GEOG 357

LECTURE 6

1

Unsupervised classification: review

Characteristics
user needs no 'a priori' knowledge of area (but it helps)
- software clusters pixels by natural DN groupings
(based on similarity and contrast = 'natural breaks')

Steps
determine input bands / channels
determine how many classes / clusters
run classifier : K-means or Isodata
assign names to classes (merge classes if needed)
- calculate accuracy x

2

2

Supervised classification

- Process uses samples of known identity to classify pixels of unknown identity
- Samples of known identity are the training areas
  - Sections on the image that can be clearly matched to areas of known identity on the image
  - Typify spectral properties of the categories they represent
  - Must be homogeneous in respect to the informational category to be classified.

Supervised Classification: Advantages

- Control of informational categories tailored to a specific purpose and geographic area.
- Class identity determined through the process of selecting training areas.
- Informational categories directly match selected classes
- Errors in classification can be identified by examining how the training data have been classified by the procedure
  - inaccurate classification of training data indicates problems in the classification or selection of training data
  - correct classification of training data does not always indicate correct classification of other data.

Supervised Classification: Disadvantages

- Imposes a classification structure on the data vs the search for "natural" classes.
  - Classes may not be distinct enough in multidimensional data space.
- Train data are often defined with reference to informational categories and only secondarily with reference to spectral properties.
  - A training area that is "100% forest" may be accurate with respect to the "forest" designation but may still be very diverse with respect to density, age, shadowing, and the like, and therefore form a poor training area.
- If the area to be classified is large, or complex it may not be representative of conditions encountered throughout the image.
- The selection of training data can be a time-consuming, expensive, and tedious undertaking for example when matching training areas on maps and aerial photographs to the image to be classified.
- The possibility of missing unique categories not represented in the training data.

5

Supervised classification

- **Characteristics:**
- User has ' a priori' info: can identify homogenous known areas
- Software groups the pixels according to these 'training areas'
- ---
- **Steps**
- determine input bands / channels
- - **identify 'training areas' for each class**
- - **Check the statistics for separability**
- run classifier - minimum distance / maximum likelihood
- - **Calculate accuracy**

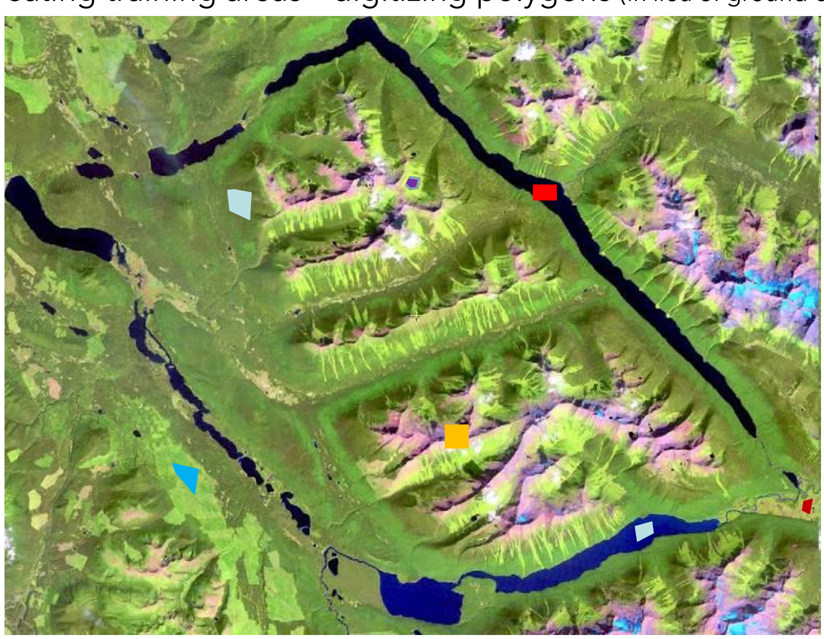
6

Understanding images for training areas selection
Reflection in visible / near IR / midIR

- **In a Landsat 4-5 TM 5-4-3 Colour composite**
- Visible = Brightness
- Near-IR= vegetation (vigour)
- MIR = dryness –low moisture
- Red = Dry, not much veg.
- Green = Healthy Vegetation
- brighter = deciduous
- Black = low reflection, water
- Purple (Red / Blue) = built-up

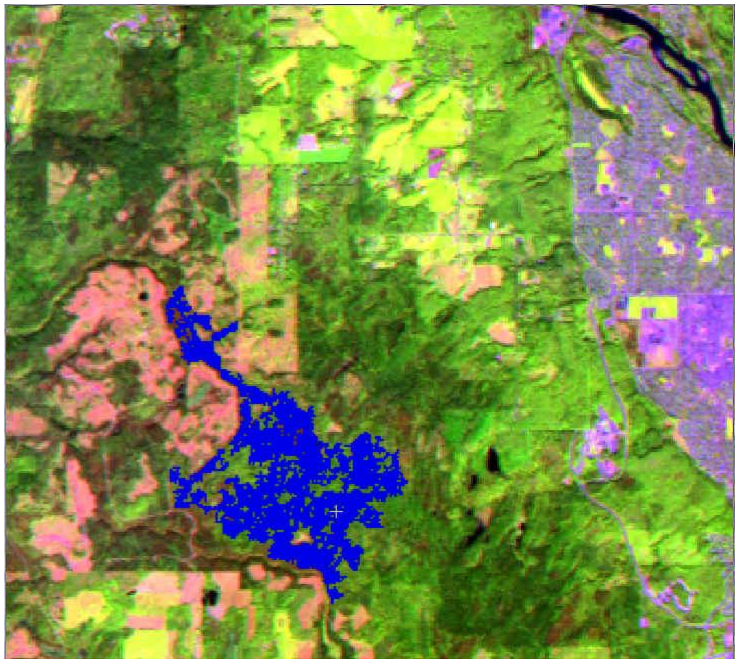
7

Creating training areas – digitizing polygons (in lieu of ground data)



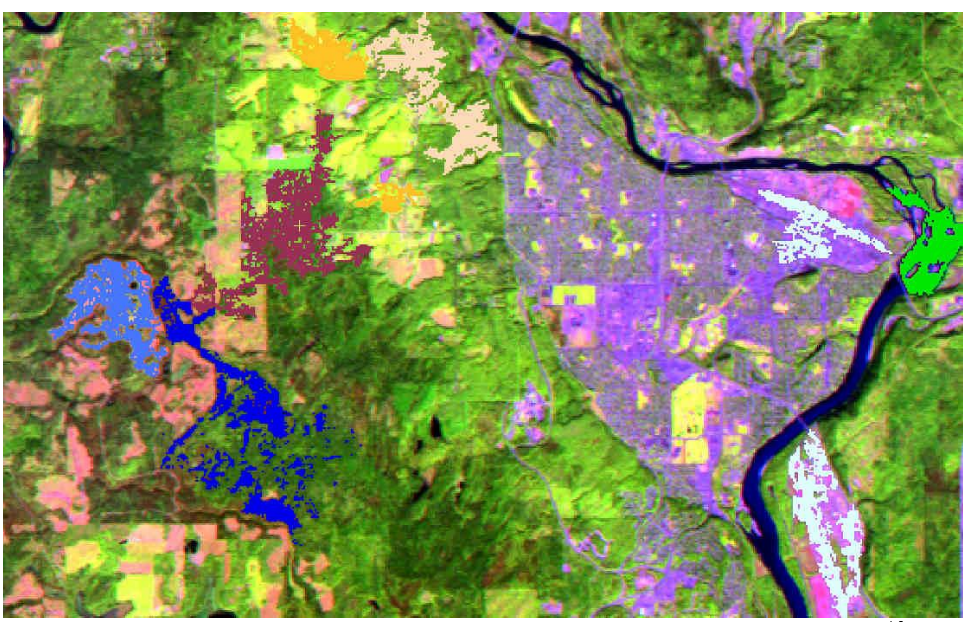
8

Raster seeding: See Catalyst Guide



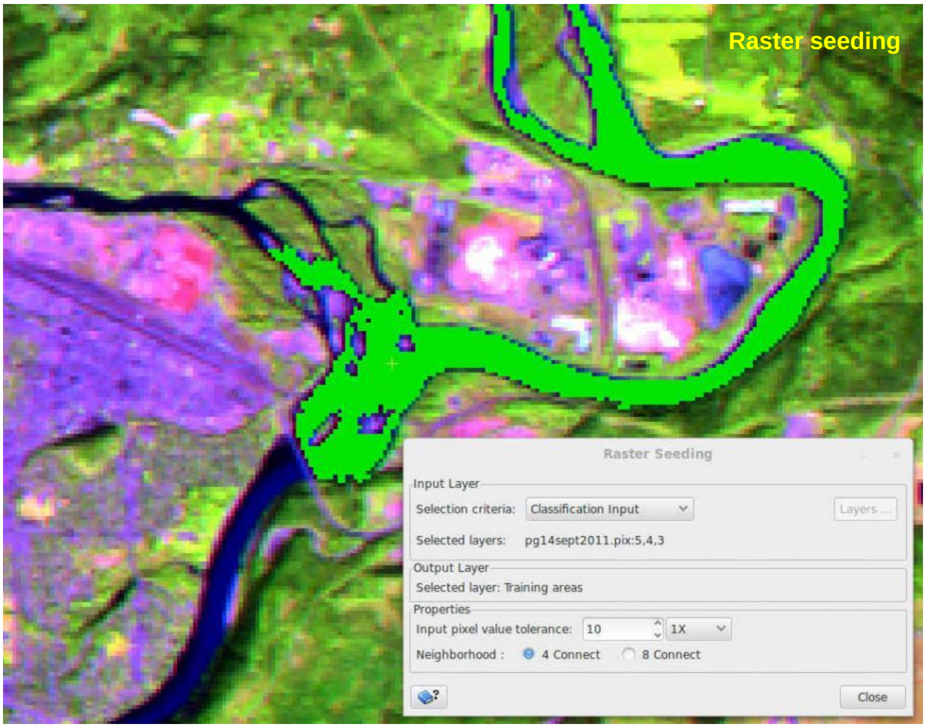
9

9



10

10



11

Supervised classification: separability

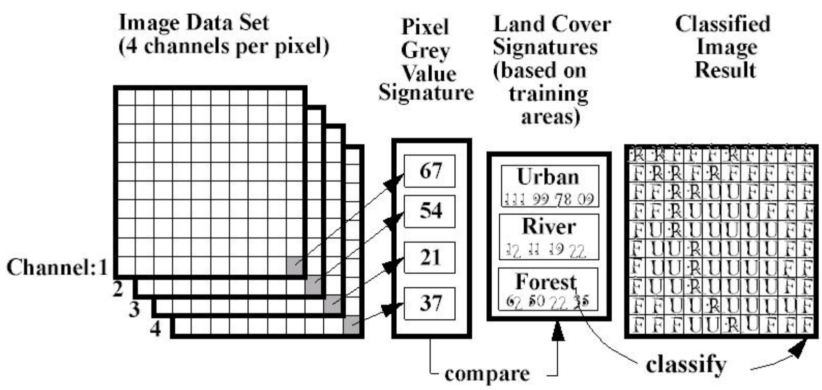
Create ground training sites for each class type (polygons, now 'seeding')

Create class signatures and check for differences (separability)

BAND:	1	2	3	4	5	6 (TH)	7	No. of Pixels
Class								
1. Seawater	57.4	16.0	12.0	5.6	3.4	112.0	1.5	2433
2. Sediments1	62.2	19.6	13.5	5.6	3.5	112.2	1.6	681
3. Sediments2	69.8	25.3	18.8	6.3	3.5	112.2	1.5	405
4. Bay Sediment	59.6	20.2	16.9	6.0	3.4	111.9	1.6	598
5. Marsh	61.6	22.8	27.2	42.0	37.3	117.9	14.9	861
6. Waves Surf	189.5	88.0	100.9	56.3	22.3	111.9	6.4	1001
7. Sand	90.6	41.8	54.2	43.9	86.3	121.3	52.8	812
8. Urban1	77.9	32.3	39.3	37.5	53.9	123.5	29.6	747
9. Urban2	68.0	27.0	32.7	36.3	52.9	125.7	27.7	2256
10. Sun Slope	75.9	31.7	40.8	43.5	107.2	126.5	51.4	5476
11. Shade Slope	51.8	15.6	13.8	15.6	14.0	109.8	5.6	976
12. Scrublands	66.0	24.8	29.0	27.5	58.4	114.3	29.4	1085
13. Grass	67.9	27.6	32.0	49.9	89.2	117.4	39.3	590
14. Fields	59.9	22.7	22.6	54.5	46.6	115.8	18.3	259
15. Trees	55.8	19.6	20.2	35.7	42.0	108.8	16.6	2048
16. Cleared	73.7	30.5	39.2	37.1	88.4	127.9	45.2	309

http://www.fas.org/irp/imint/docs/rst/Sect1/Sect1_17.html

12



Per pixel classifiers

13

Supervised classification methods: a. Minimum distance

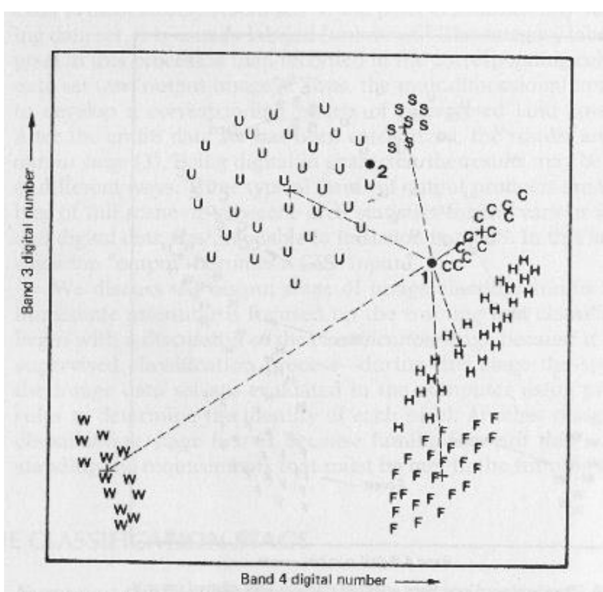


Figure 7.40 Minimum distance to means classification strategy.

The graphic is 2D

Letters indicate a training pixel

Think in n-dimensions:

The screen can only display 3 bands but a classifier can input many more

14

Supervised classification methods: b: Parallelepiped

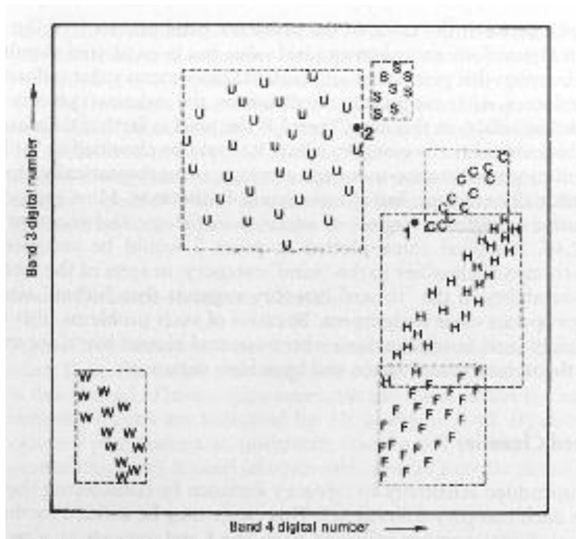


Figure 7.41 Parallelepiped classification strategy.

Less used due to overlap of training areas – conflict of assigning pixels to classes

15

Supervised classification methods c: Maximum likelihood

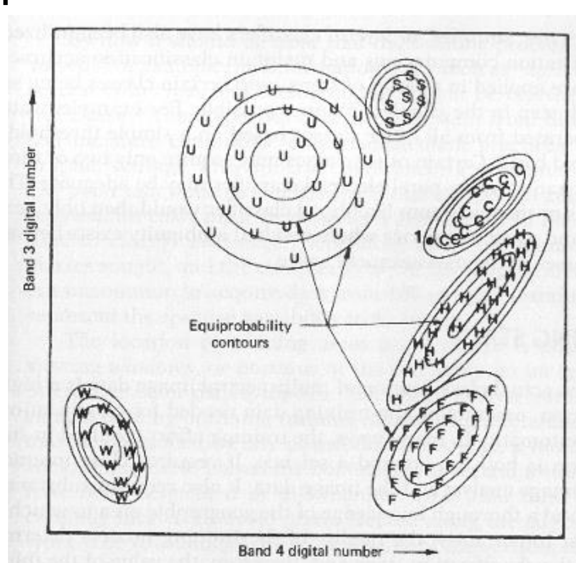


Figure 7.44 Equiprobability contours defined by a maximum likelihood classifier.

With or without null class

16

Supervised classification: how it works

Minimum distance: each pixel is assigned to the class whose mean is closest to data point

(in n-dimensions)

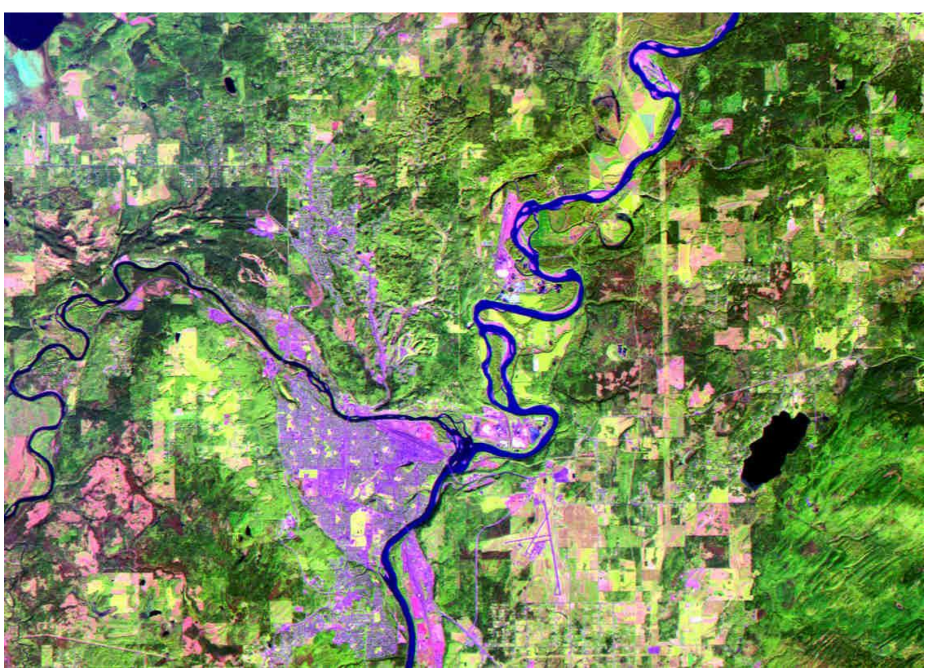
Parallelepiped: Each pixel is assigned to the class whose range it falls in

(overlap = double assignment)

Maximum Likelihood: each pixel is assigned to the class for which it has the highest probability.

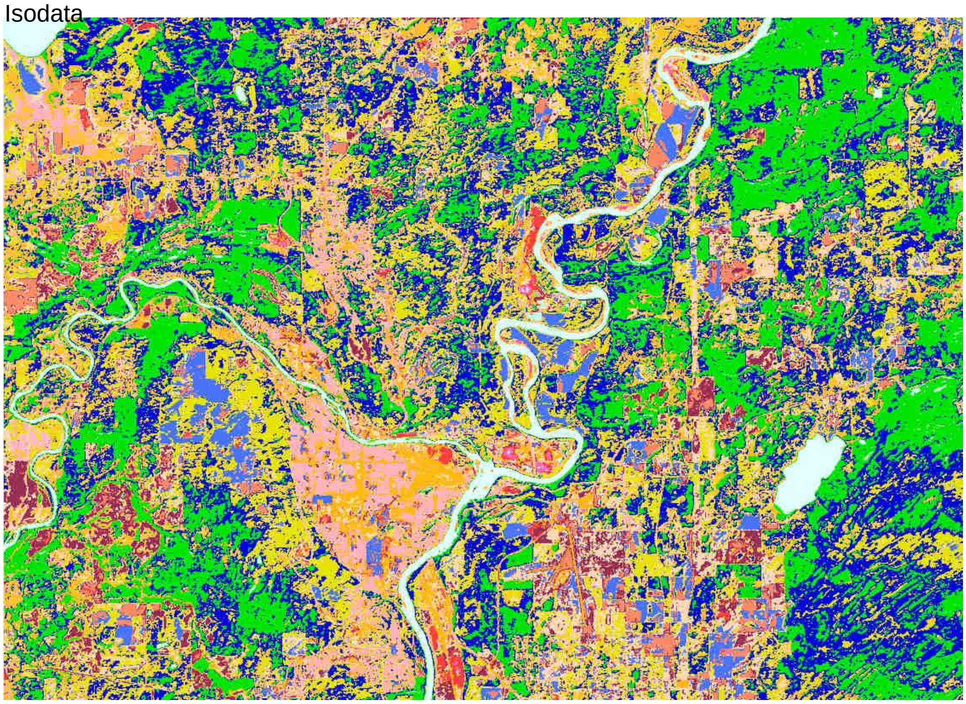
Max. likelihood can be run with a 'null class'
(where some pixels have no assignment to any class)

17



18

Isodata



19

Merging and adding classes

Merging

a. if classes overlap spatially or b. are not distinguishable spectrally.

Splitting / adding: one class covers too much area
[Unsupervised: - run again with more clusters]

Supervised:– create new training class or delete some training areas

Areas are unclassed – create new training class

20

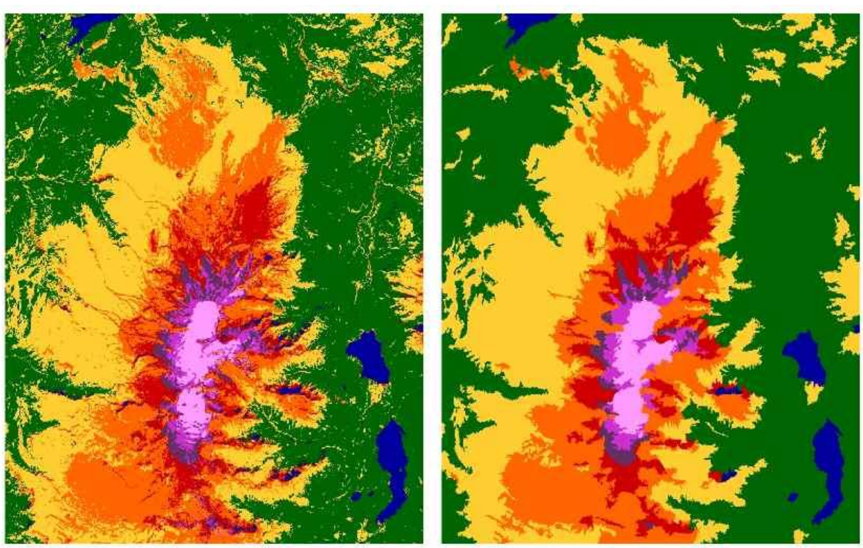
Post-classification steps

- **Check the display**
- **Merge / add classes**
- **Sieve … to remove isolated pixels**
- **Accuracy assessment**

- Conversion of results to vectors – see lab 7

21

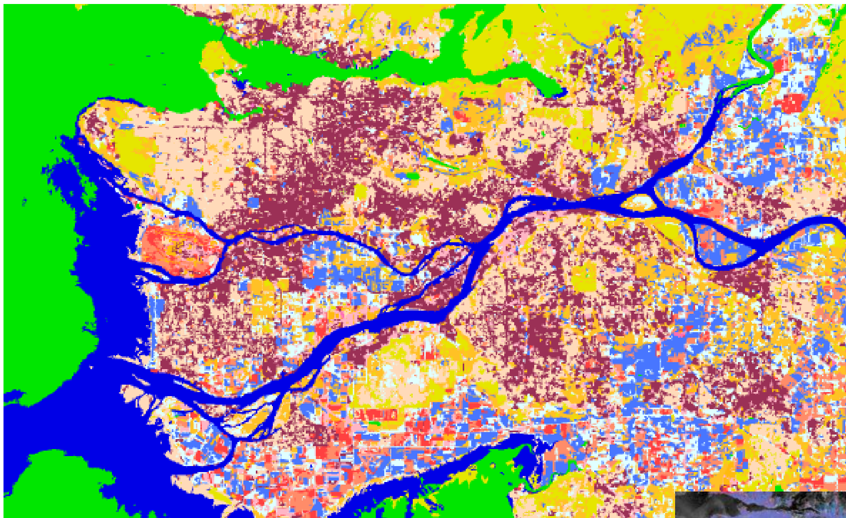
Mt. Edziza – classification and **sieve – removing isolated pixels**



- recognises connectivity of adjacent pixels in the same class
- special classes e.g. wetlands can be specified and preserved
- removes small sub-areas; does not 'blur' edges like filtering

22

Supervised classification –GEOG432 project



Classification on Metro Vancouver, September 2011

23

Accuracy assessment

This requires knowing what is reality at some pixels (ground truthing), and how they were classified. This generates a **'confusion matrix'**

	Class	Reference test information				Row total	User's Accuracy
		Road	Building	Green	Bare		
Remote sensing classification	Road	101	0	25	20	146	69.18%
	Building	0	128	0	17	145	88.28%
	Green	10	0	104	1	115	90.43%
	Bare	2	4	2	105	113	92.92%
	Column total	113	132	131	143	519	
	Producer's accuracy	89.38%	96.97%	79.39%	73.43%		

Overall accuracy = 84.4%, Kappa coefficient: 0.825.

The diagonal represents pixels correctly classified
An off diagonal column element = an 'error of omission'
An off diagonal row element = 'error of commission'

http://www.gisdevelopment.net/application/nrm/overview/mma09_Mustapha.htm

24

Measuring accuracy

The overall yardstick of 85% accuracy is held up as a (rarely achieved) ideal.
Producer's accuracy: based on ground truth pixels
User's accuracy: based on classified pixels
Kappa: a composite accuracy index

Error (Confusion) Matrix Classified

Reference DataData

	water	ice	snow	conif	decid	alpine	rock	deglac	TOTALS
water	2	0	0	0	0	0	0	0	2
ice	0	5	0	0	0	0	1	0	6
snow	0	1	6	0	0	0	2	0	9
conif	0	0	0	14	1	0	0	0	15
decid	0	0	0	2	4	0	0	0	6
alpine	0	0	0	0	1	1	0	0	2
rock	0	0	0	0	0	0	4	0	4
deglac	0	1	0	0	1	1	5	3	11
Totals	2	7	6	16	7	2	12	3	55

	Producer's accuracy	User's	Kappa
water	100.000%	100.000%	1.0000
ice	71.429%	83.333%	0.8090
snow	100.000%	66.667%	0.6259
conif	87.500%	93.333%	0.9060
decid	57.143%	66.667%	0.6181
alpine	50.000%	50.000%	0.4811
rock	33.333%	100.000%	1.0000
deglac	100.000%	27.273%	0.2308

25

Classification review

Unsupervised classification:
clustering into classes
identification of classes by user

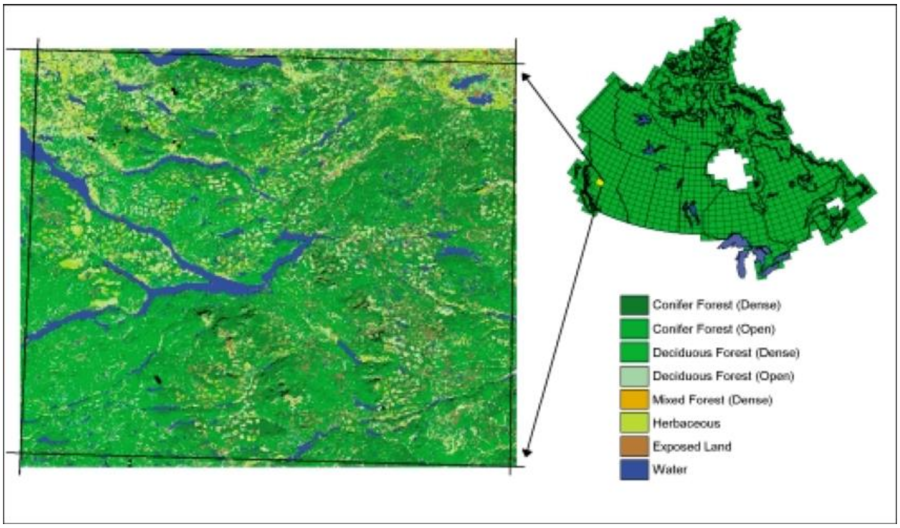
Supervised classification:
training areas to 'train' the classification,
check the statistics of the classes created
check resulting coverage for errors and accuracy

Unsupervised	Supervised
Unknown classes beforehand	Pre-defined classes
Clusters may not match desired classes	Defined classes may not match natural classes
Desired clusters may be unidentifiable	Selected training areas may be inadequate
a posteriori' cluster identification time-consuming	'a priori' training is time consuming
Unexpected categories may be revealed	Only predefined classes will be found
Immediate execution, quick	Takes longer, but better directed

26

EOSD Earth Observation for Sustainable Development of Forests

80% Canada mapped from Landsat 7 ~2000
- using supervised classification, 480 Landsat scenes, 630 1:250,000 map sheets



27

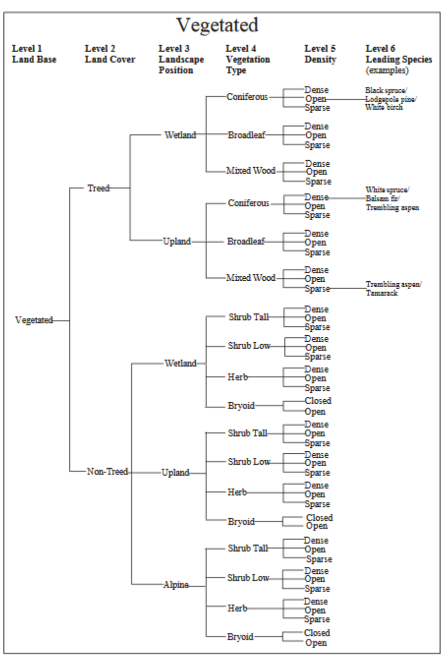


Figure 1 - Structure of the Land Cover Classification Scheme - Vegetated polygons

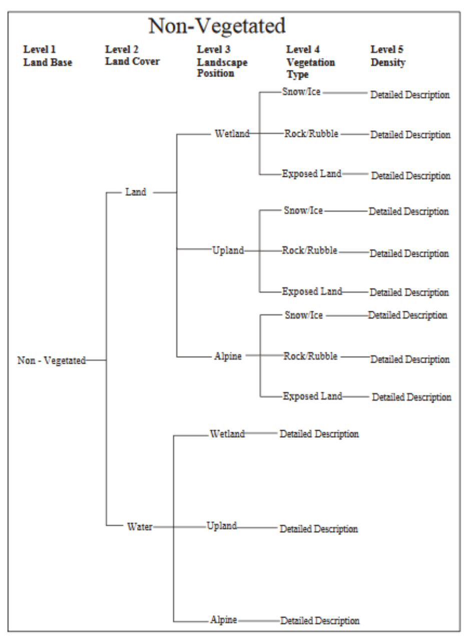


Figure 2 - Structure of the Land Cover Classification Scheme - Non-Vegetated polygons

EOSD Earth Observation for Sustainable Development of Forests ...
http://ftp.maps.canada.ca/pub/nrcan_rncan/vector/geobase_lcc_csc/shp_en/

28

Classification summary

There are many articles on classification approaches:

▪Input channel combinations (see the next lectures)

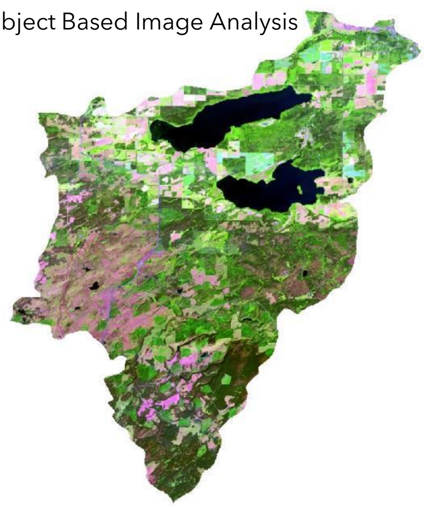
▪Best algorithms unsupervised and supervised

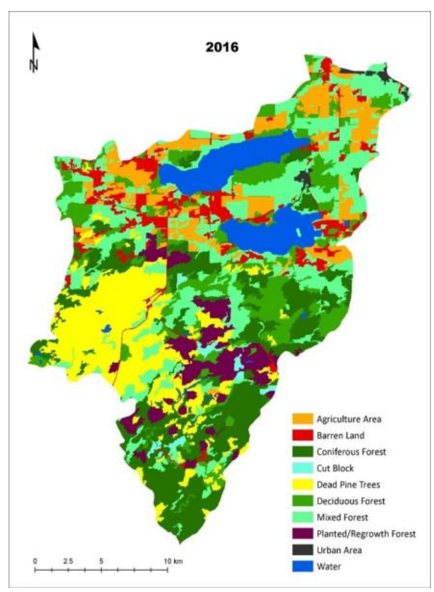
▪New approaches e.g. include texture, shape etc.

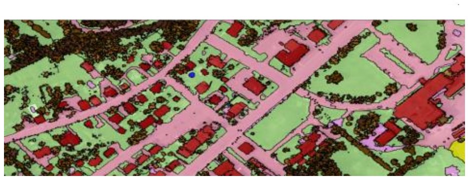
▪ Object based image analysis (not pixel based) .. See next slide

29

Object Based Image Analysis







More complex than per pixel classifiers; used in GEOG457 And by graduate students – identifies objects or shapes first

30