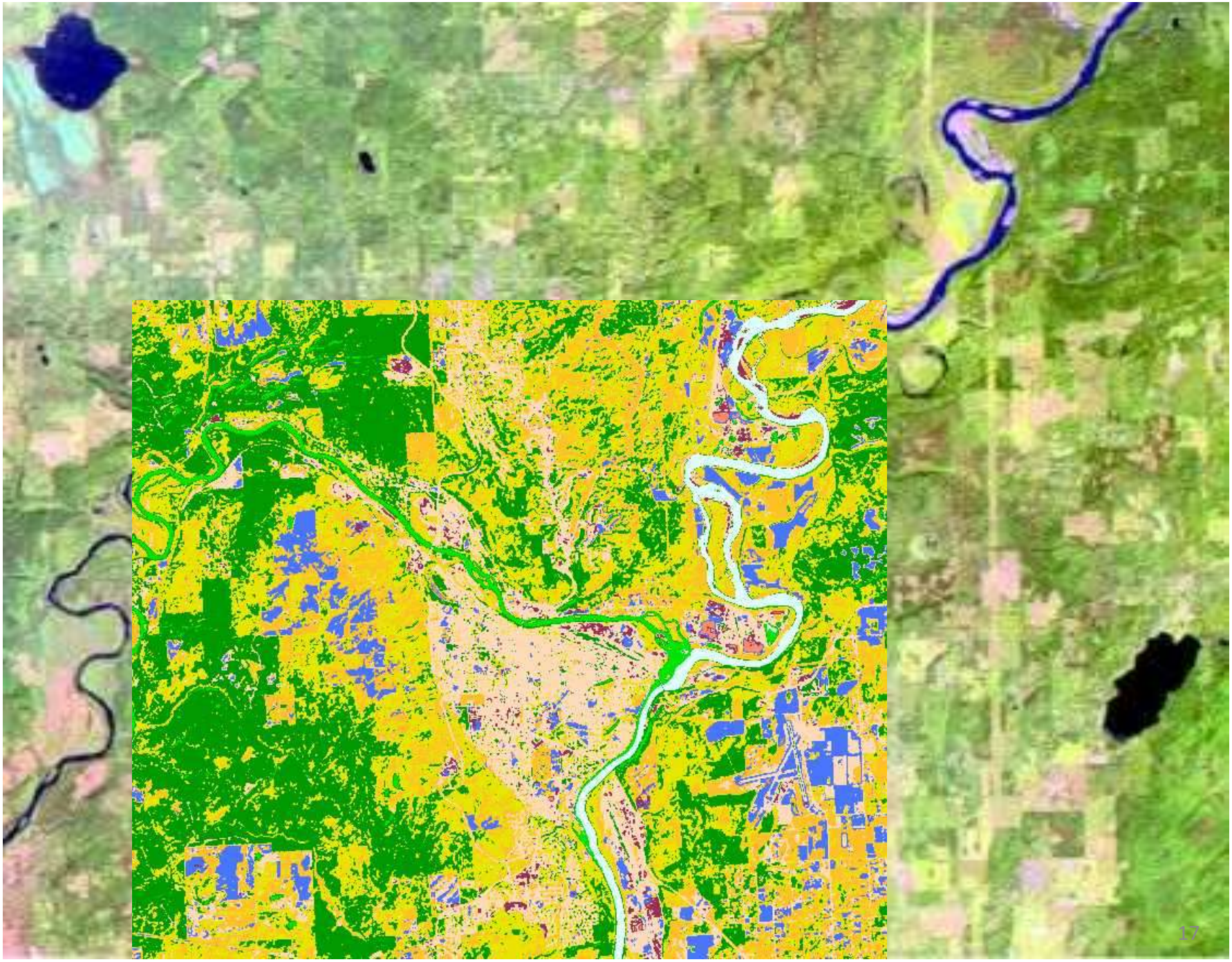


This is a new channel in your .pix file

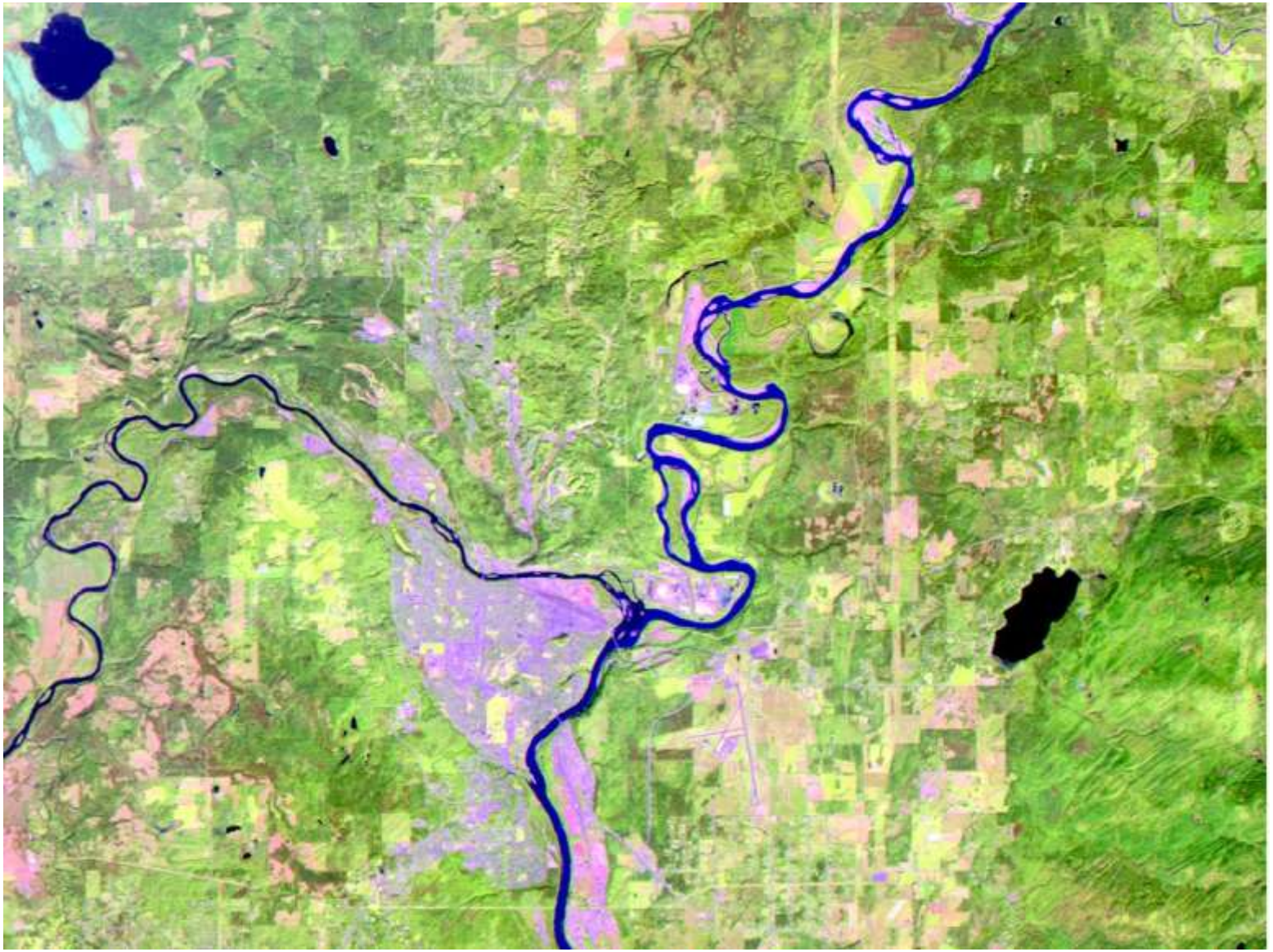
- it's not a band
- - displayed in pseudocolors

Colours are random

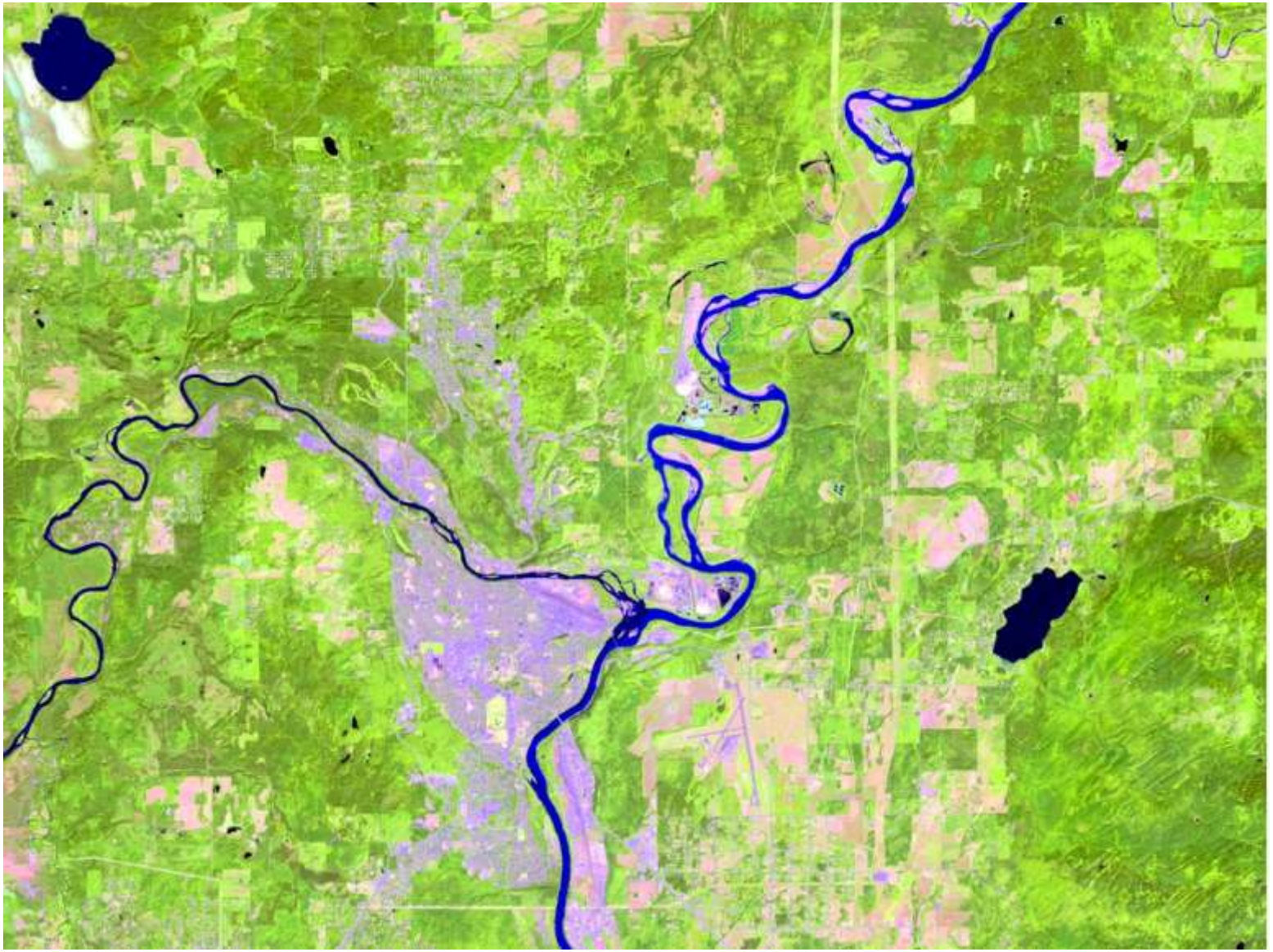
Note: urban classification is often NOT easy!



14 September 2011



20 July 2023 – note less shadows, logged areas recovering, swamp is ‘drier’



Cutblocks won’t distinguish as well as in 2011; deciduous vs coniferous has less contrast

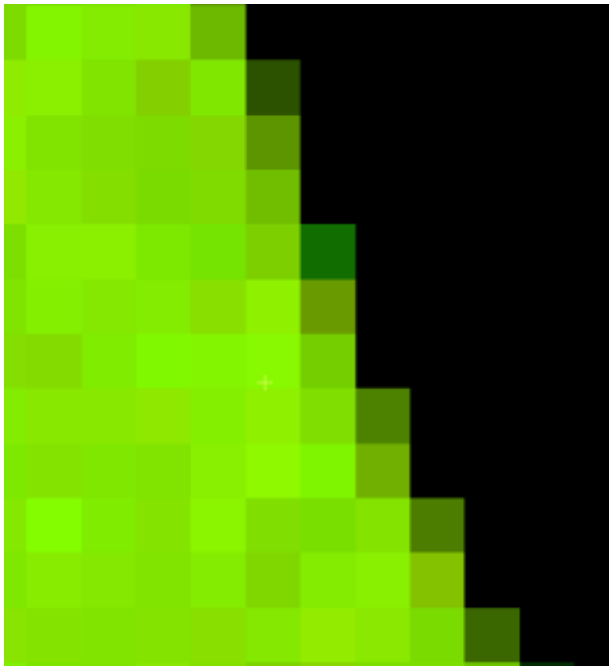
pixel size: Pure and Mixed Pixels

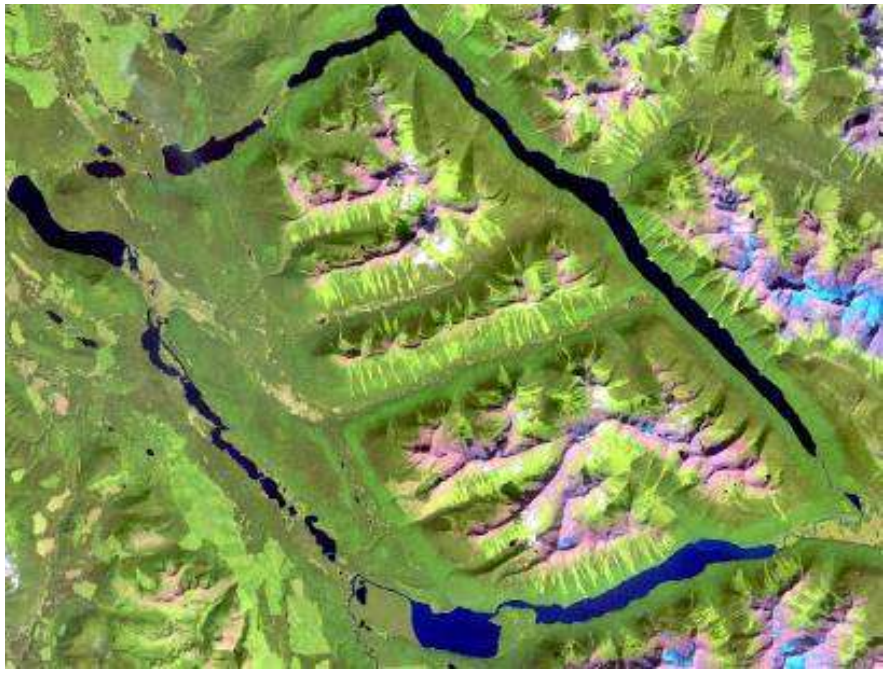
One pixel = one digital number value per layer

Remote sensing data and raster GIS data give the impression that a pixel has one uniform value across its width. This may be true for a small pixel or homogenous cover, such as a large lake, or field, but what we are seeing is an average value for a variable forest or a mixture of different surface covers.

Landsat example: Bowron Lakes with mix along the lake edges

1 pixel = 30 x 30m

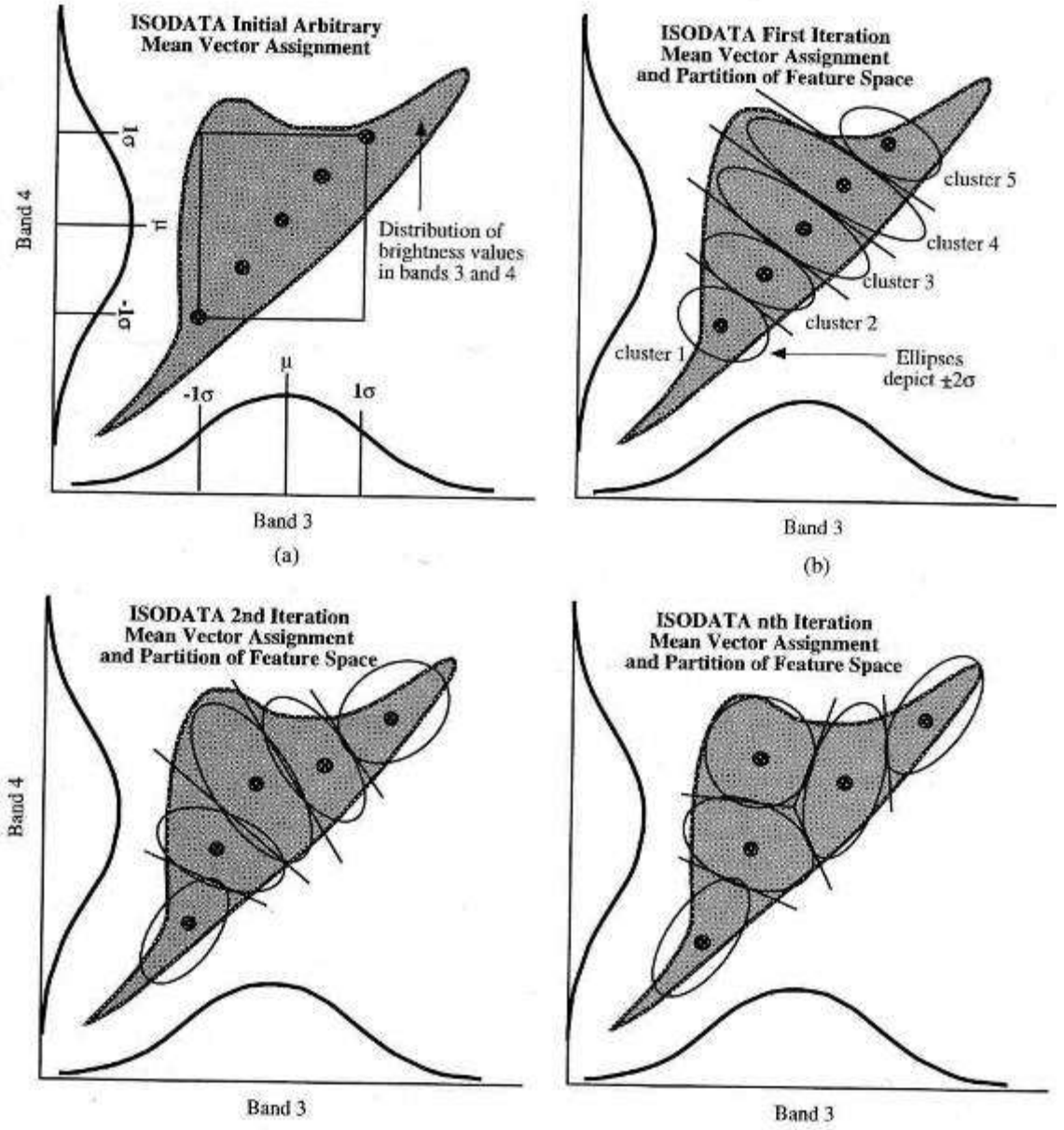




This trail and service road may be ~30m wide, but the pixels may include the edges

Unsupervised – how it works …. YIKES! (do we need to know this?)

- ❑ Algorithm starts with statistical seed points
- ❑ Assigns each pixel to the closest seed
- ❑ Calculates group mean …. in 'n-dimensional' space
- ❑ Re-assigns pixels to the closest group mean
- ❑ Re-calculates group mean
- ❑ Iterates (10 ?) until relatively little change and fixes groupings

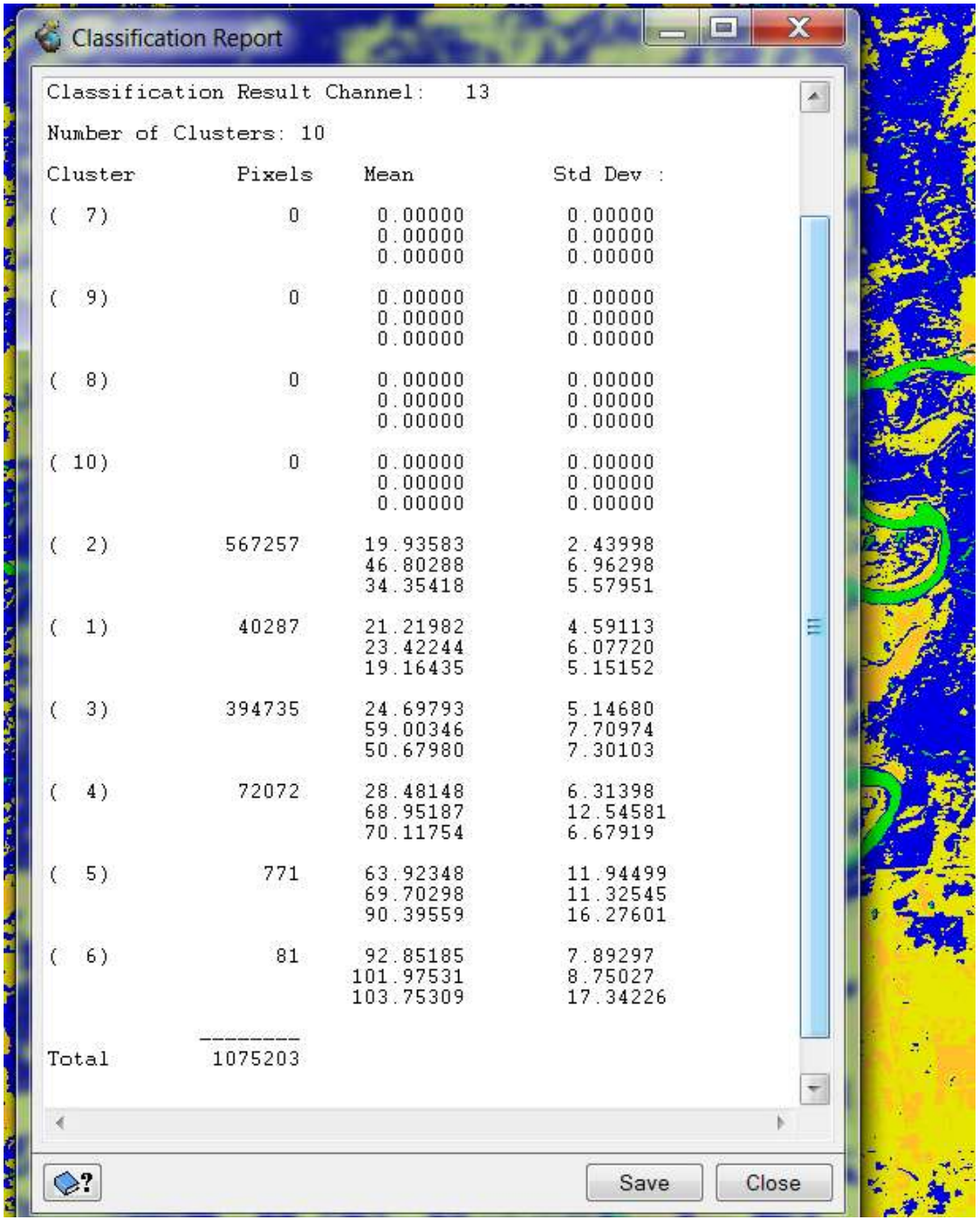


classification report
1 iteration

Note:
clusters with 0 pixels

DN values for 3 bands averages ..
Distribution is 1-2 dimensional

Final step will be Assigning names to the clusters (and merge/ split some)



After 16 iterations and 16 classes/clusters

Classification Report

```
Classification Algorithm: Fuzzy K-Means Unsupervised
Classification Input Channels:   3,4,5
Classification Result Channel:   9

Number of Clusters: 16
```

Cluster	Pixels	Mean	Std Dev :
(2)	296774	23.24140	8.24662
		44.77742	8.91783
		32.44915	10.04080
(3)	292356	24.48324	7.14404
		67.65602	10.67916
		49.51679	9.53926
(4)	155525	24.75149	5.03961
		107.39487	18.20386
		74.22362	13.35878
(1)	135750	42.07941	13.08230
		26.82458	8.12628
		16.47926	11.00162
(5)	151100	42.87475	9.25817
		60.36603	13.13133
		89.47187	18.67191
(7)	86198	84.79987	12.60066
		59.46275	9.37685
		20.03181	14.03484
(8)	85354	122.33620	10.20878
		85.34046	8.68392
		19.40815	16.75611
(9)	79592	151.05591	9.22842
		105.45887	8.13428
		20.59924	20.61743
(10)	60789	175.72850	8.52307
		125.70449	8.00244
		25.36989	25.79001
(11)	55539	201.23238	7.64972
		142.52280	7.16980
		17.82207	16.81002
(12)	54187	225.29511	7.08180
		159.06710	6.45671
		16.18565	13.71707
(13)	56164	247.23974	6.06449
		172.45732	4.56608
		13.11189	7.50873
(14)	113965	254.84619	0.99330
		185.52277	4.41832
		12.44097	4.33174
(15)	51887	254.93781	0.76476
		202.95095	5.59141
		14.38227	7.67388
(16)	33140	254.99879	0.08035
		225.13265	8.29124
		13.28431	3.35810

Fuzzy classification – each pixel has potential membership in more than one cluster

Unsupervised classification –algorithms and iterations
(PCI .. Fuzzy K-means is less common in GIS software)

1. K-means minimises 'within cluster range' of DNs

2. Fuzzy K-means enables mixed membership, based on distribution of the clusters

3. Isodata can also merge or split clusters, so the number of clusters is more flexible

Merging clusters

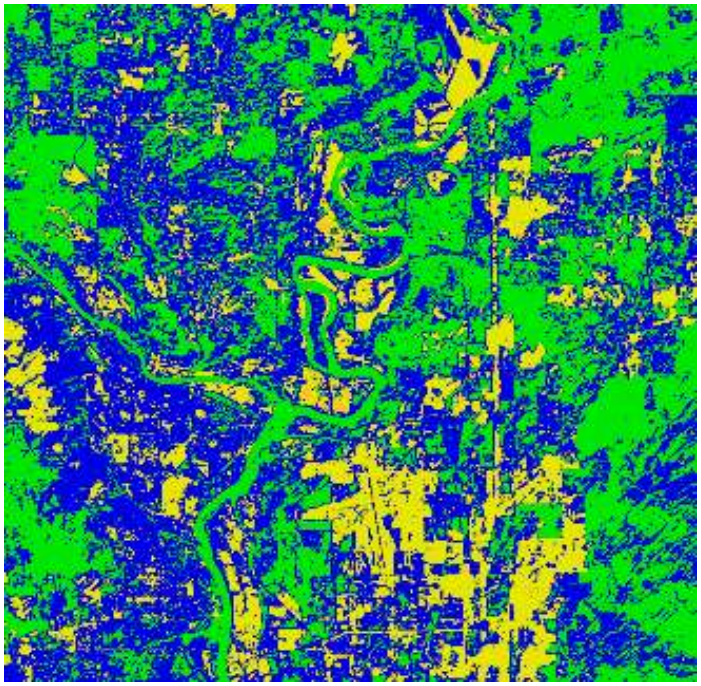
Merging – if clusters are not really separate features; Clusters are merged if they overlap spatially or are similar spectrally.
(visually examine image)

Splitting / adding clusters

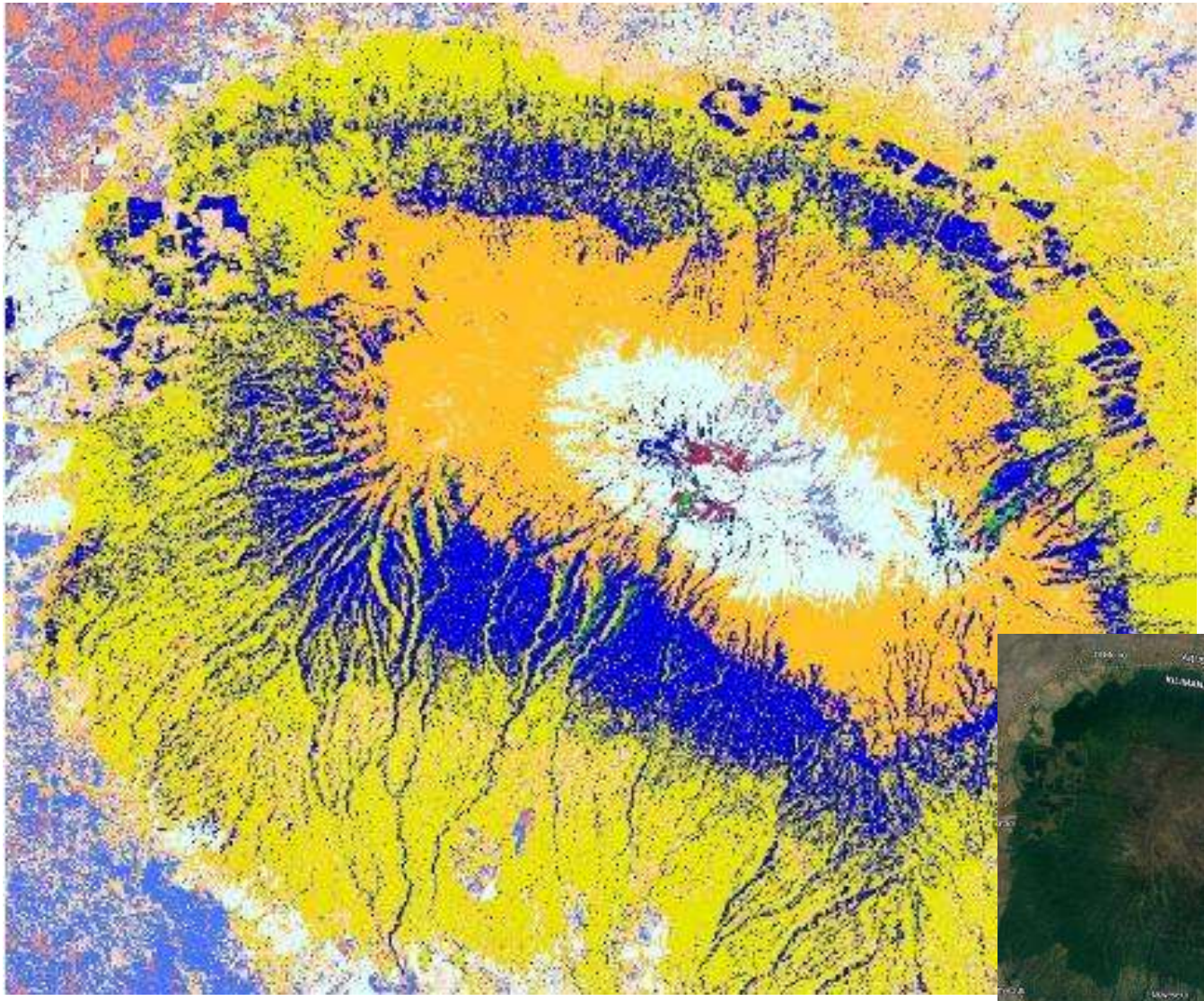
If one cluster covers too much area – run again with more clusters

One can also generate many clusters, and then group merge later …

One ploy is to make many clusters (e.g. 50-100 and plan to merge)



Too much detail / isolated pixels – ‘salt and pepper’ effect

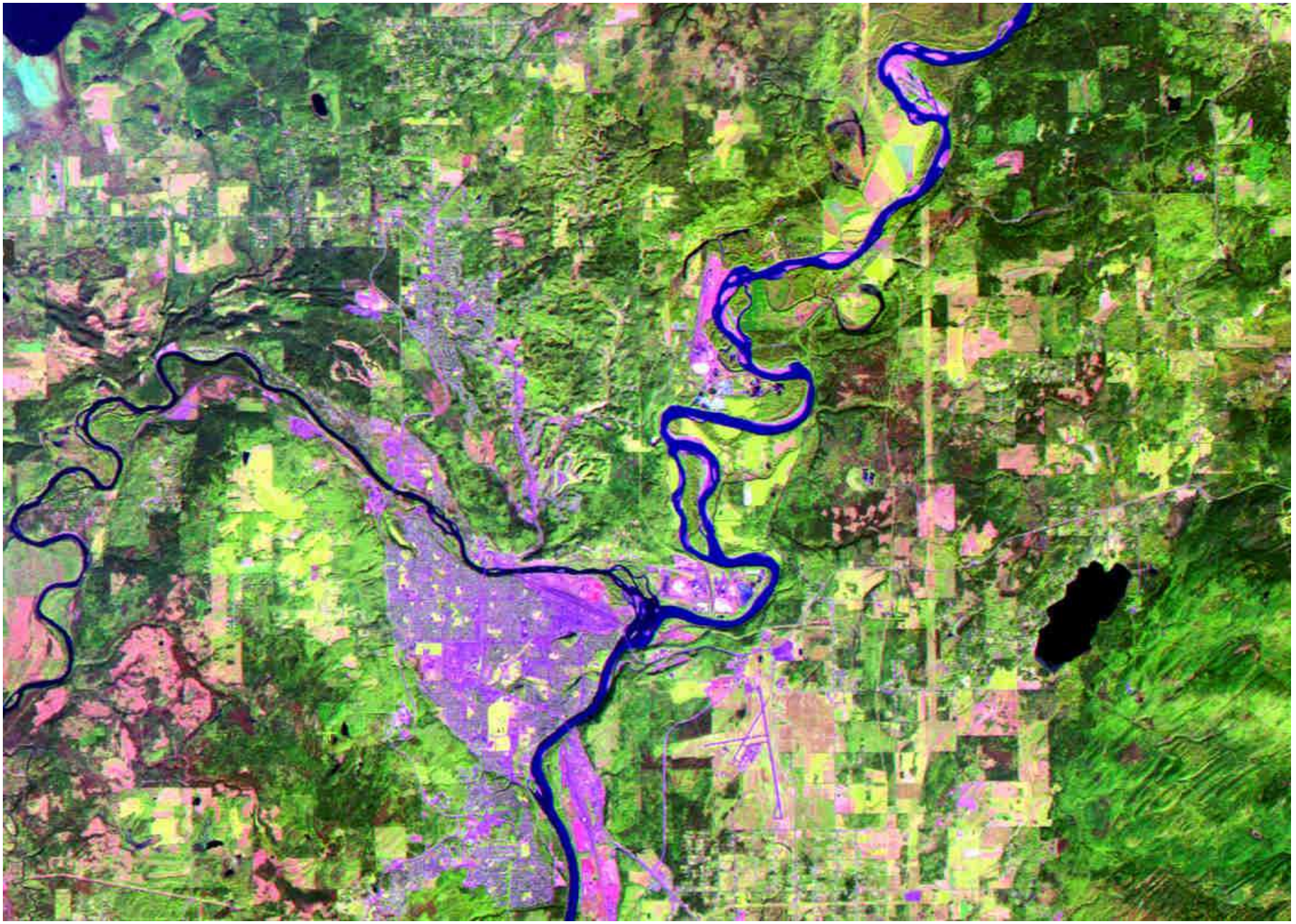


Mt. Kilimanjaro

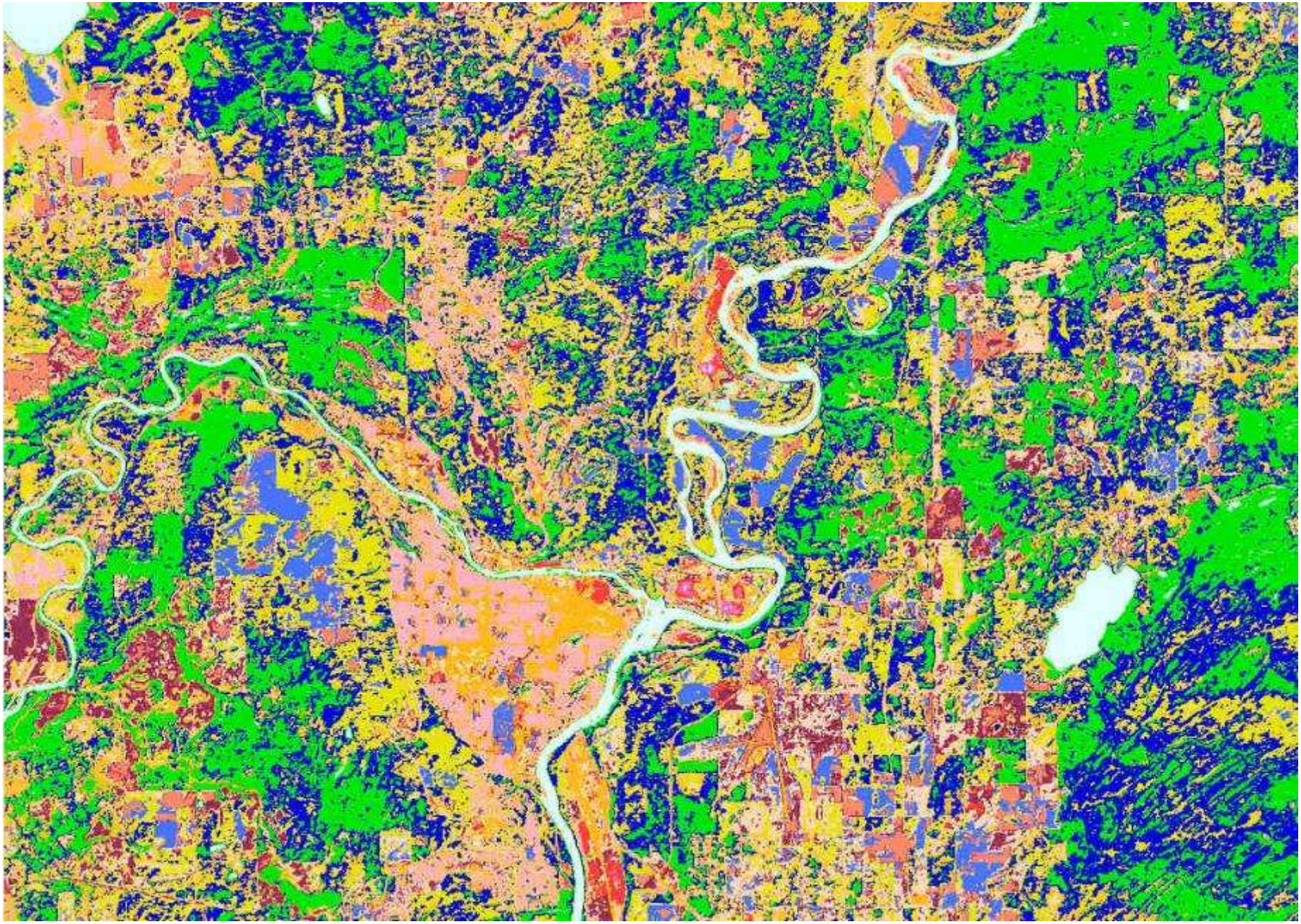
Always due to :

- fine local DN variations
- ‘**per-pixel**’ classifiers

Image- TM 543 September 2011



Isodata classification



reduced by Sieve tool

SIEVE
Merges isolated pixels into the adjacent class

Minimum cluster = ?

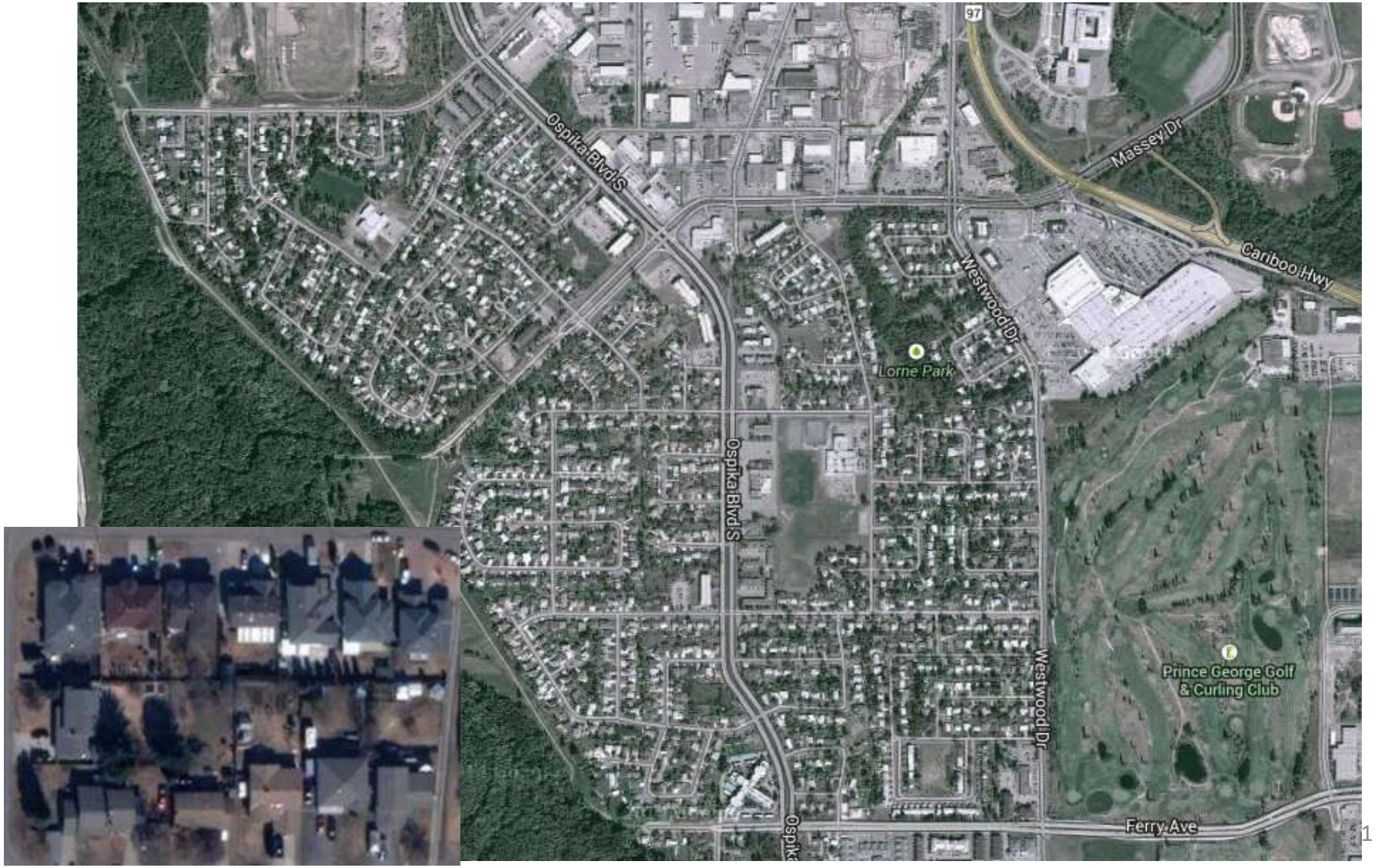
1 ha (100x100m) = 10,000m^2

Pixel = 30x30m (900m^2)
so 1ha = ~ 11 pixels

Or use 2 or 5 ha ?
22 / 55 pixels

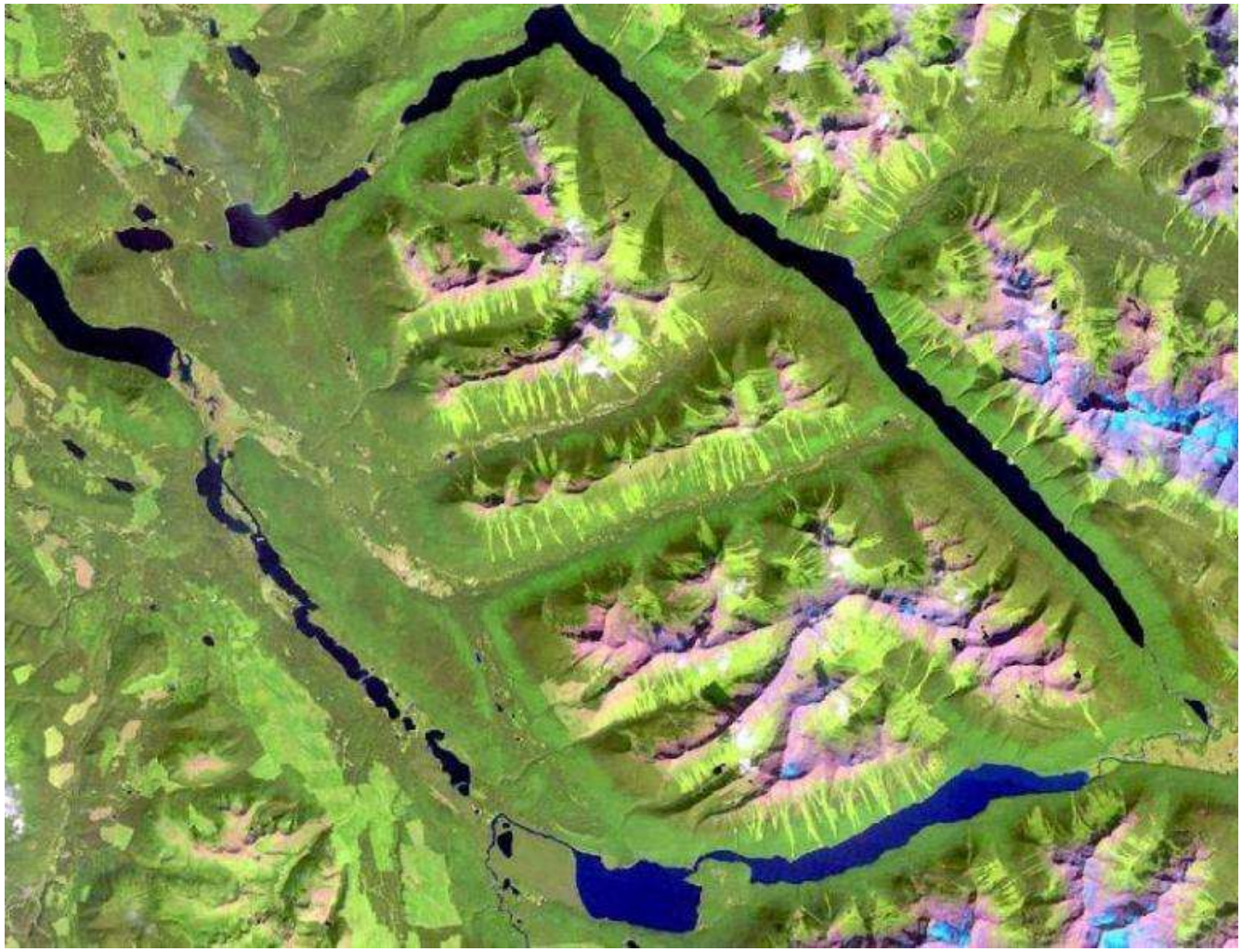
Challenges in classification – natural range of DN values

Urban / Residential – mosaic of smaller features inside a 30 metre pixel
- amount of grass, types of material, roofing colour, sun angle (building shape)



Challenges in classification (why it doesn't always easily beat digitizing)

There are many spatial variations in reflectance (a range of DNs for a feature)
e.g. stand purity, understory, age/maturity, density, disease, sun angle, **topography**



Classes: water, bare rock, glaciers, deciduous, coniferous, cutblocks, regrown

Overall summary on classification

- It is always complex – many classes and contrasts
- There are many causes of spatial variations in reflectance
- Most (natural) features are continuous, not discrete

- When using only pixel DNs:
- Any land cover types have a range of values
- Conversely, different cover types can appear similar

Further complications for all images:

a. moisture (recent events)
b. edge (mixed) pixels
c. sun angle (illumination)

Textbook classification goal: ~ 85% accuracy

Manual digitizing may not do any better