

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 ?

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

Understanding images for training areas selection

Reflection in visible / near IR / midIR

In a Landsat TM 5-4-3 Colour composite

Visible = Brightness

Near-IR= vegetation (vigour)

MIR = dryness –low moisture

Red = Dry, not much veg.

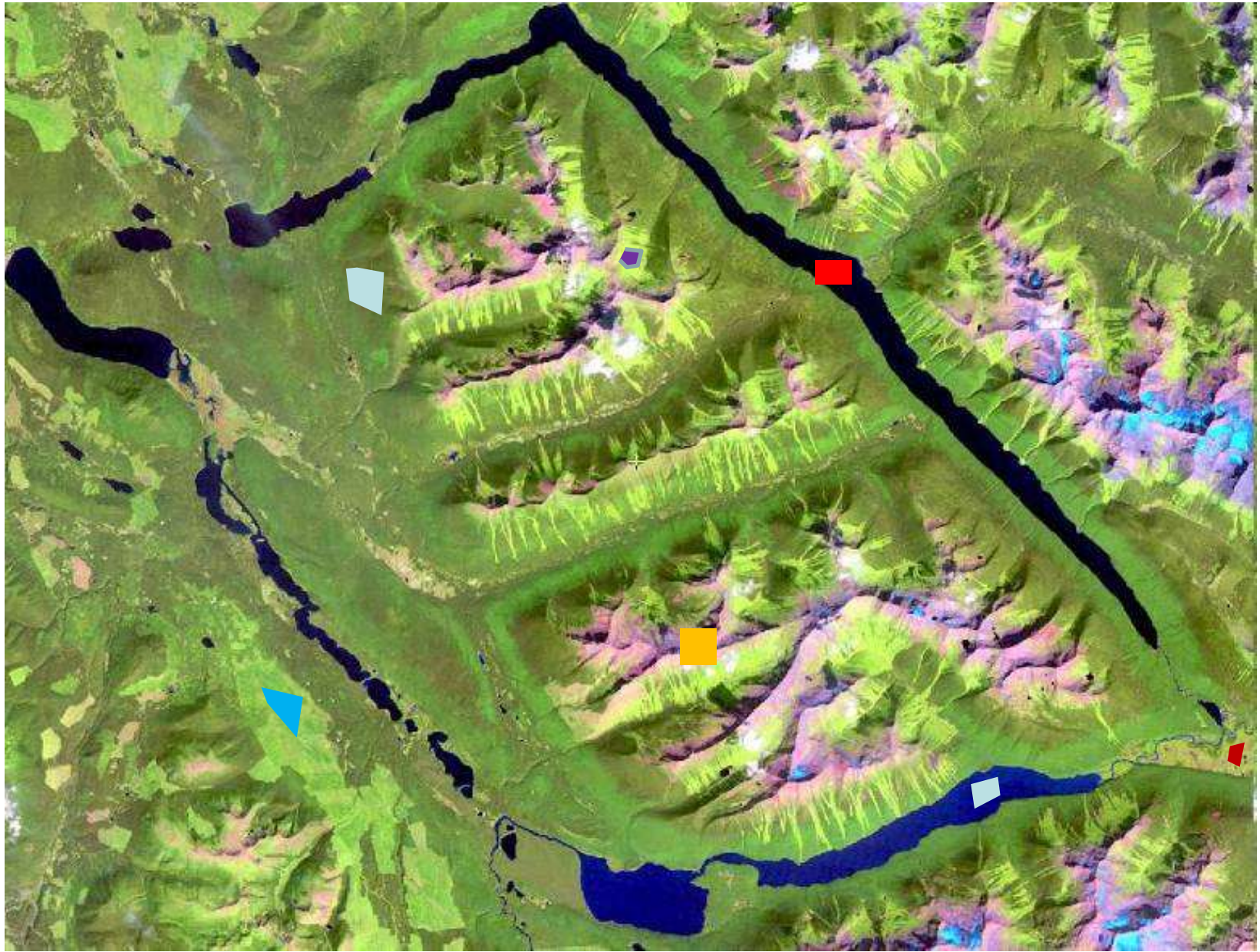
Green = Healthy Vegetation
brighter = deciduous

Black = low reflection, water

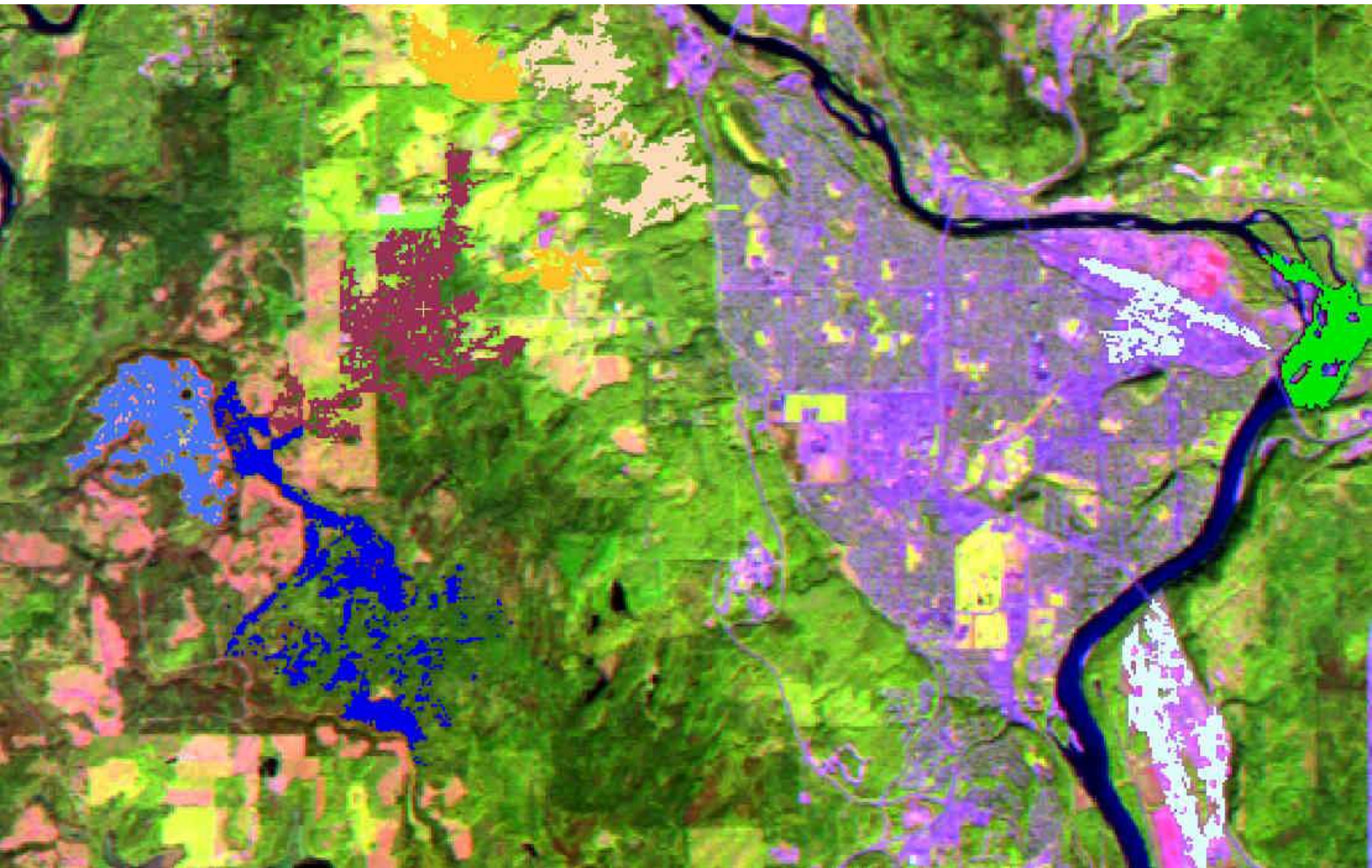
Purple (Red / Blue) = built-up



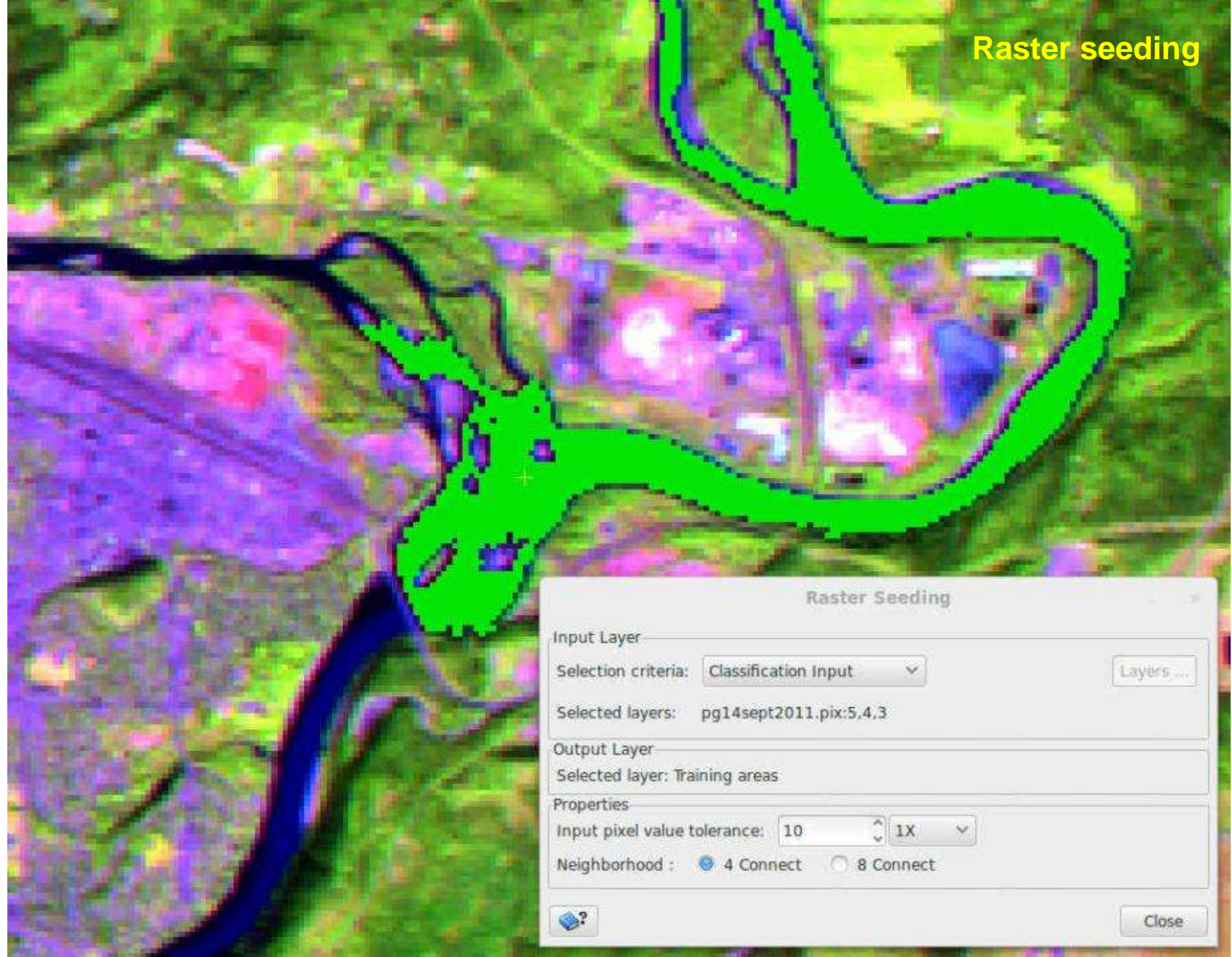
Creating training areas - digitizing polygons (in lieu of ground data)



Raster seeding – sample areas, don't try to fill it all



Raster seeding



Raster Seeding

Input Layer
Selection criteria: Classification Input Layers...
Selected layers: pg14sept2011.pix:5,4,3

Output Layer
Selected layer: Training areas

Properties
Input pixel value tolerance: 10 1X
Neighborhood : ☒ 4 Connect ☐ 8 Connect

? Close

Supervised classification: separability

Create ground training sites per class
Create class signatures and
check for differences (separability)

BAND:	1	2	3	4
Class				
1. Seawater	57.4	16.0	12.0	
2. Sediments1	62.2	19.6	13.5	
3. Sediments2	69.8	25.3	18.8	
4. Bay Sediment	59.6	20.2	16.9	6.0
5. Marsh	61.6	22.8	27.2	42.0
6. Waves Surf	189.5	88.0	100.9	56.3
7. Sand	90.6	41.8	54.2	43.9
8. Urban1	77.9	32.3	39.3	37.5
9. Urban2	68.0	27.0	32.7	36.3
10. Sun Slope	75.9	31.7	40.8	43.5
11. Shade Slope	51.8	15.6	13.8	15.6
12. Scrublands	66.0	24.8	29.0	27.5
13. Grass	67.9	27.6	32.0	49.9
14. Fields	59.9	22.7	22.6	54.5
15. Trees	55.8	19.6	20.2	35.7
16. Cleared	73.7	30.5	39.2	37.1

	Feature 1	Feature 1	Feature 1
Bad features	3.4	111.9	1.6
Good features	37.3	117.9	14.9
Best features	22.3	111.9	6.4
	86.3	121.3	52.8
	53.9	123.5	29.6
	52.9	125.7	27.7
	107.2	126.5	51.4
	14.0	109.8	5.6
	58.4	114.3	29.4
	89.2	117.4	39.3
	46.6	115.8	18.3
	42.0	108.8	16.6
	88.4	127.9	45.2
			598
			861
			1001
			812
			747
			2256
			5476
			976
			1085
			590
			259
			2048
			309

Transformed Divergence - Battacharaya Distance measure

$0.0 < x < 1.0$ (poor separability)

$1.0 < x < 1.9$ (moderate separability)

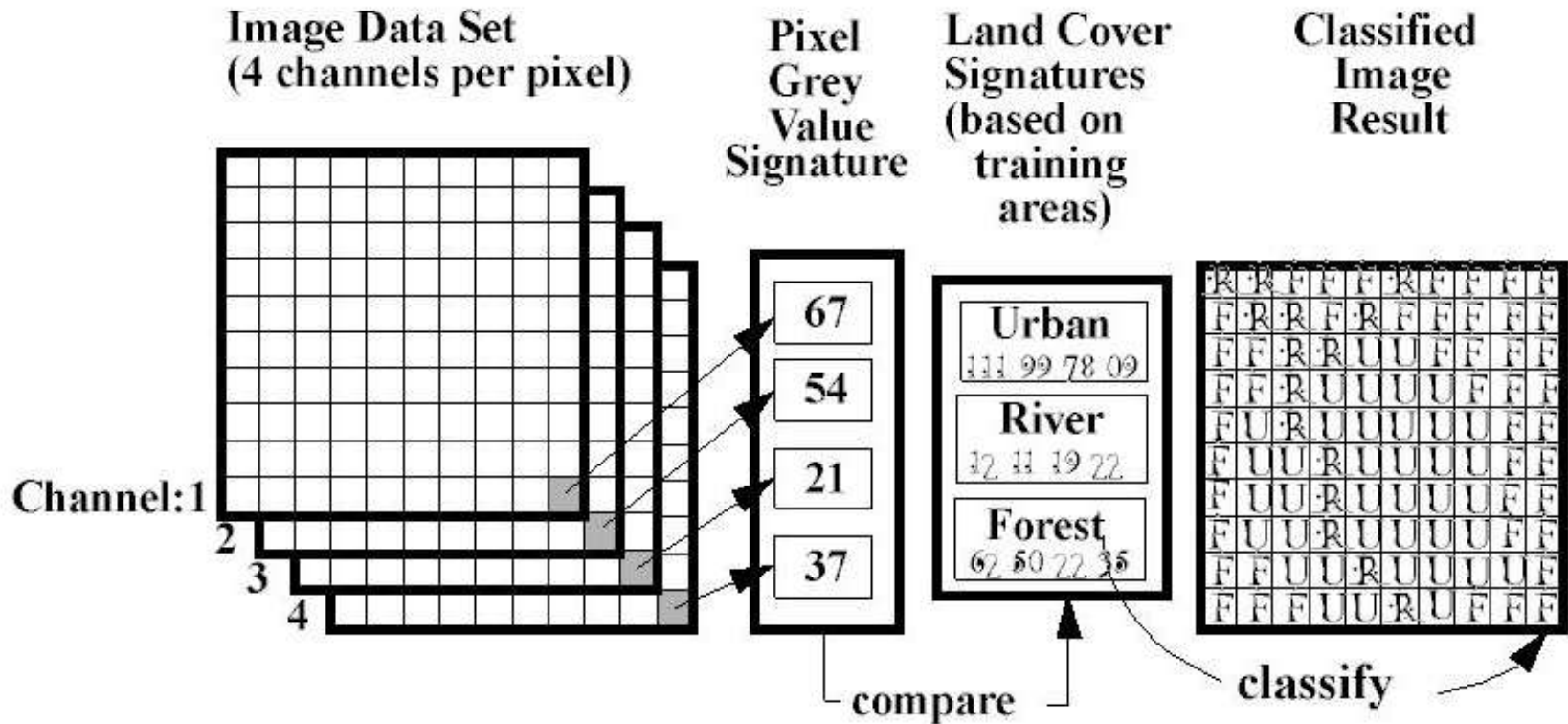
$1.9 < x < 2.0$ (good separability)

Poor separability ($0.0 < x < 1.0$) indicates that the two signatures are statistically very close to each other. You have two options:

One signature can be discarded (suggested when the separability is closer to 0), or the two signatures can be merged using **Merge** option (suggested when the separability is closer to 1).

OK ? ... ready to run the classifier

Supervised - class assignment



Per pixel classifiers

Supervised classification methods: a. Minimum distance

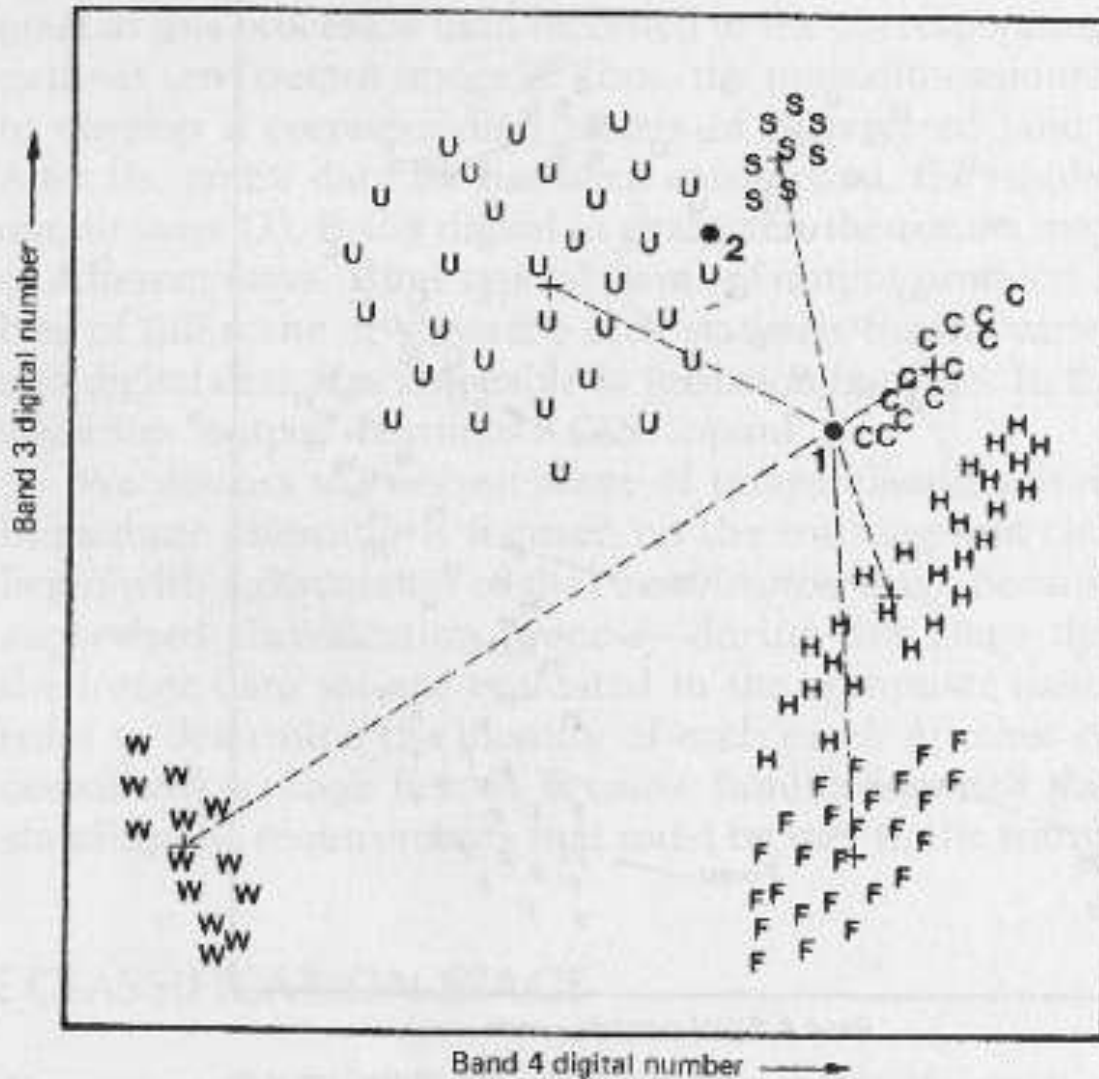


Figure 7.40 Minimum distance to means classification strategy.

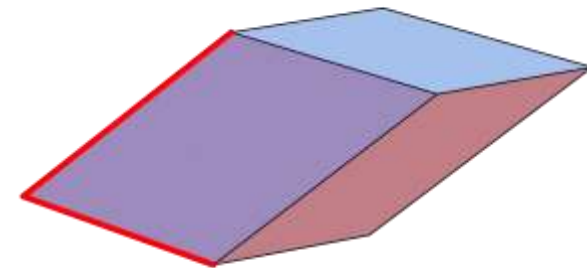
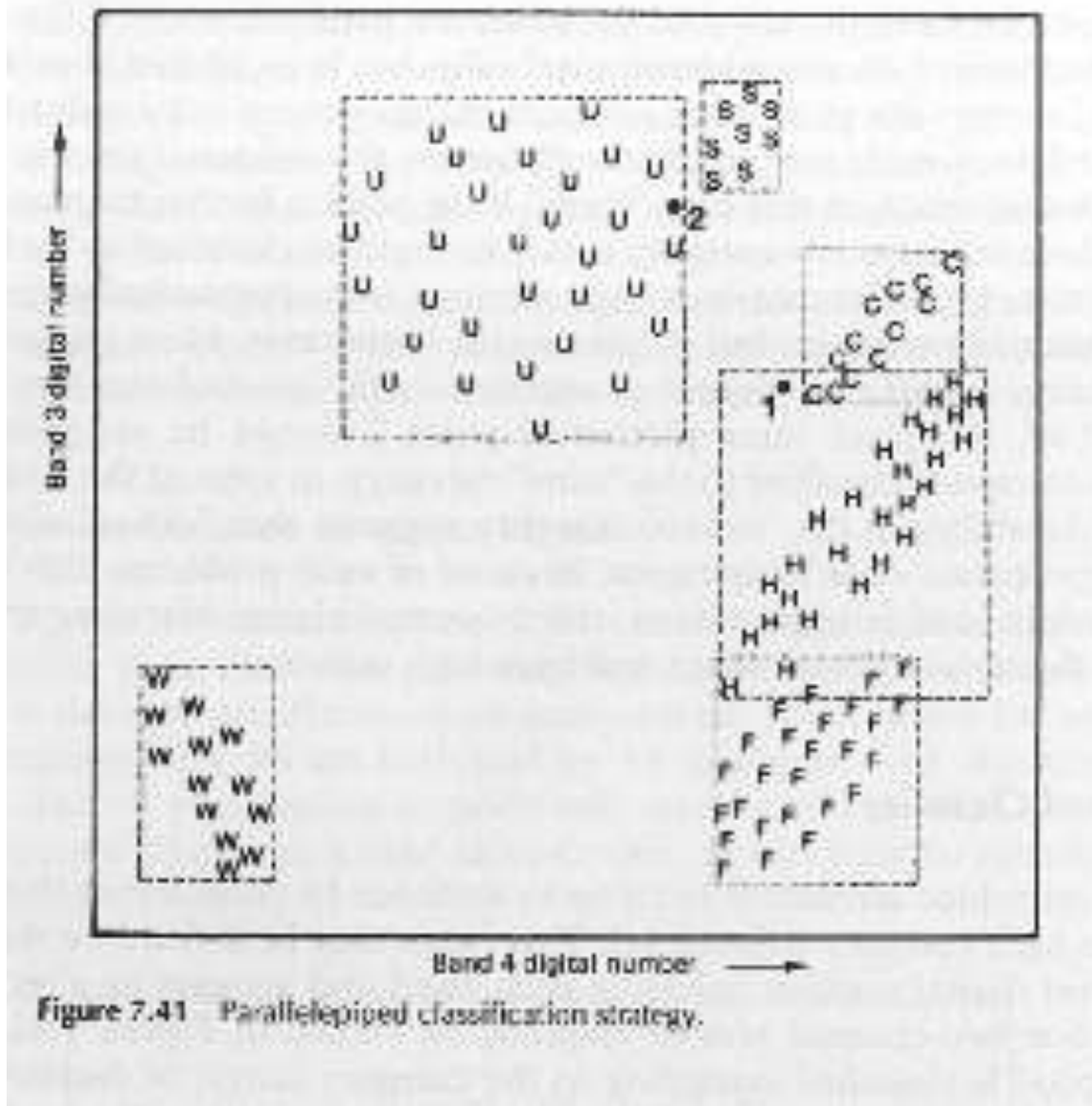
The graphic is 2D

Letters indicate a training pixel

Think in n-dimensions:

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

Supervised classification methods: b: Parallelepiped ??



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

Supervised classification methods

c: Maximum likelihood

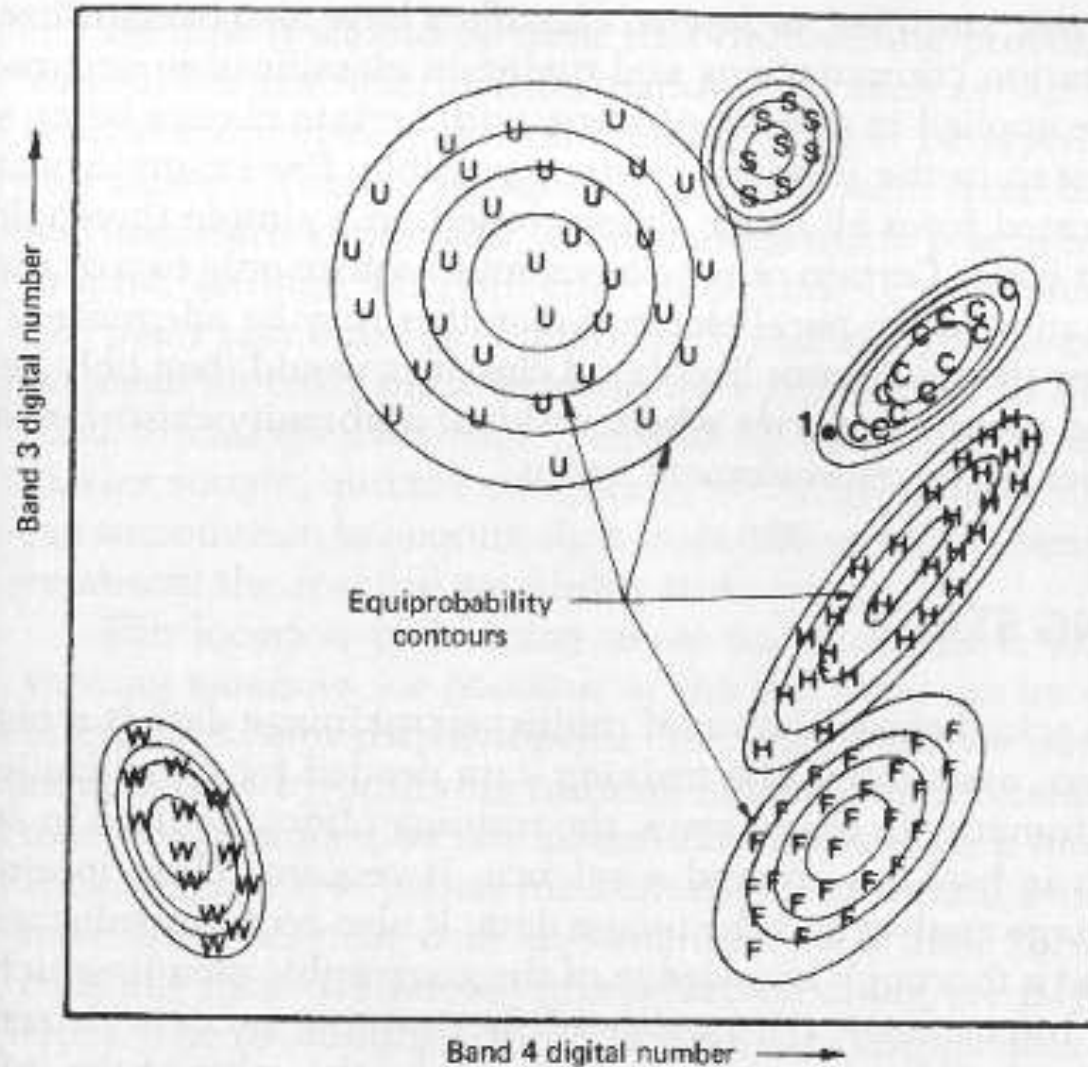


Figure 7.44 Equiprobability contours defined by a maximum likelihood classifier.

With or without null class

Supervised classification: how it works

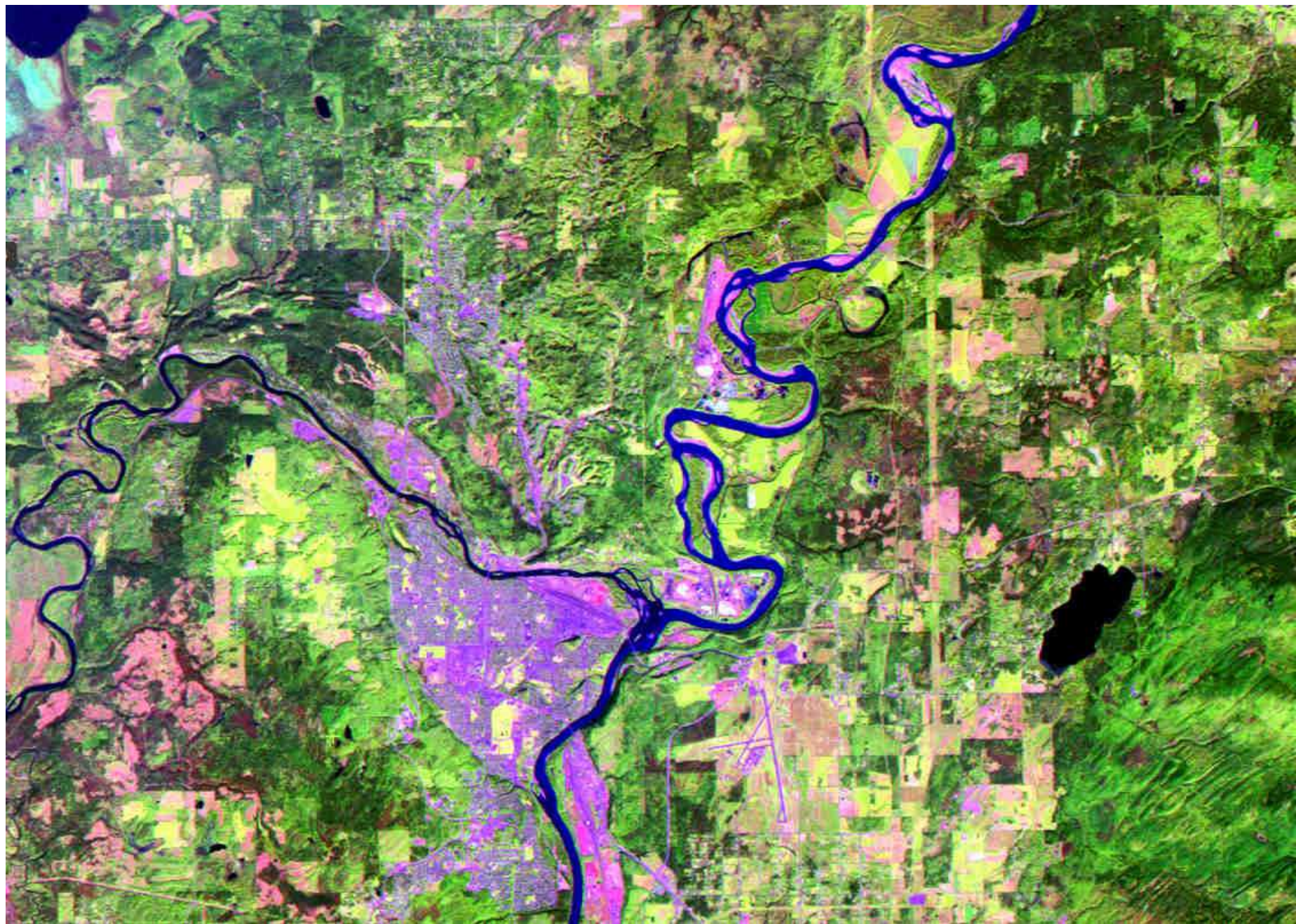
Minimum distance: each pixel is assigned to the class whose mean is closest to data point

(in n-dimensions)

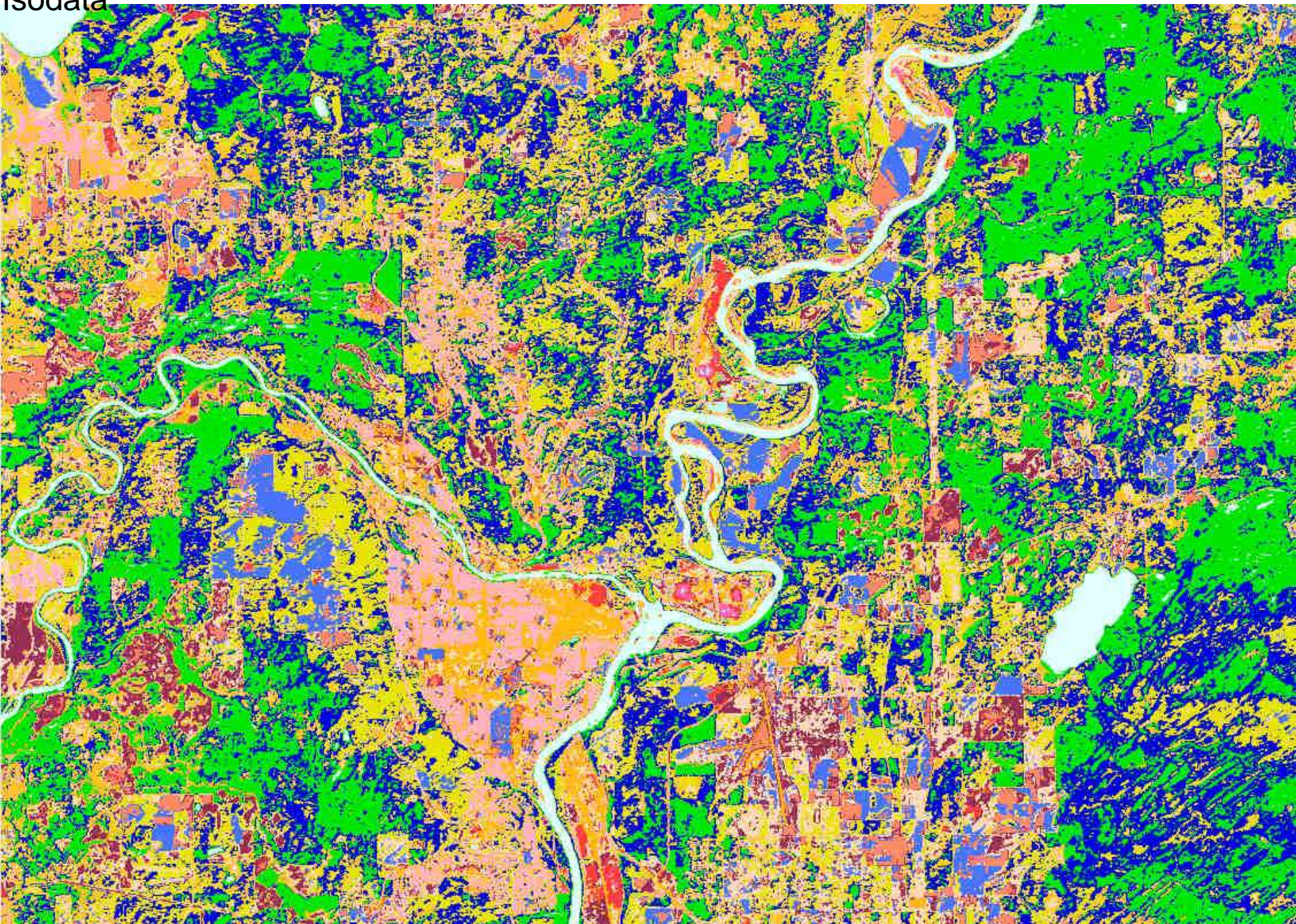
Parallelepiped: Each pixel is assigned to the class whose range it falls in
(overlap = double assignment)

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)



Isodata



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]

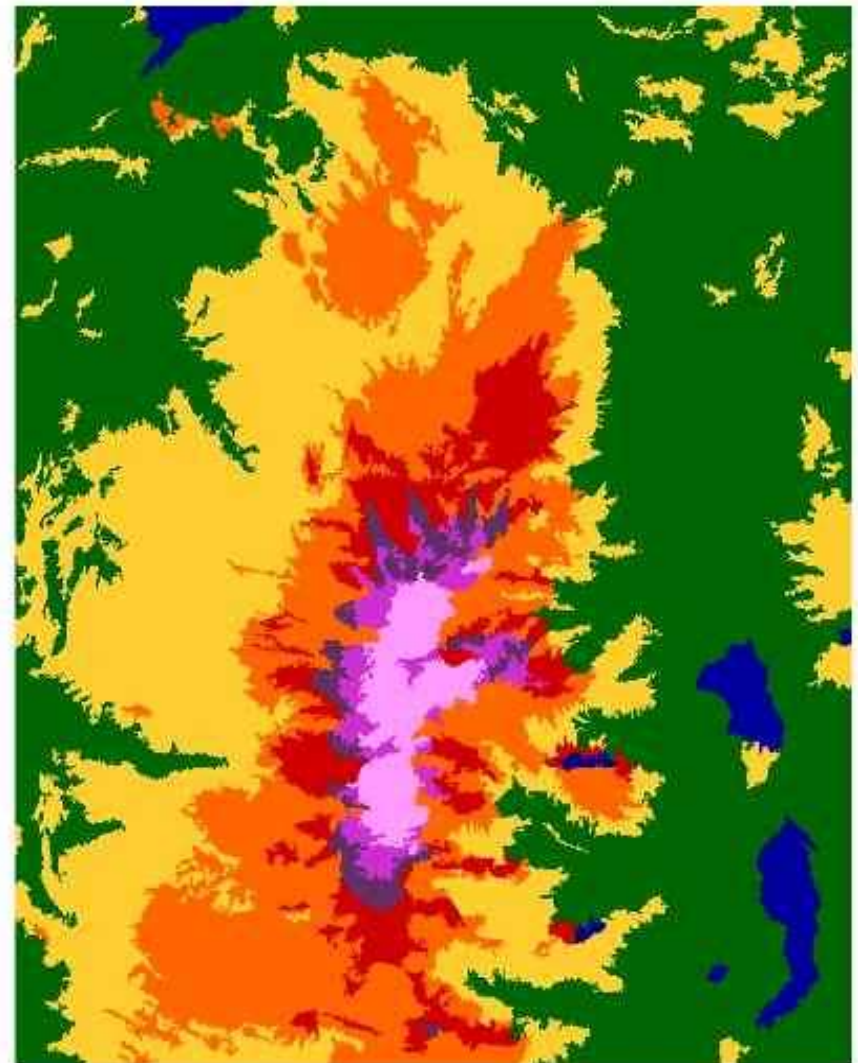
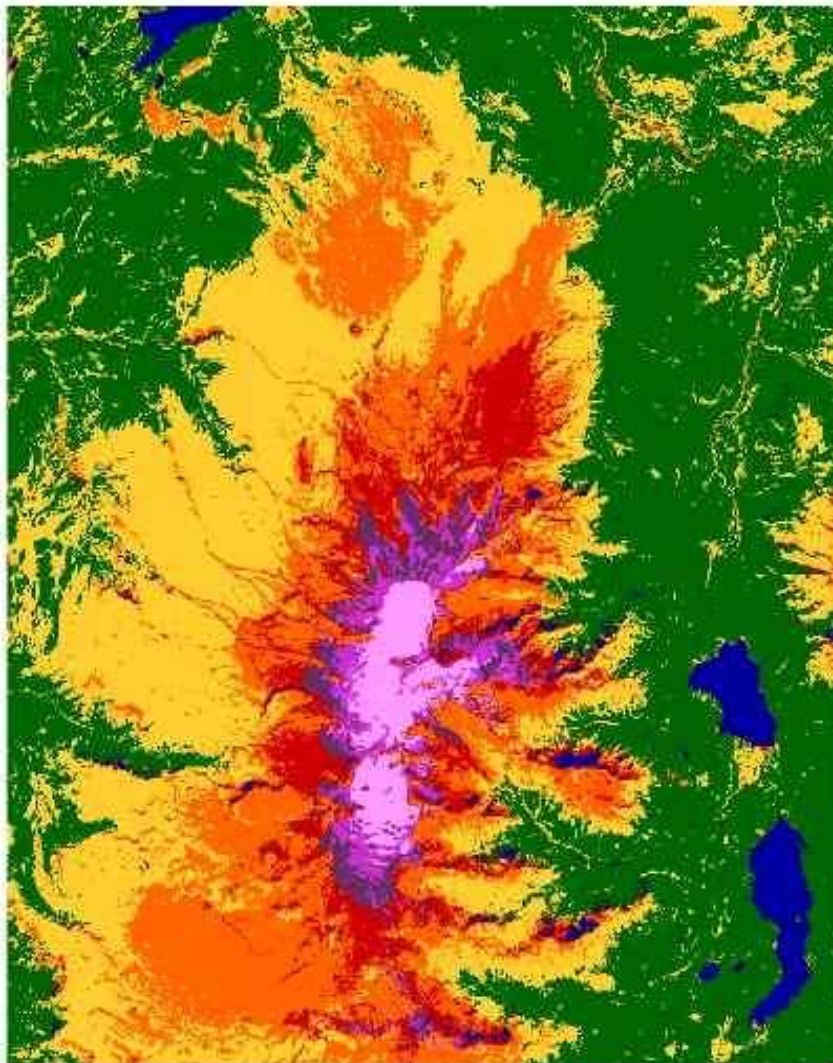
Supervised:- create new training class or delete some training areas

Areas are unclassified - create new training class

Post-classification steps

- Check the display
- Merge / add classes
- Sieve ... to remove isolated pixels
- Accuracy assessment
- Conversion of results to vectors - see lab 6

Mt. Edziza - classification and sieve - removing isolated pixels



- recognises connectivity of adjacent pixels in the same class
- special classes e.g. lakes or wetlands can be specified and preserved

Accuracy assessment

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

		Reference test information				Row total	User's Accuracy
		Class	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'

Error matrix – pixel classifications, right or wrong ?

Columns – ground truth data – what were these pixels ?

Rows: how were these pixels classified ?

Classes	River	Built Up Area	Vegetation	Agri Land	Waterbody	Vacant /Soil	Classified Totals	Omission Error	Commissions Error
River	36	0	1	0	4	1	42	14%	14%
Built Up Area	1	38	0	0	0	3	42	10%	10%
Vegetation	0	0	37	4	0	1	42	17%	12%
Agri Land	1	0	4	35	1	1	42	14%	17%
Waterbody	4	0	1	0	37	0	42	12%	12%
Vacant/Soil	0	4	1	2	0	35	42	14%	17%
Classified Totals	42	42	44	41	42	41	252		

Producer's accuracy: based on ground truth pixels

User's accuracy: based on classified pixels

Measuring accuracy

Producer's accuracy: based on ground truth pixels

User's accuracy: based on classified pixels

Error (Confusion) Matrix Classified

Reference Data	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

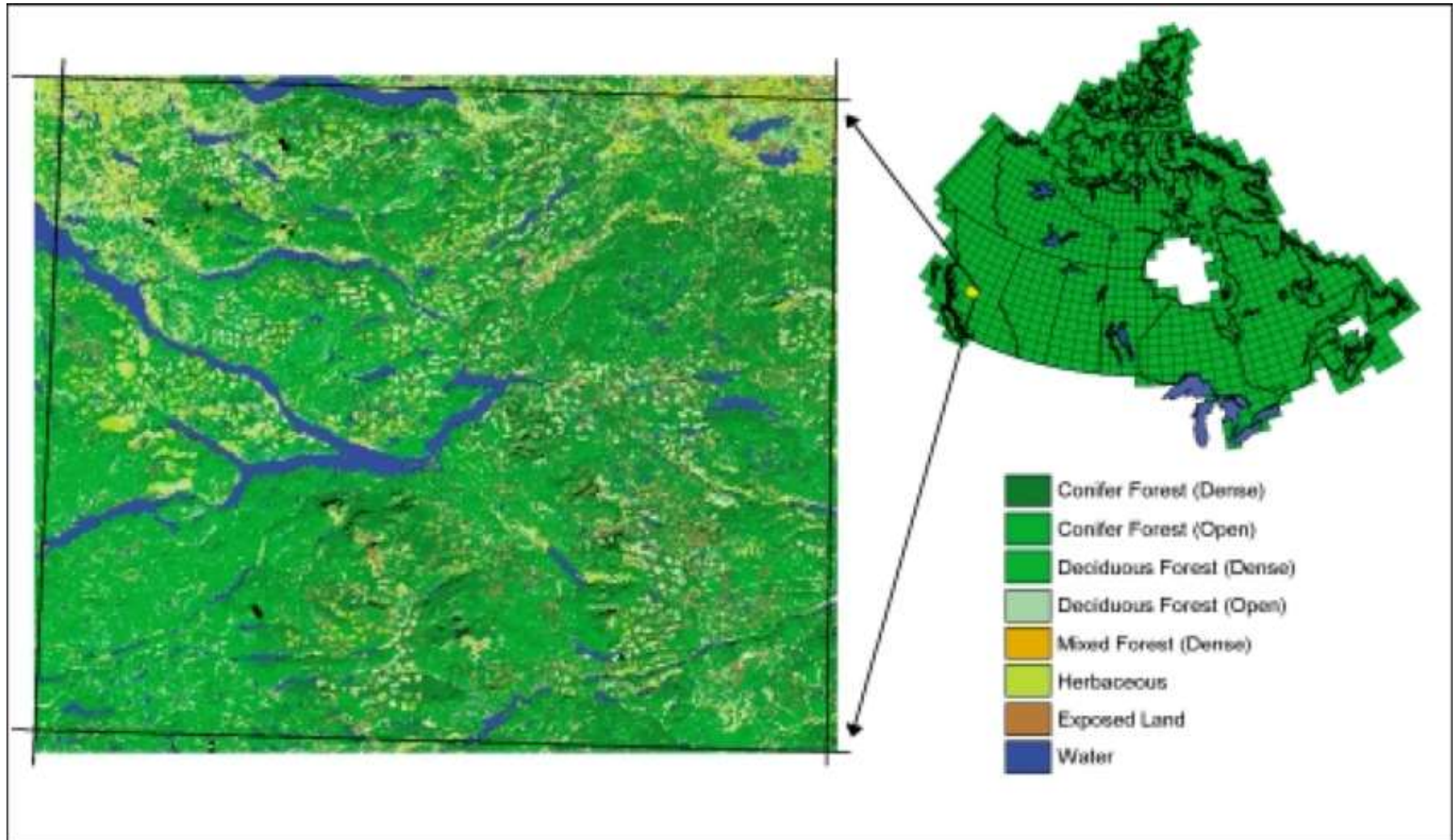
Kappa: a composite accuracy index:

- 0.7 = good;
- < 0.2 = no agreement

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



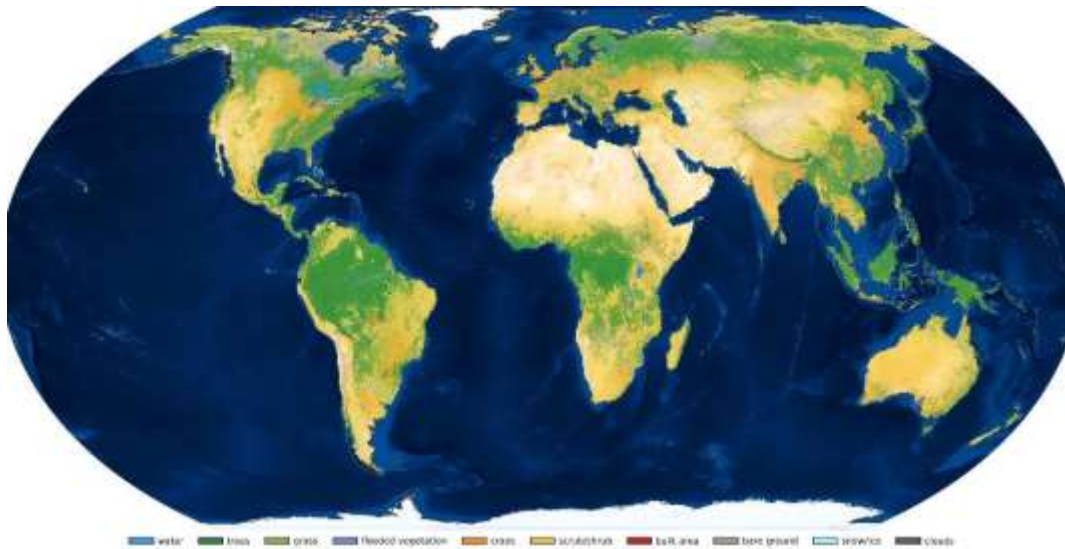
Global Sentinel classification (Esri)

<https://www.bbc.com/news/science-environment-57615408>

<https://www.arcgis.com/home/item.html?id=d6642f8a4f6d4685a24ae2dc0c73d4ac>

Global viewer

<https://caitlin-kontgis.medium.com/mapping-the-world-in-unprecedented-detail-7c0513205b90>



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

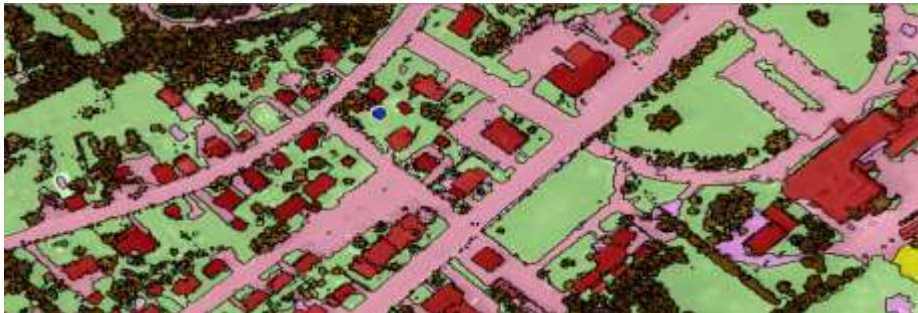
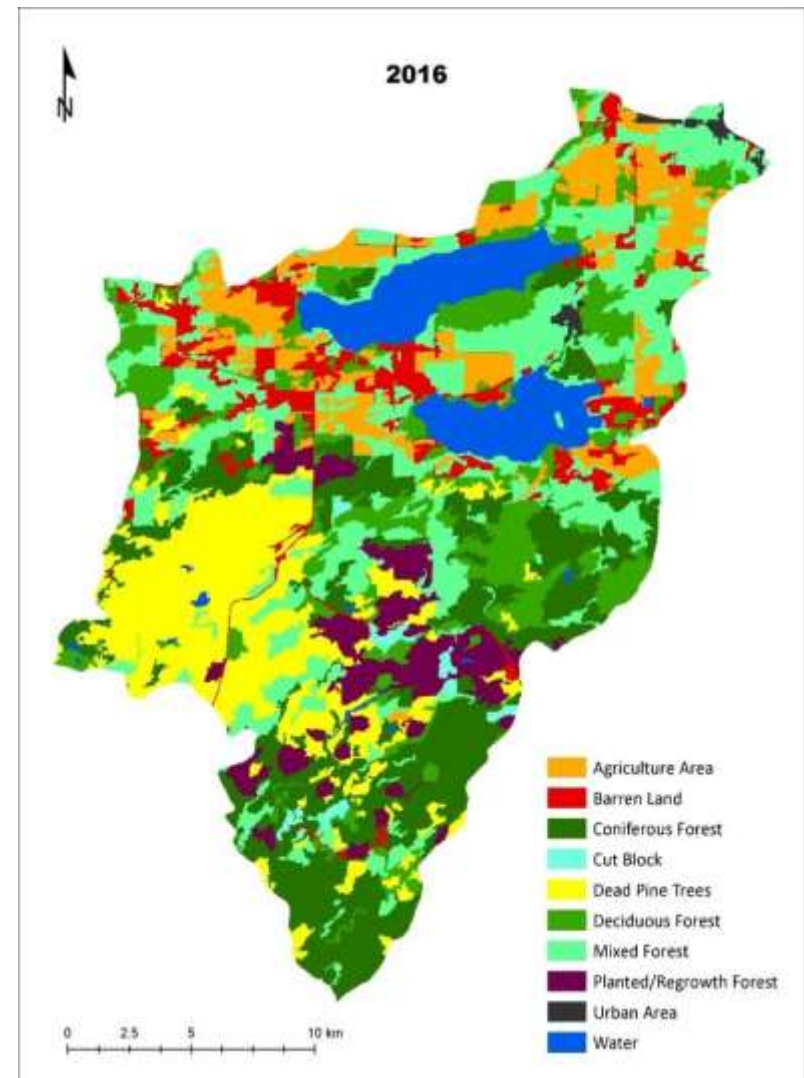
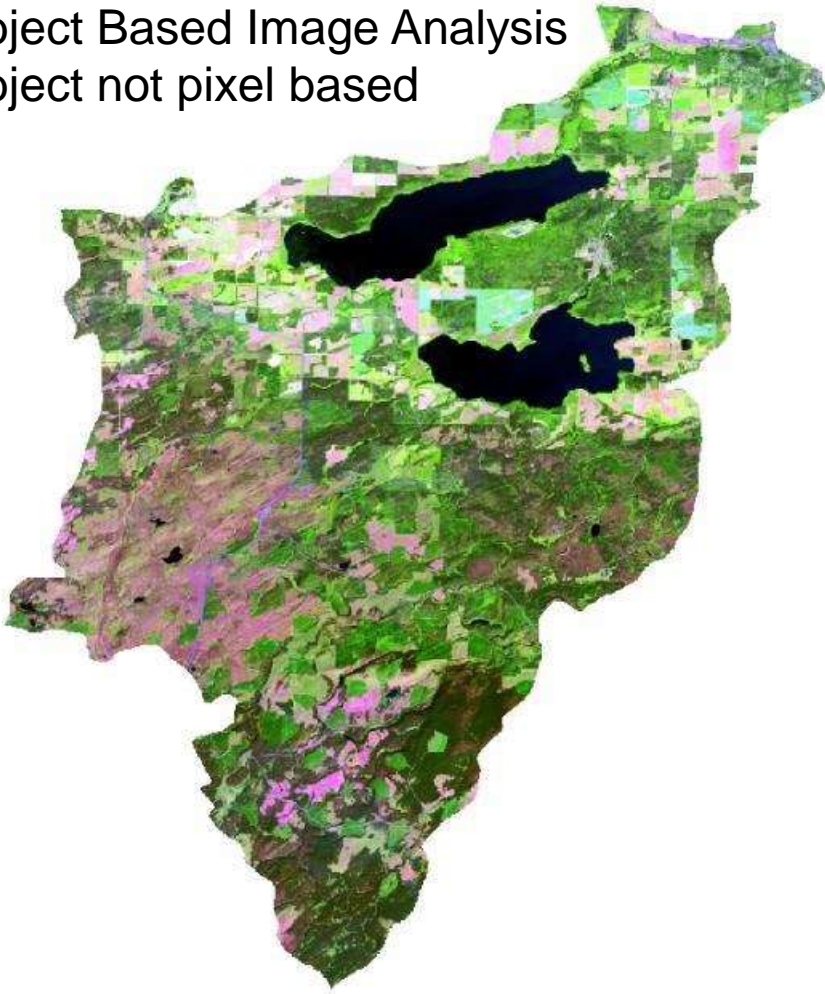
Classification summary

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

Object Based Image Analysis
Object not pixel based



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