Dirty REMOTE SENSING

Lecture 4: Classification

Stuart Green

Earthobservation.wordpress.com

Stuart.Green@Teagasc.ie

Classifying

- Replacing the digital numbers in each pixel (that tell us the average spectral properties of everything in the pixel) , with a single number-a code that might represent the majority landuse/cover in the pixel, a biophysical property in the pixel (amount of biomass) or a relative value for a Landover (percentage of pixel that is forestry).

Main Routes to creating a Thematic Landcover Map

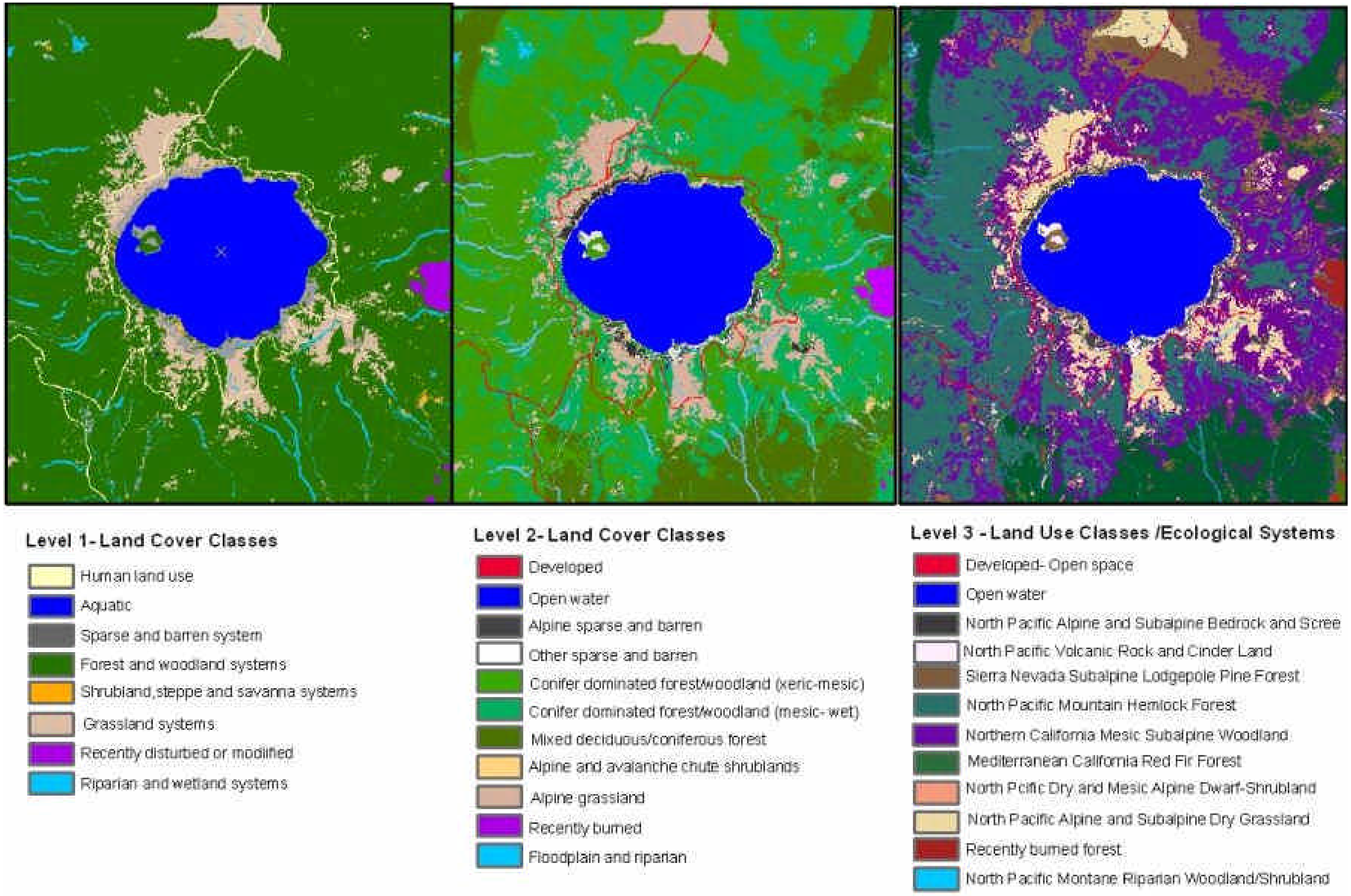
- Manual Digitisation and Interpretation
- Photomorphic Labelling
- Unsupervised Classification
- Supervised Classification
- Object Orientated
- Hybrid Approaches

Whats the differnce between landuse and landcover

- **http://oceanservice.noaa.gov/facts/lclu.html**

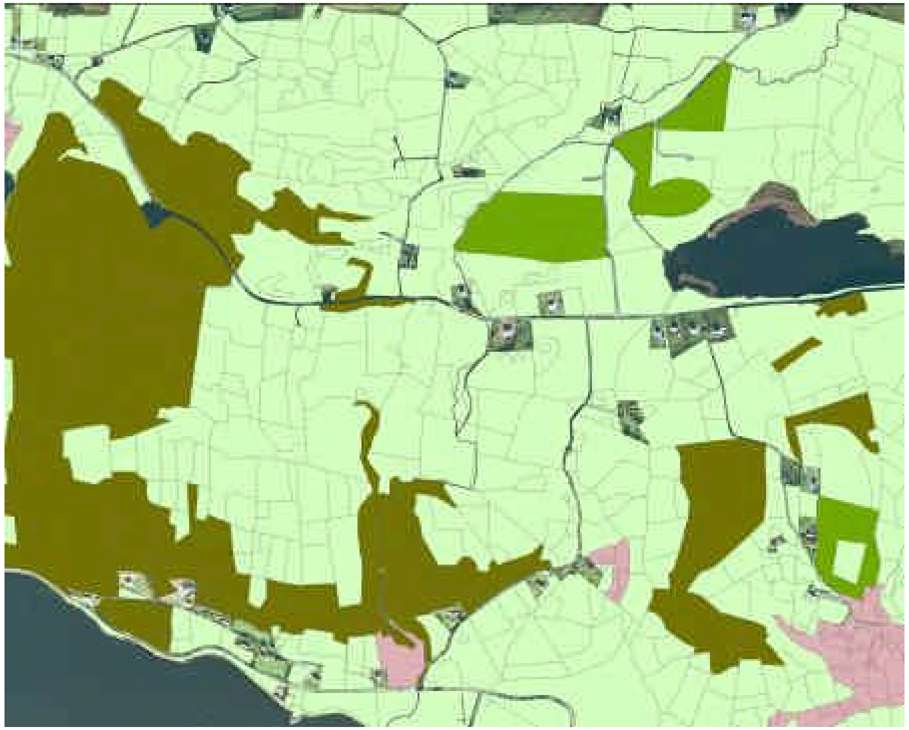
- **Land cover indicates the physical land type such as forest or open water whereas land use documents how people are using the land**

Or landcover is what is under your feet- landuse is what you might do with that



http://conservationmaven.com/frontpage/first-detailed-national-map-of-us-land-cover-vegetation-rele.

Manual Digitisation as in LPIS for SFP



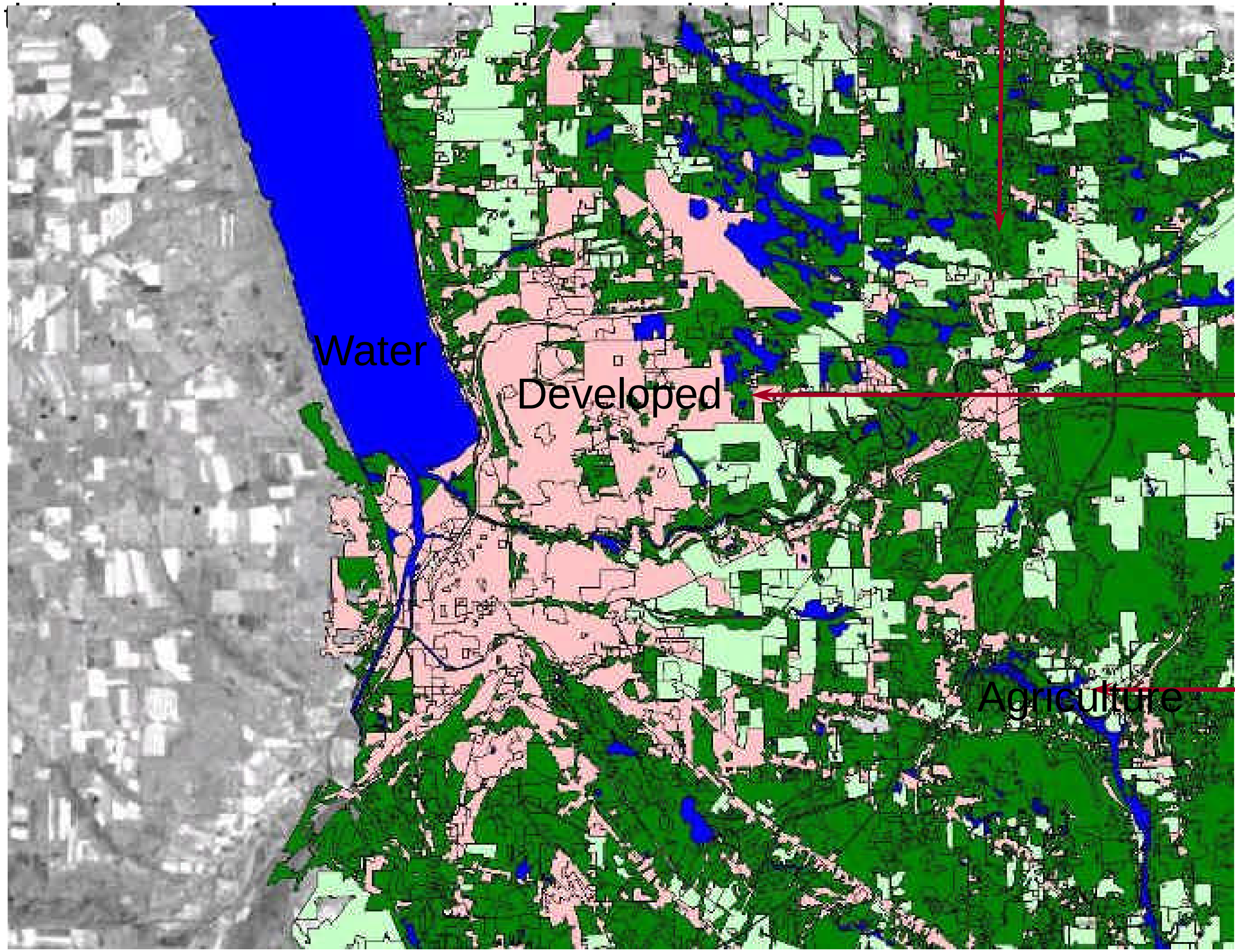
UNSUPERVISED CLASSIFICATION

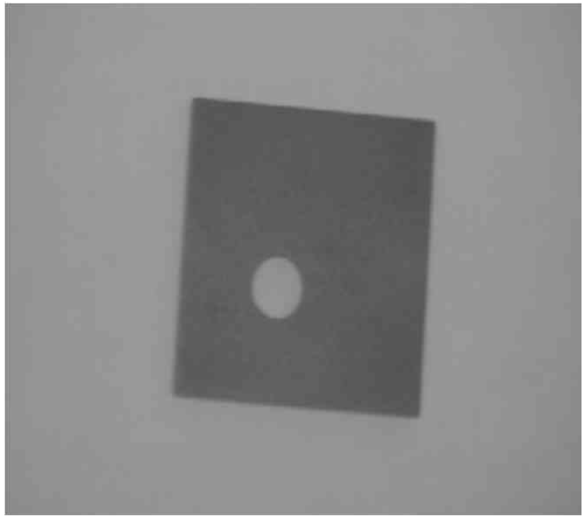
The computer performs a clustering exercise on the image:
The user tells the computer how many clusters to look for
and the computer then analysis the image to \produce this number
of statistaclly sound clusters.
Most commonly the ISODATA algorithm is used

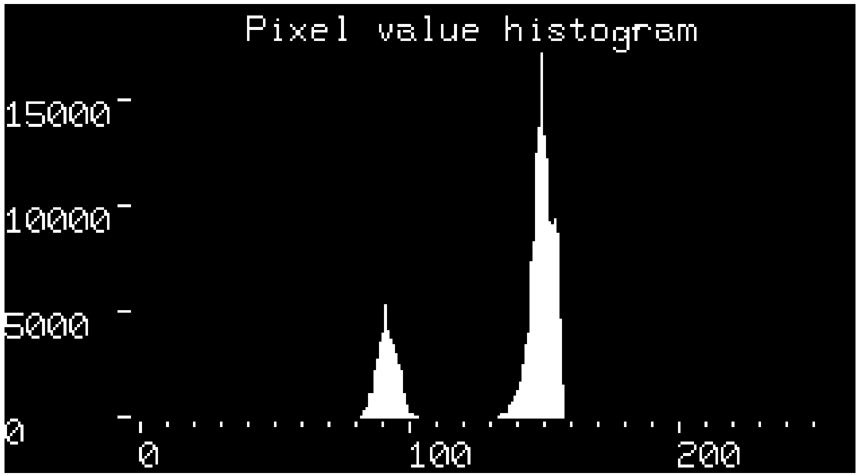
Principle of Classification

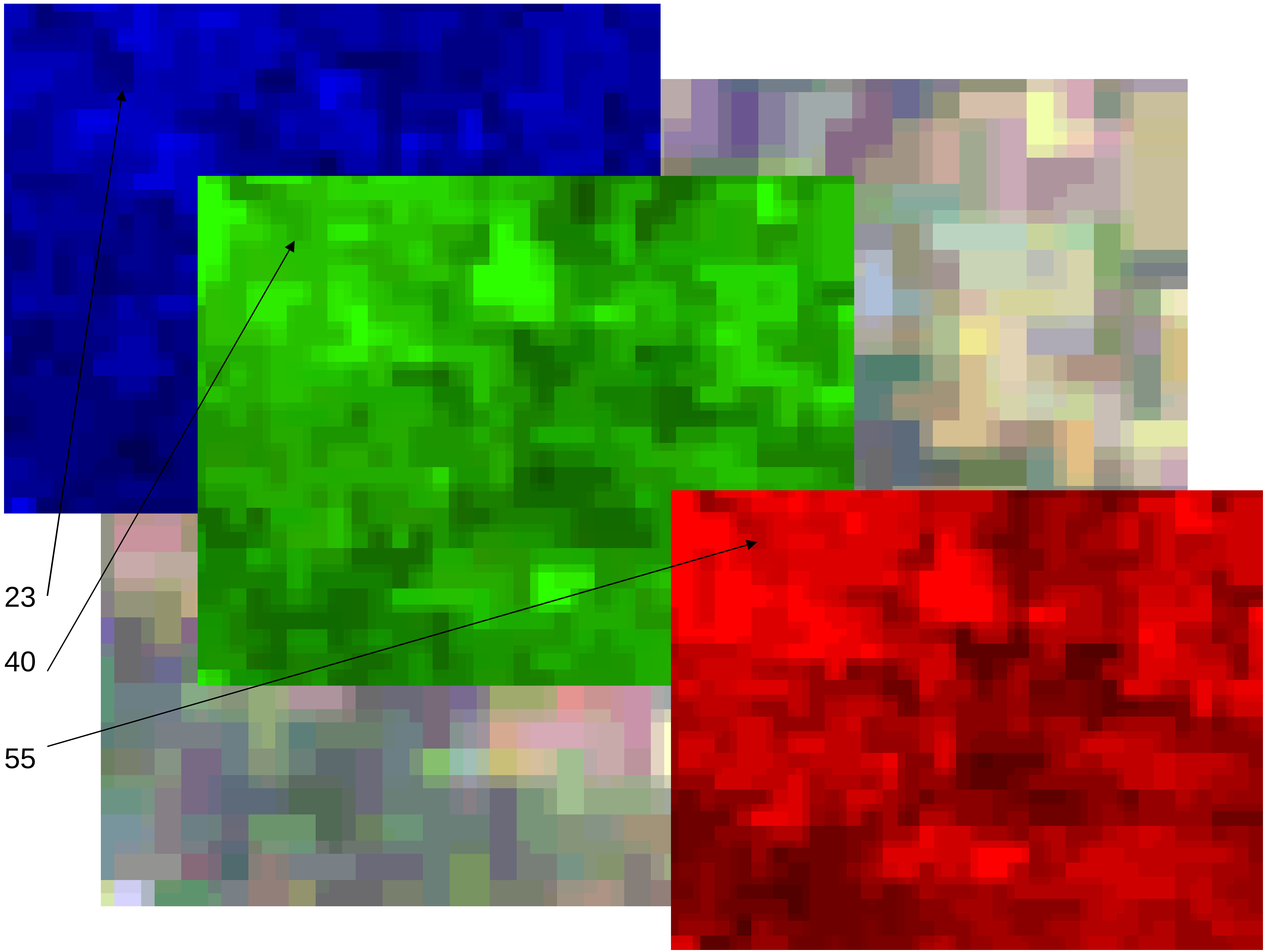
- Each class is known as a *cluster (or a theme)*. It is possible, using statistics draw boundaries between clusters.
- Unsupervised classification does not require prior knowledge. This type of classification relies on a computed algorithm which clusters pixels based on their inherent spectral similarities.

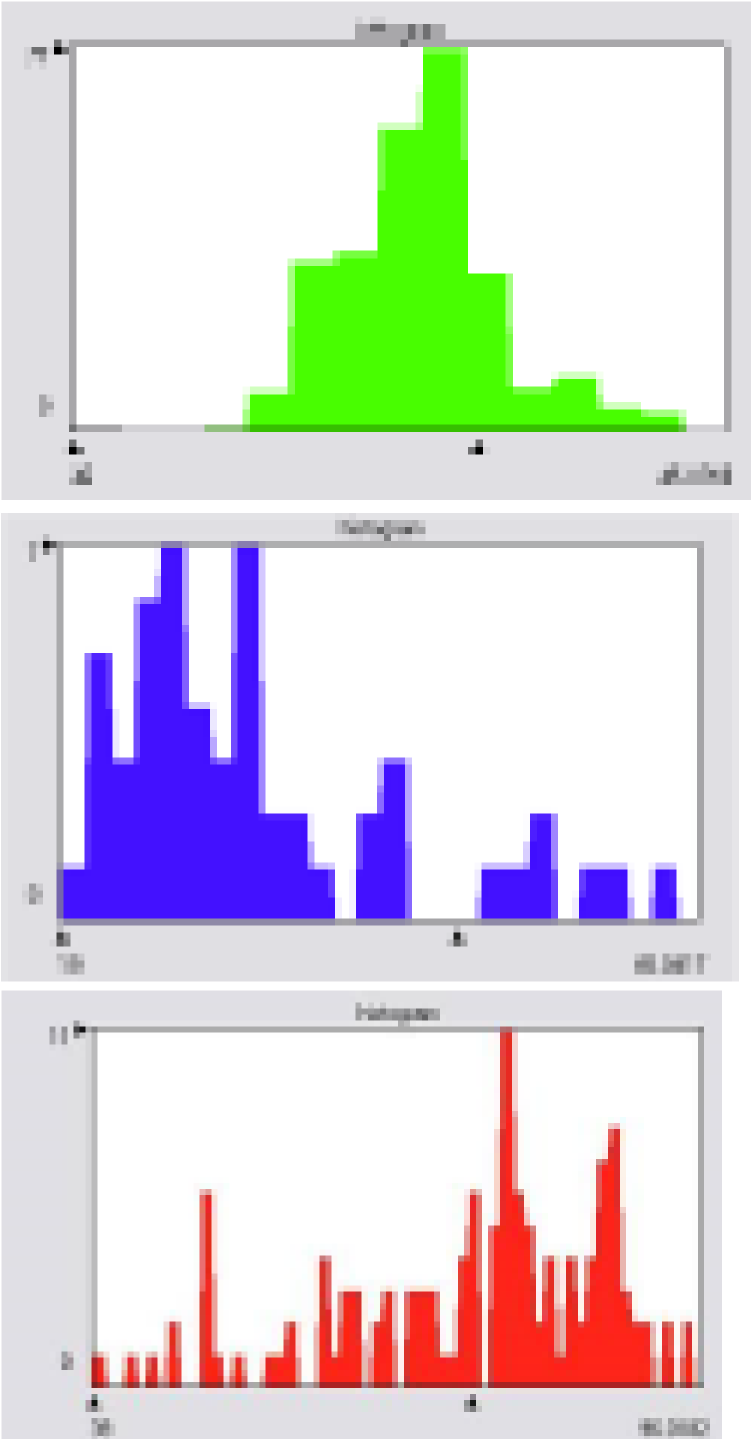
- the concept of image classification can also be described as:
- using the brightness values in one or more spectral bands, and classifies each pixel based on its spectral information The goal in classification is to assign remaining pixels in the image to a designated class such as water, forest, agriculture, urban, etc.
- The resulting **classified** image is comprised of a collection of pixels, colorcoded to represent a particular theme. The overall process then leads to the creation of a

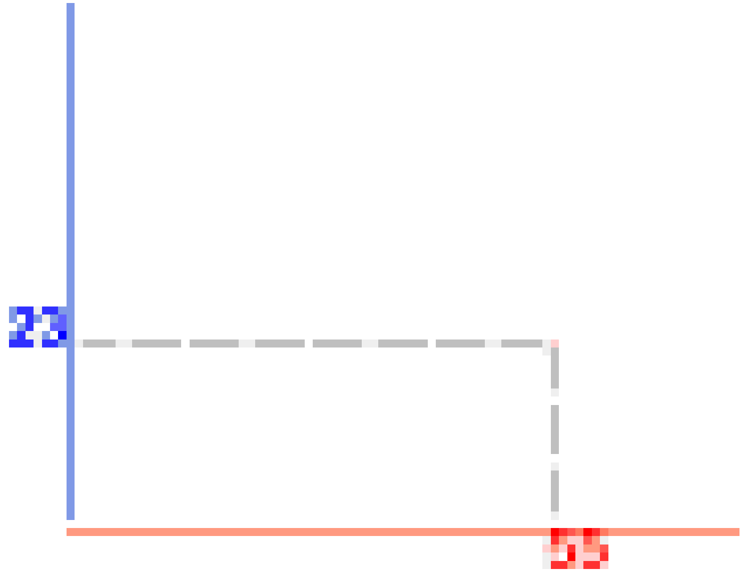


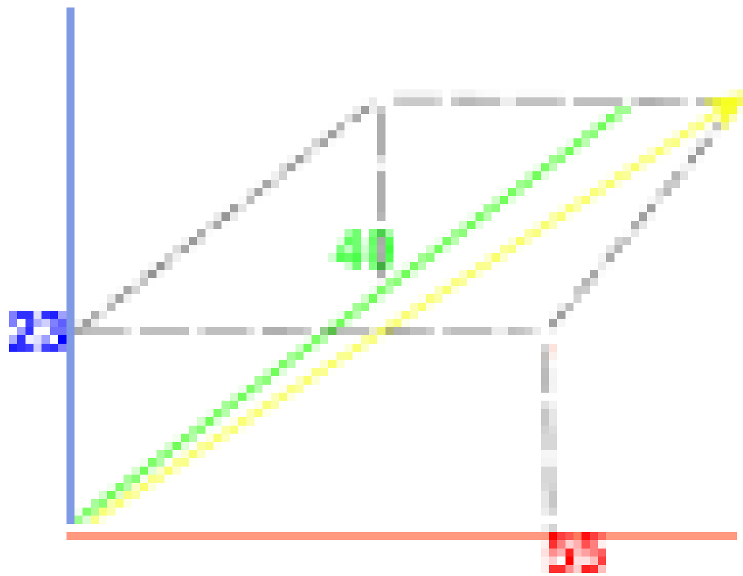



Pixel value histogram
15000
10000
5000
0
0
100
200

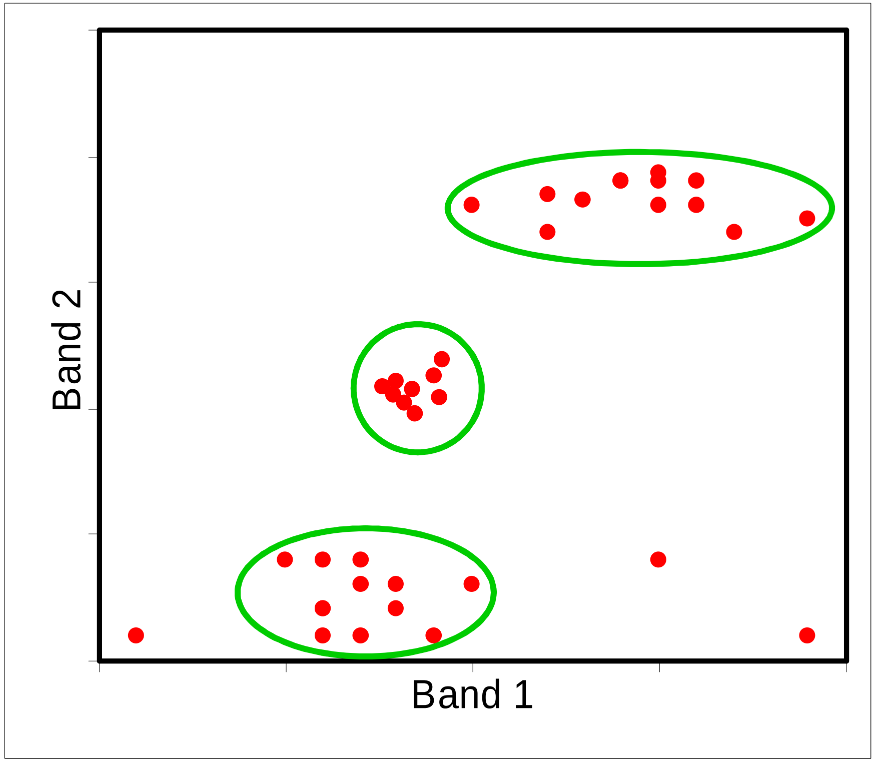

23
40
55




23
55


40
23
55

Example spectral plot



- Two bands of data.
- Each pixel marks a location in this 2d spectral space
- Our eye's can split the data into clusters.
- Some points do not fit clusters.

Rory Hutson
Remote Sensing Group
Plymouth Marine Laboratory

ISODATA clustering procedure

• Begin by distributing N arbitrary cluster means.

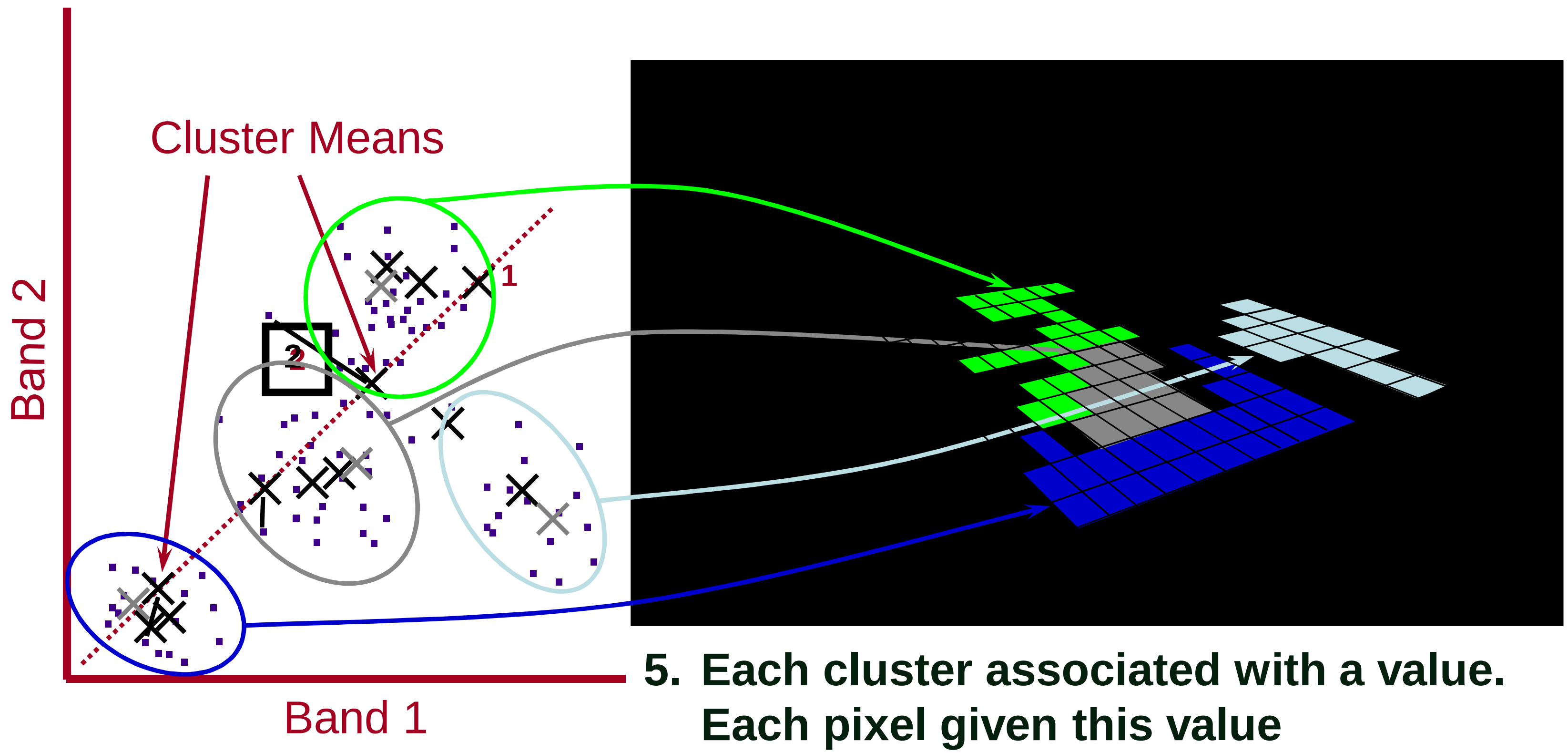
• The spectral distance between the candidate pixel and each cluster mean is calculated. The pixel is assigned to the cluster whose mean is the closest.

• After each iteration, the means for each cluster are recalculated, based on the actual spectral locations of the pixels in the cluster, causing them to shift in feature space. These new cluster means are used for defining clusters in the next iteration.

• The process will be repeated until either the convergence threshold T or the maximum number of iterations M is reached.

Separate Data into Groups with unsupervised classification

- Unsupervised classification- the objective is to group multi-band spectral response patterns into clusters that are statistically separable
- Our example uses 3 bands
  - More bands can be used, but it can’t be shown in this 3-D plot
- A = Agriculture; D= Desert; M = Mountains; W = Water



5. Each cluster associated with a value. Each pixel given this value

We can modify these clusters, so that their total number can vary arbitrarily.

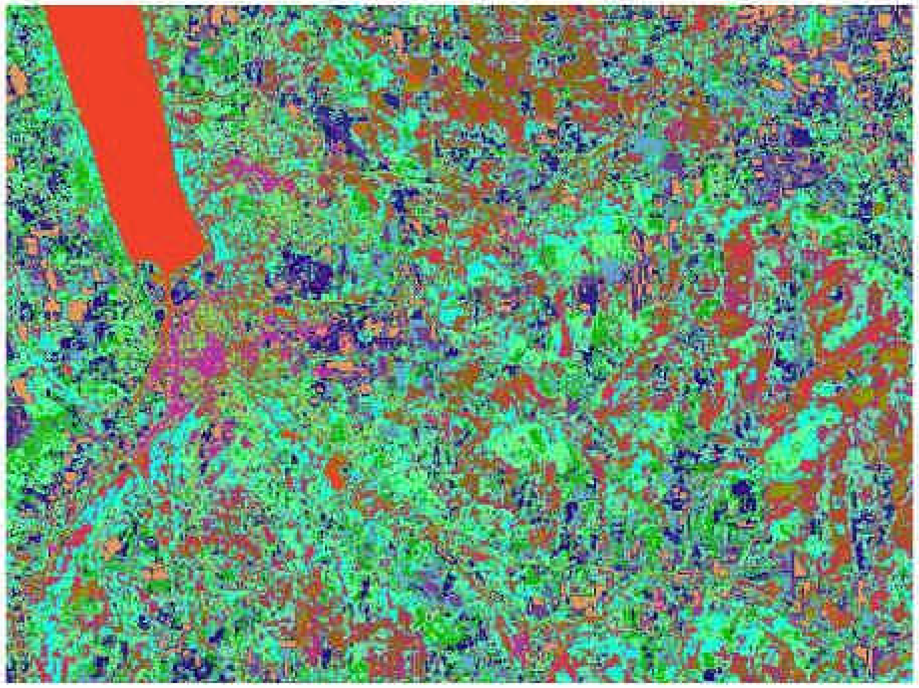
When we do the separations on a computer, each pixel in an image is assigned to one of the clusters as being most similar to it in DN combination value.

Generally, in an area within an image, multiple pixels in the same cluster correspond to some (initially unknown) ground feature or class .

The trick then becomes one of trying to relate the different clusters to meaningful ground categories. We do this by either being adequately familiar with the major classes expected in the scene, or, where feasible, by visiting the scene (**ground truthing**) and visually correlating map patterns to their ground counterparts.

Classify Data into Groups

Unsupervised classification using 20 different categories was carried out. Now, the task will be to group these categories into some kind of smaller grouping. In our case we have been using 5 classes: Agriculture, Developed, Natural, Forest, and Water.



Obviously, the red is water, we can see the Lake. Also, the purple looks like a city, so we would call that developed.

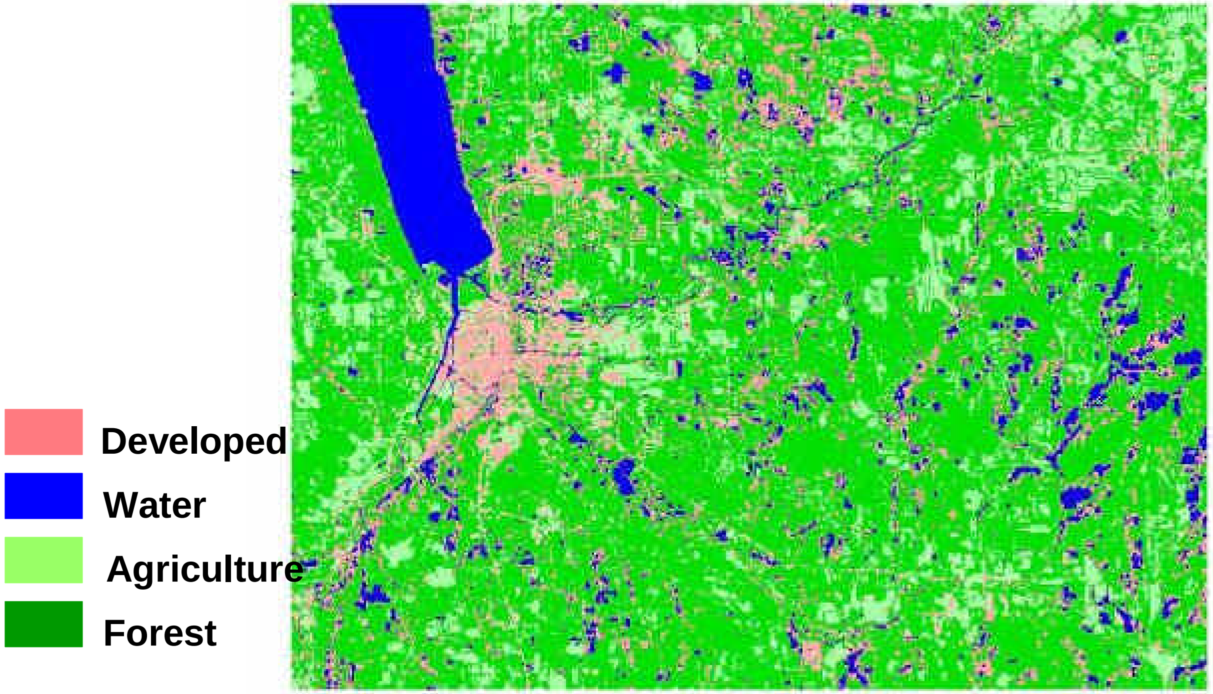
The rest of the colors are anyone's guess. So, the laborious process of assigning a category to the different classes (colors) will now begin.

Assign a name to each group

After about 20 minutes, I was able to assign the classes with four of the categories to create a final land use map.

But where is Natural? This is sometimes a problem in digital image processing. Natural can look like the other classes
  And, based on the digital numbers, we were unable to discriminate the spectral differences

This is known as spectral confusion. We may be able to discriminate between the extractive and developed if we chose more classes, but even then it might not be enough. So, that is part of the struggle we will have as image processors.



Supervised Classification

Supervised classification is much more accurate for mapping classes, but depends heavily on the cognition and skills of the image specialist. The strategy is simple: specialist must recognize conventional classes from prior knowledge.

Training sites are areas representing each known land cover category that appear fairly homogeneous on the image (as determined by similarity in tone or color within shapes delineating the category). Specialists locate and circumscribe them with polygonal boundaries drawn (using the computer mouse) on the image display. For each class thus outlined, mean values and variances of the DNs for each band used to classify them are calculated from all the pixels enclosed in the site.

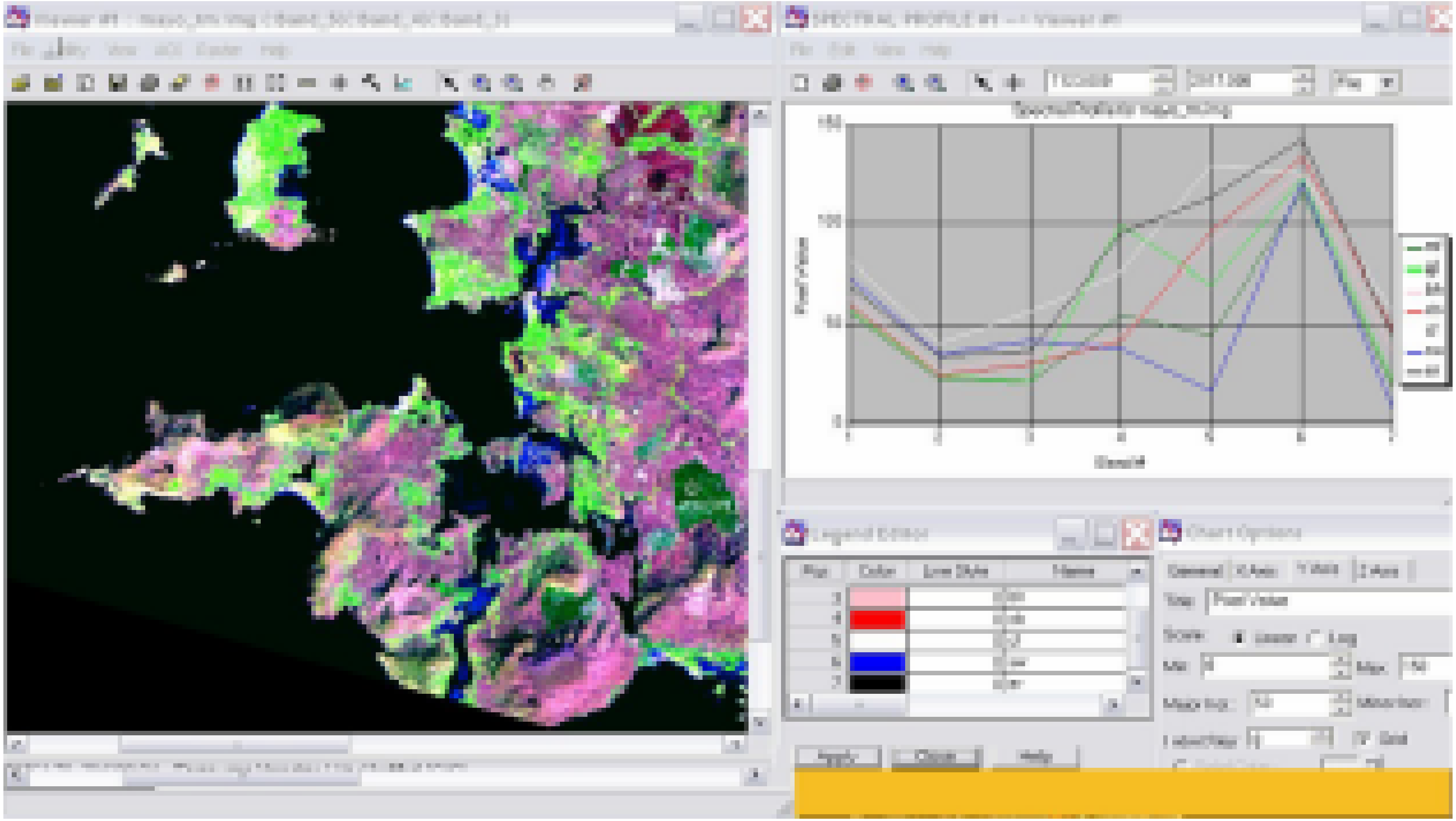
When DNs are plotted as a function of the band sequence (increasing with wavelength), the result is a spectral signature or spectral response curve for that class.

Classification now proceeds by statistical processing in which every pixel is compared with the various signatures and assigned to the class whose signature comes closest. A few pixels in a scene do not match and remain unclassified, because these may belong to a class not recognized or defined).

Supervised Classification

Steps

1. Decide on classes.
2. Choose “training areas” which represent these classes. These areas have to have satisfy certain spectral statistics.
3. Use the training area with a classifier algorithm to run classification.
4. Test classification, if necessary adjust training areas and re-run until acceptable levels of accuracy are achieved (usually >85%)



Pick themes you want to map. These have to be “mappable”. Ie within the capabilities of being detected with the sensor/image you are using.

So they have to be greater than the resolution of your image (in our case 25m)

They have to be observable (for optical systems this means “visible” on the surface)

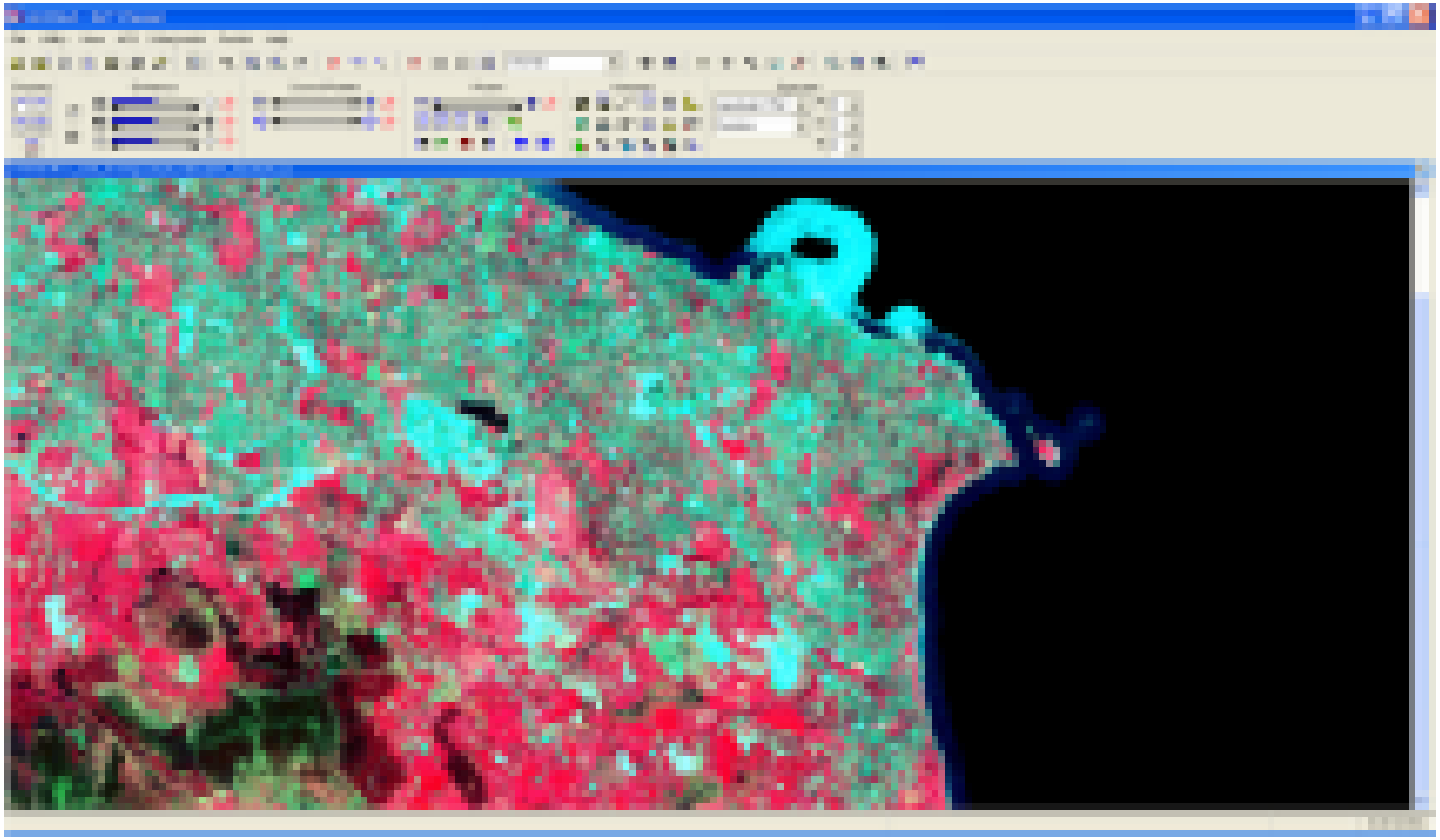
TEGASC LANDCOVER MAP

Bog & Heath
Cut Bog
Wet Grasslands
Dry Grasslands
Rocky Complexes
Bare Rock
Mature Forest
Immature Forest & Scrub
Water
Built Land
Coastal Complexes
Cut&eroding bog

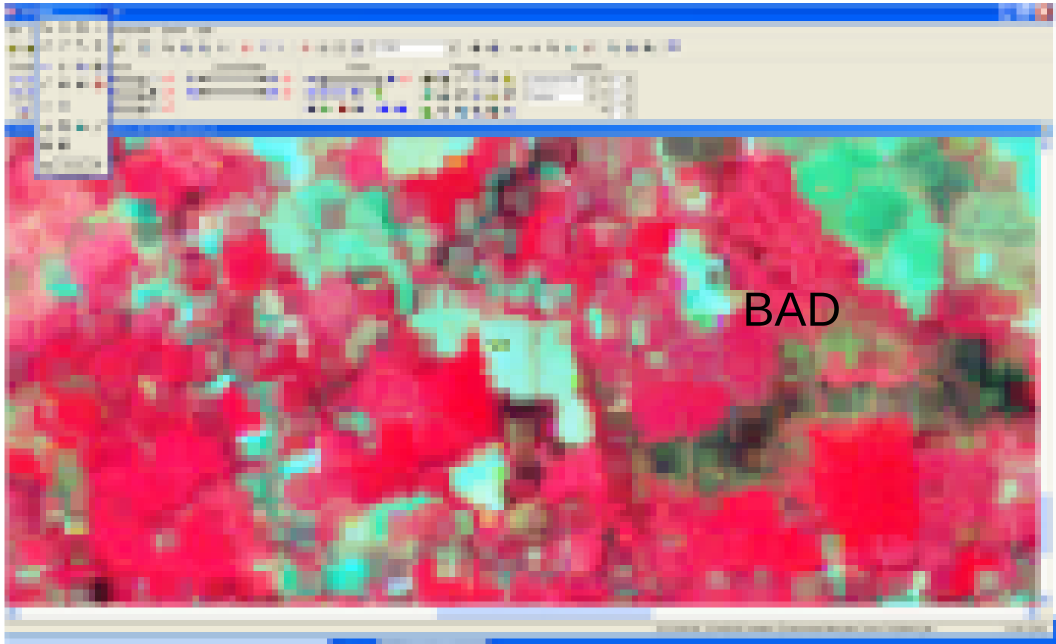
- Why bog & Heath ?*.datortho sing

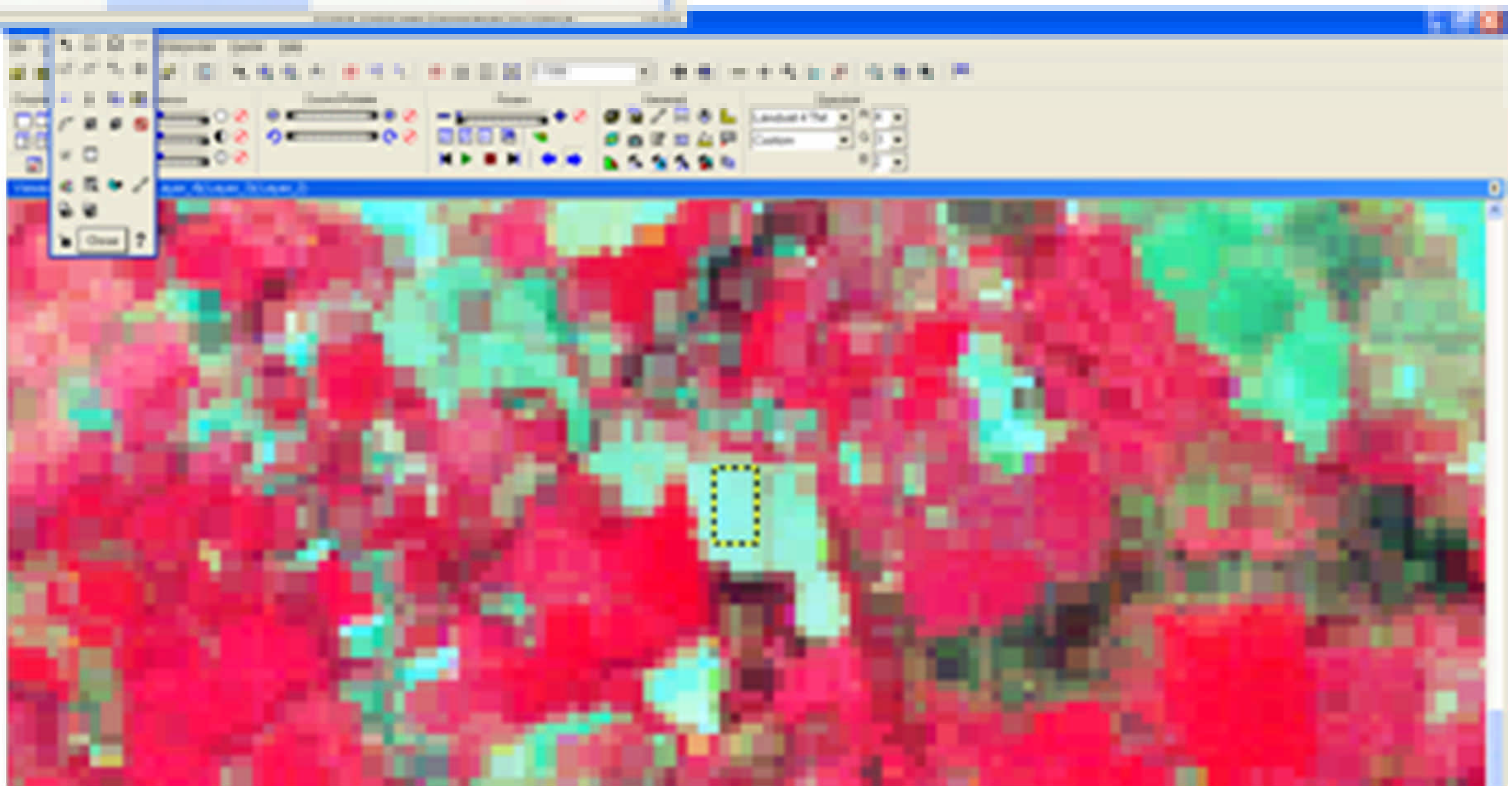
Identify and delinate *training areas*

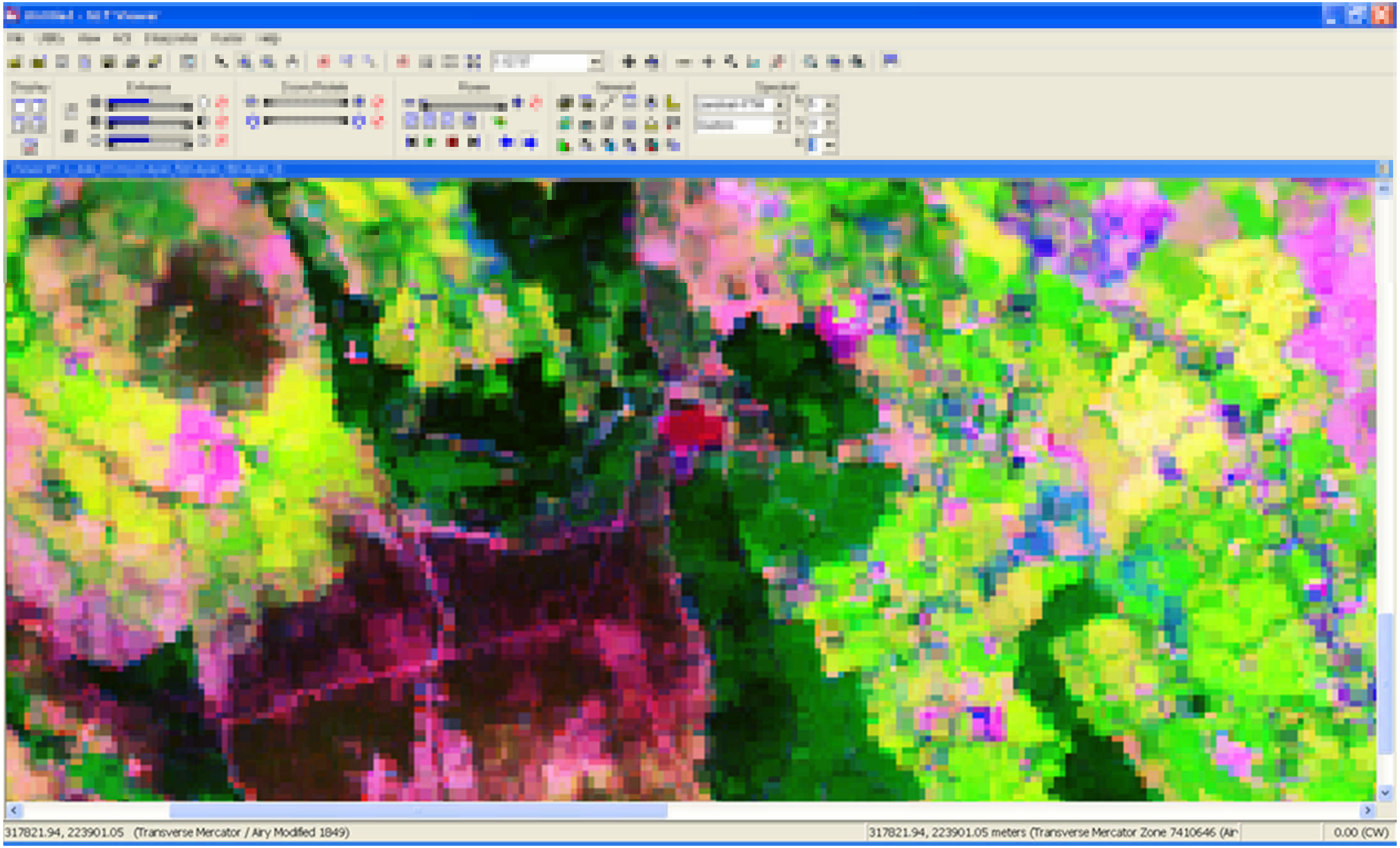
- These have to be good, homogenoues areas that truly represent the cover type you are interested in. Most cover types will have a number of traing areas to fully define the range of values present in the image.

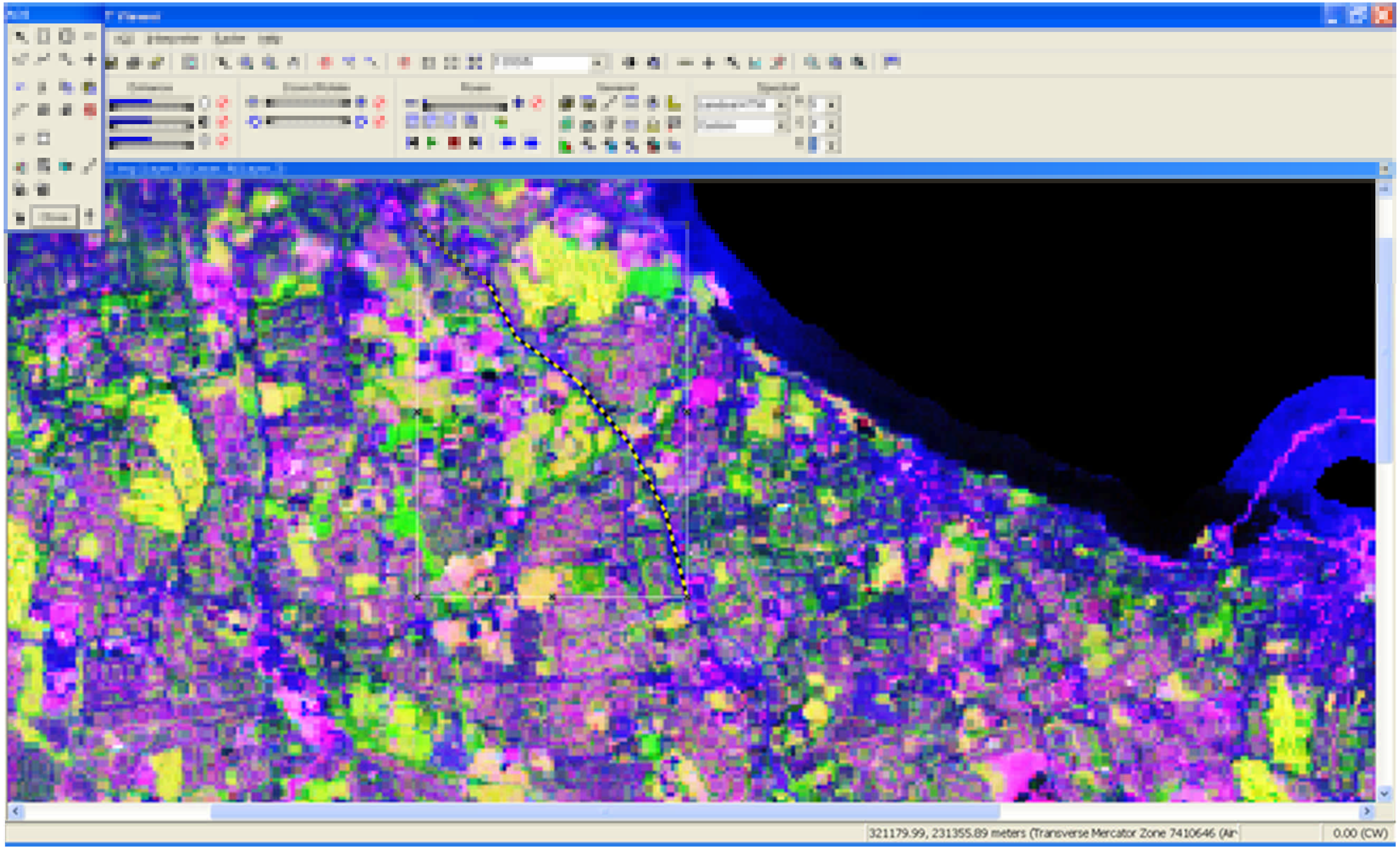


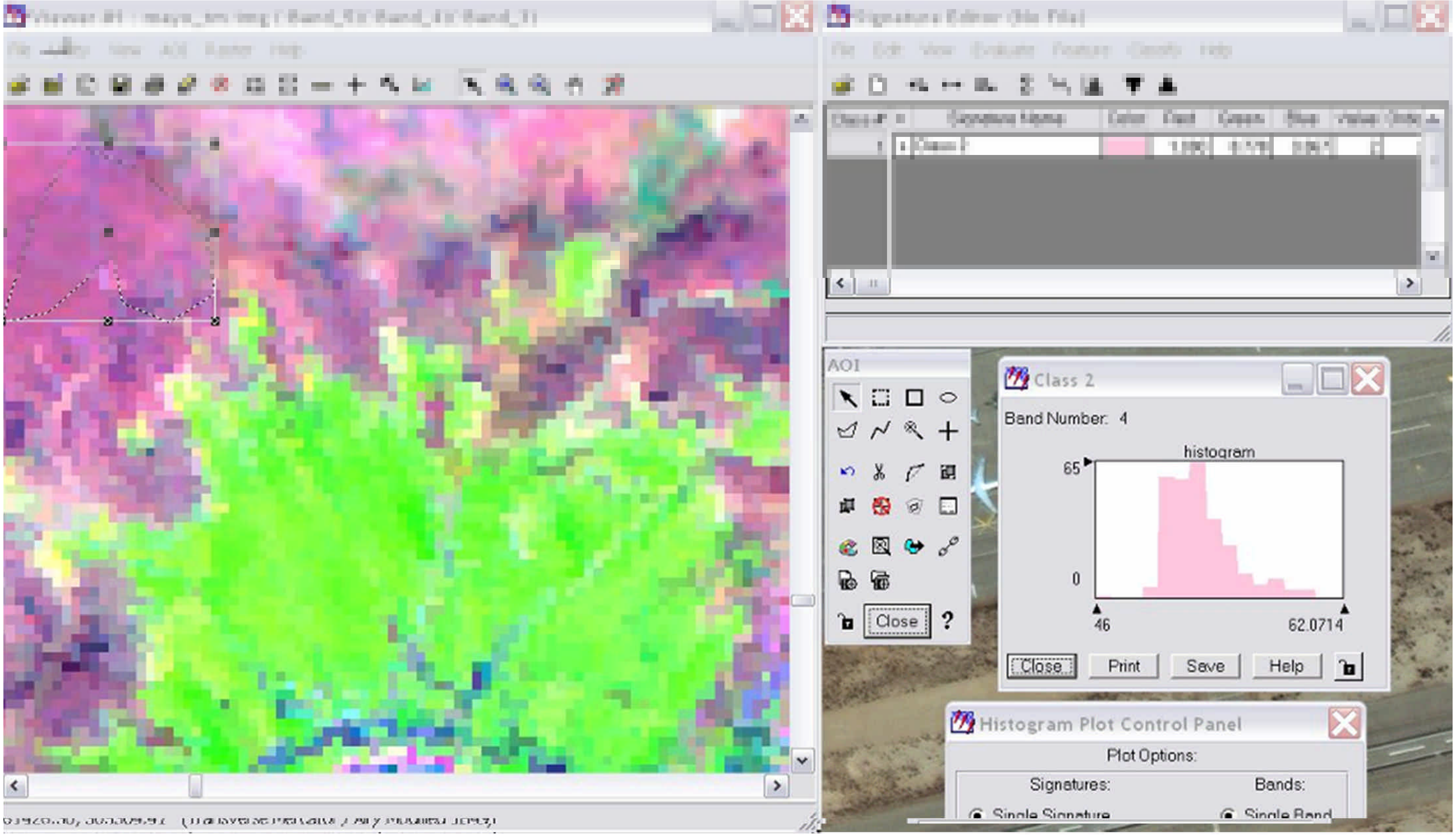

BAD


BAD




317821.94, 223901.05 (Transverse Mercator / Airy Modified 1849)
317821.94, 223901.05 meters (Transverse Mercator Zone 7410646 (Air
0.00 (CW)


321179.99, 231355.89 meters (Transverse Mercator Zone 7410646 (Air
0.00 (CW)



AOI
Close
Class 2
Band Number: 4
histogram
65
0
46
62.0714
Close
Print
Save
Help
Histogram Plot Control Panel
Plot Options:
Signatures:
Bands:

Chose the algotithm that will allocate pixels into your classess

Some Types of Classifiers

- Non-parametric
  - Parallelepiped classifier (BOX or PPD)
  - Minimum-distance-to-means (MDM)
  - Expert Systems

Parametric

- Maximum likelihood (ML)
- Neural networks

A parametric signature is based on statistical parameters (e.g., mean and covariance matrix) of the pixels that are in the training sample or cluster. Supervised and unsupervised training can generate parametric signatures. A set of parametric signatures can be used to train a statistically-based classifier to define the classes.

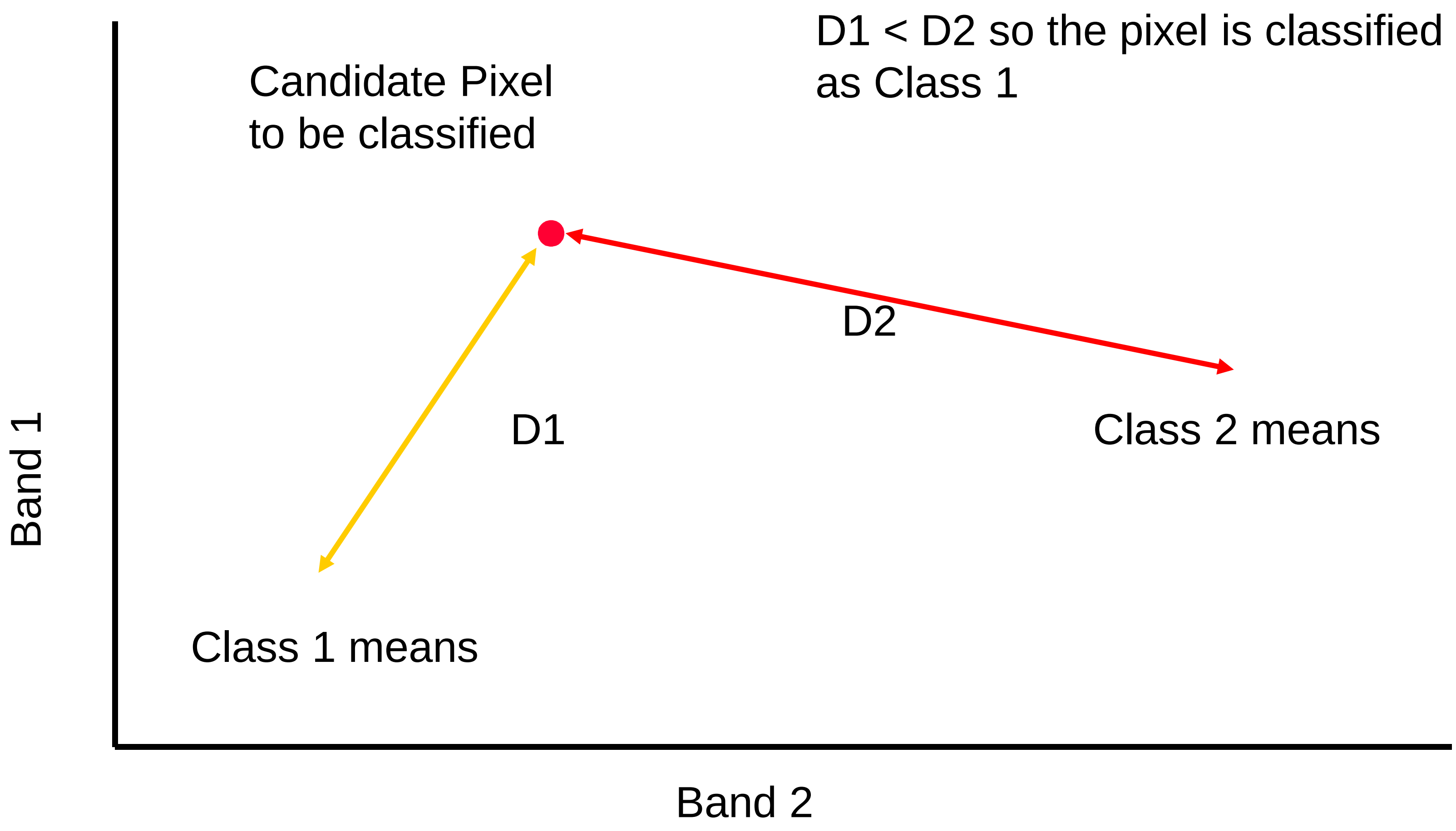
A nonparametric signature is not based on statistics, but on discrete objects (polygons or rectangles) in a feature space image. These feature space objects are used to define the boundaries for the classes.

A nonparametric classifier uses a set of nonparametric signatures to assign pixels to a class based on their location either inside or outside the
area in the feature space image. Supervised training is used to generate nonparametric signatures (Kloer 1994).

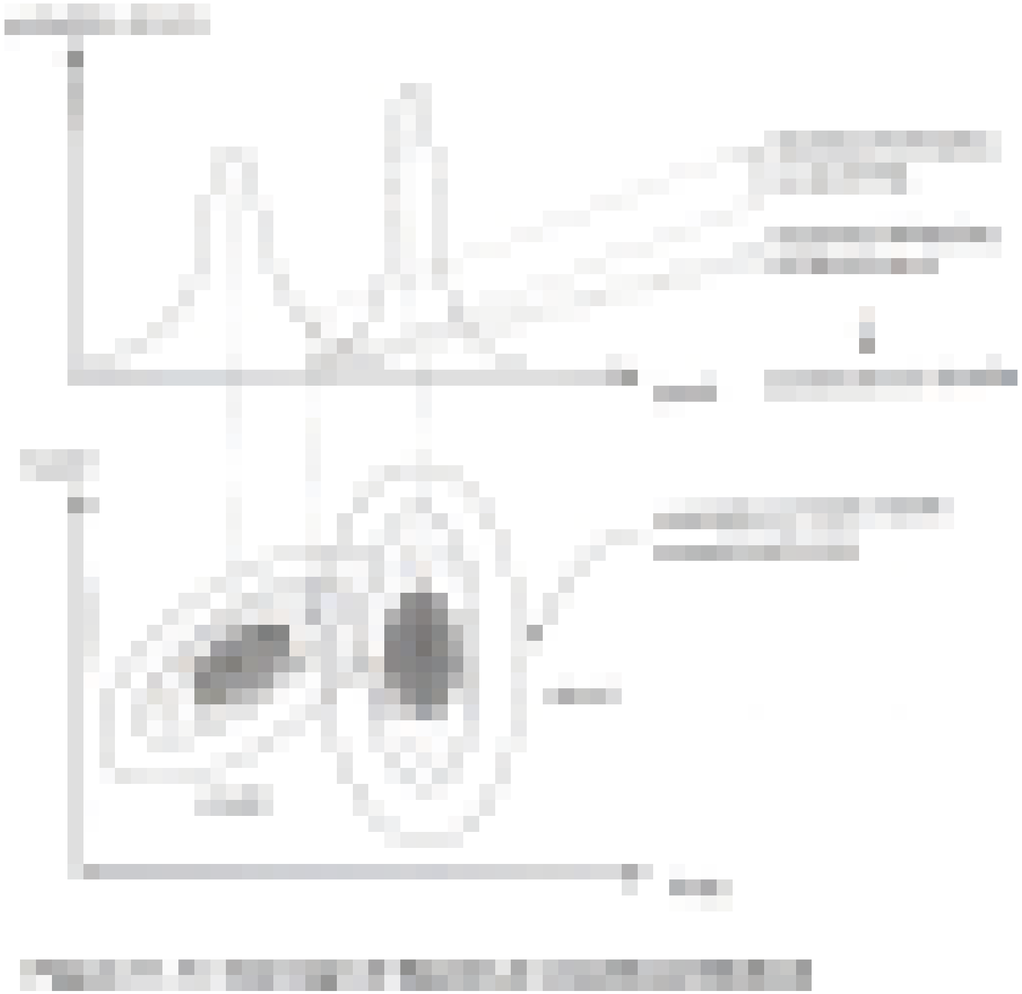
Parallelepiped

- The minimum and maximum DNs for each class are determined and are used as thresholds for classifying the image.
- Benefits: simple to train and use, computationally fast
- Drawbacks: pixels in the gaps between the parallelepipes can not be classified; pixels in the region of overlapping parallelepipes can not be classified.

Minimum-Distance-to-Means



Maximum Likelihood

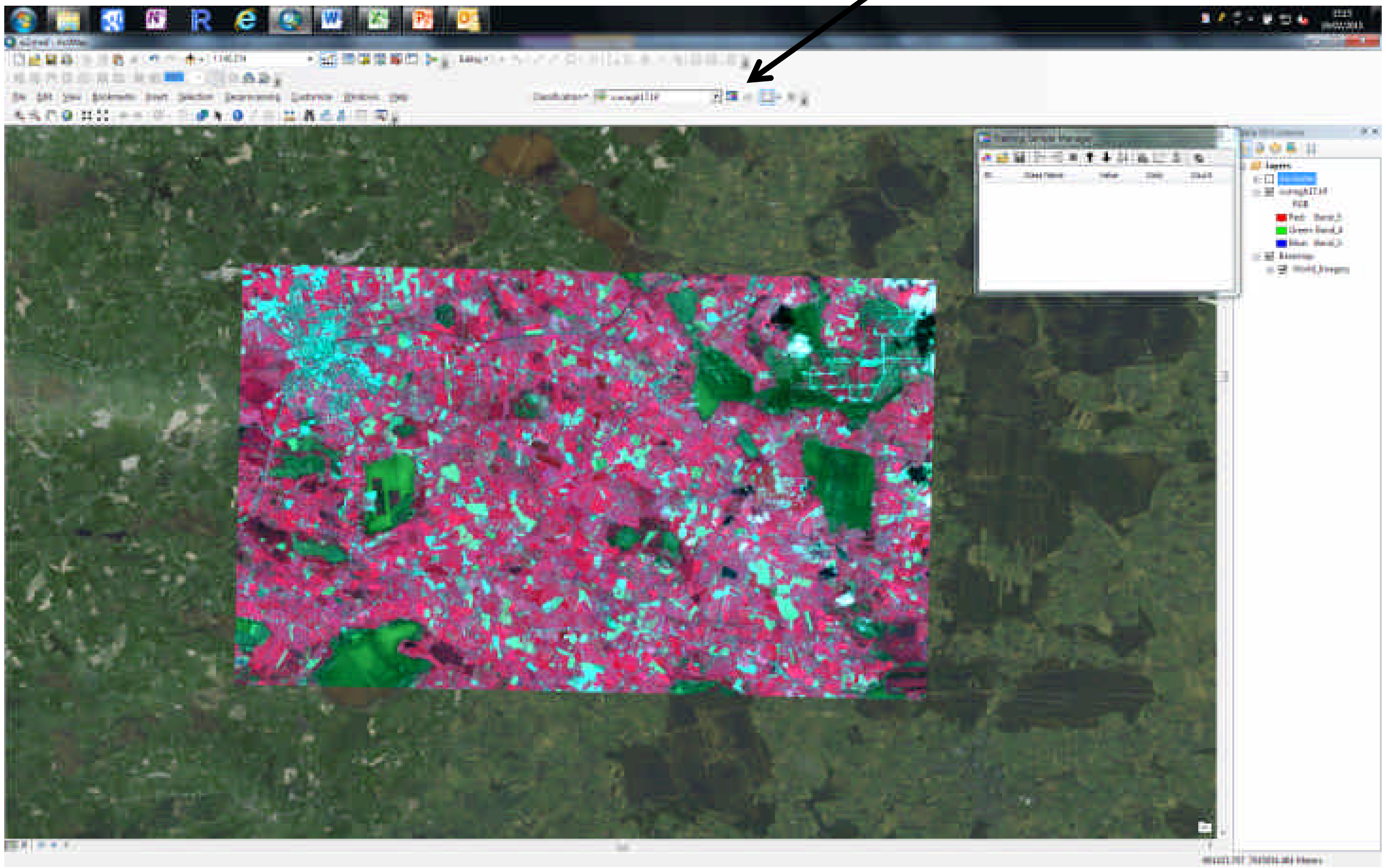


Practical& Assessment

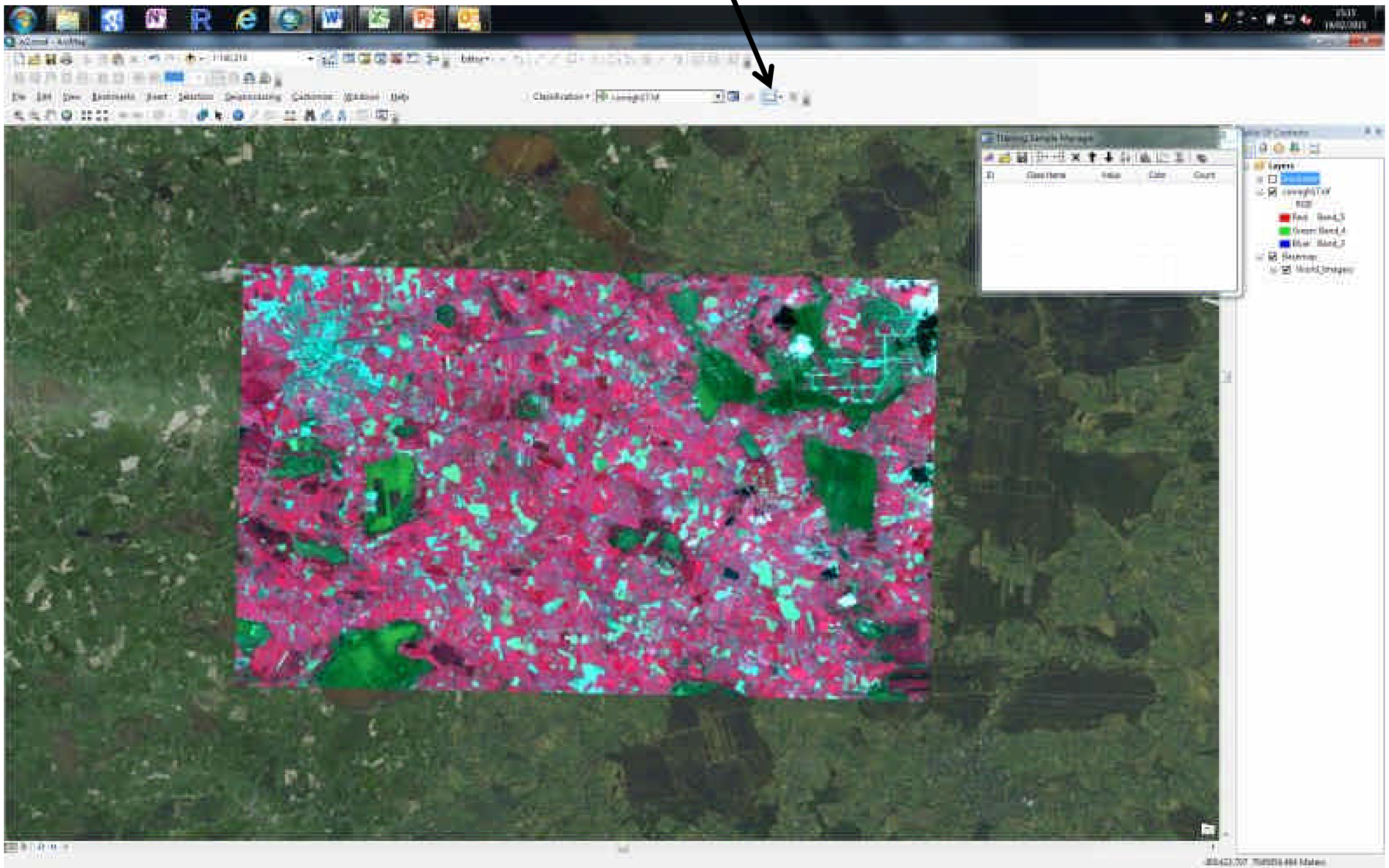
Simple map of Curragh17.tif:

- Grassland
- Urban
- Peatland
- Crop
- Forest

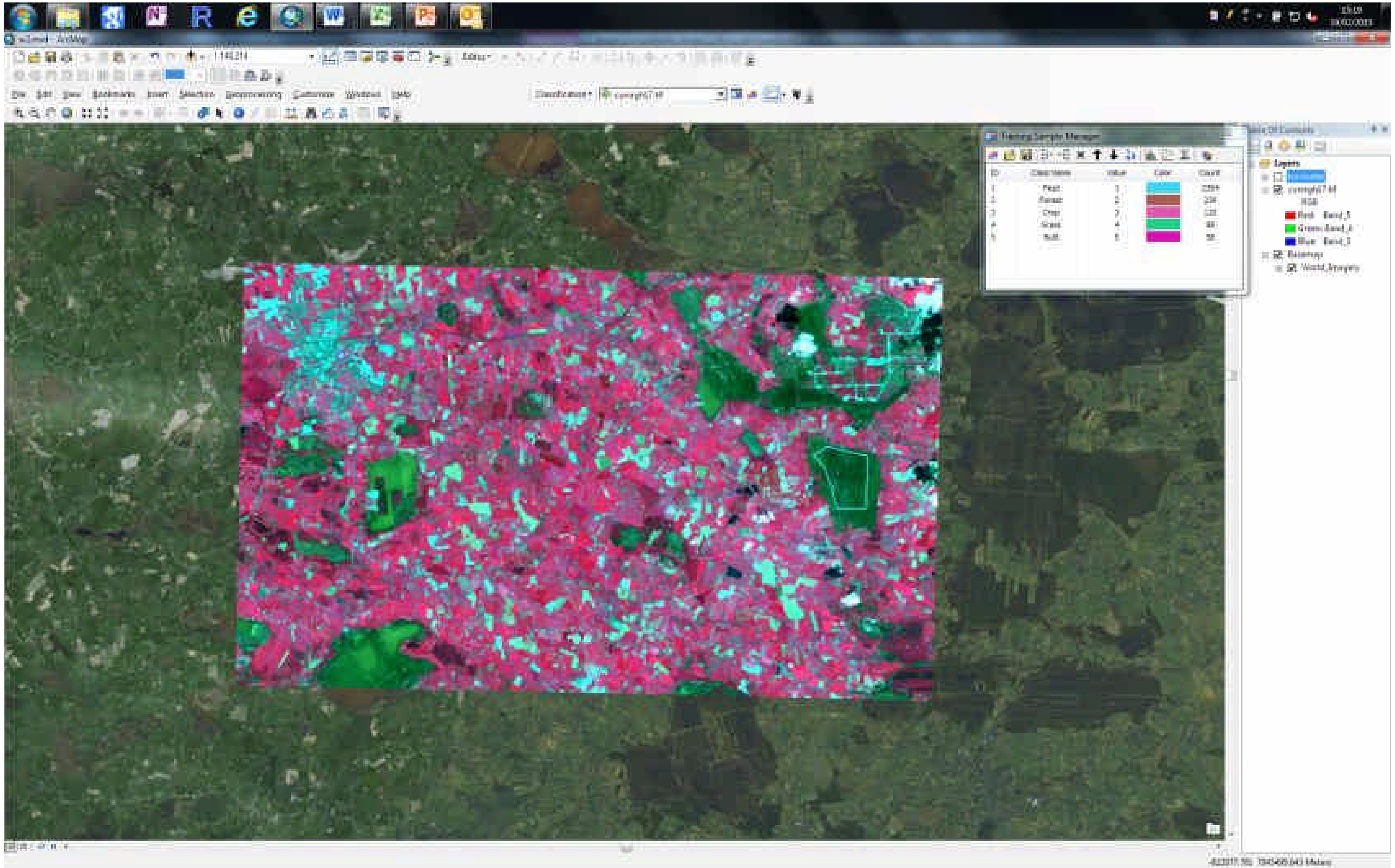
Supervised Classification in ArcMAP
Click the Training Sample Manager Button



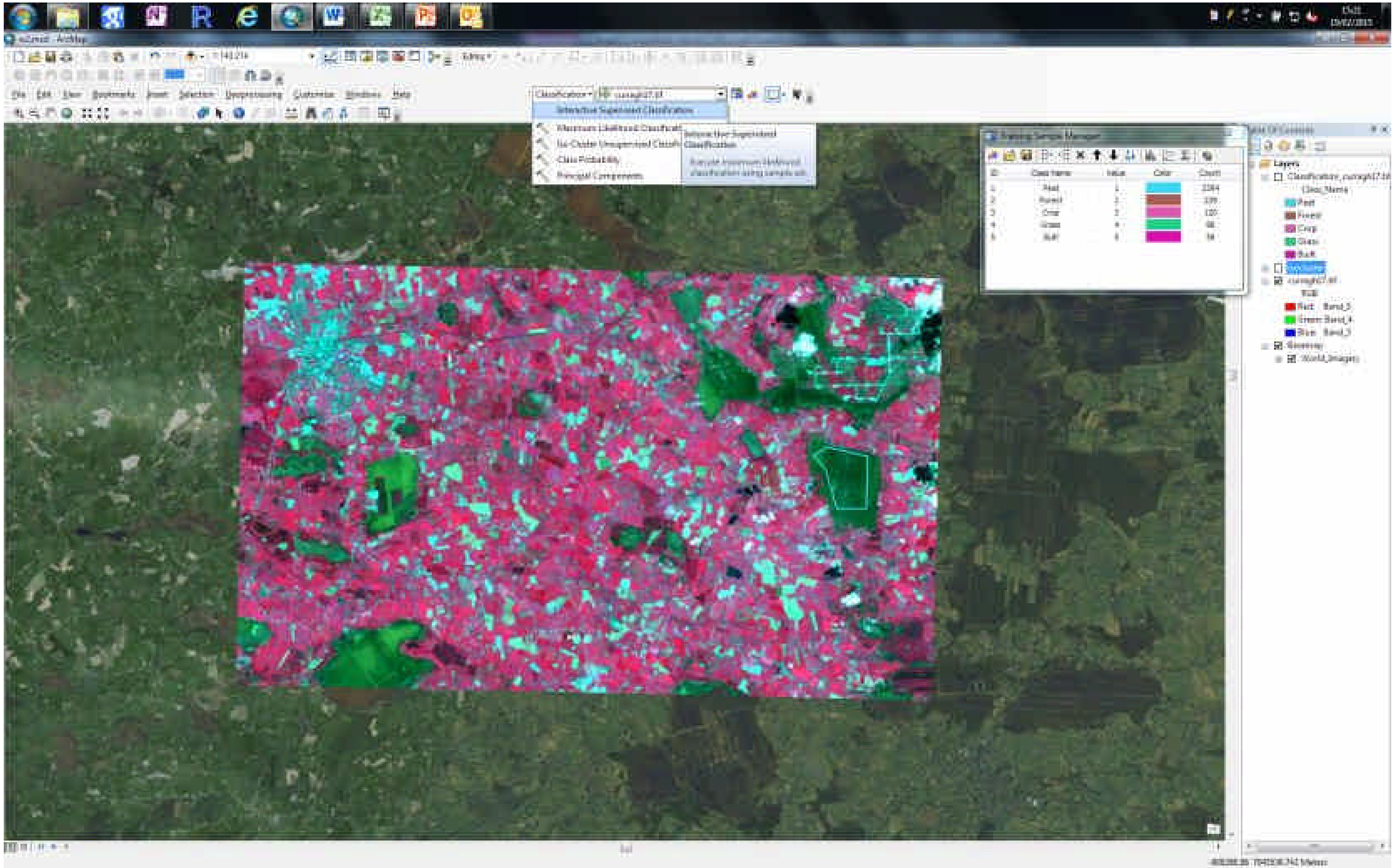
Click the Draw Polygon Icon



Digitise 5 Training areas- one for each theme



Click on Interactive Supervised Classifcation



Assessment

- Complete the map of the 5 themes
- Produce a report on how it was done (Introduction, method, results, assessment)
- To assess: pick 20 random points and compare the classes observed in the basemap air photos with the class your map has produced
- The report should be at least 1500 words
- 50% of the mark is for producing a good map with legend
- 30% for the write up
- 20% for assessment
- This should take no more than 3 hours to produce
- All submissions to me by email (stuart.green@teagasc.ie) in word or pdf.
- **DEADLINE MARCH 30th**

Accuracy Assessment

We may *define* accuracy, in a working sense, as *the degree of correspondence between observation and reality*. We usually judge accuracy against existing maps, large scale aerial photos, or field checks. We can pose two fundamental questions about accuracy:

Is each category in a classification really present at the points specified on a map?

Are the boundaries separating categories valid as located?

Various types of errors diminish the accuracy of feature identification and category distribution. We make most of the errors either in measuring or in sampling. When quantifying accuracy, we must adjust for the lack of equivalence and totality, if possible. Another, often overlooked point about maps as reference standards, concerns their intrinsic or absolute accuracy. Maps require an independent frame of reference to establish their own validity.

As a general rule, the level of accuracy obtainable in a remote sensing classification depends on diverse factors, such as the suitability of training sites, the size, shape, distribution, and frequency of occurrence of individual areas assigned to each class, the sensor performance and resolution, and the methods involved in classifying (visual photointerpreting versus computer-aided statistical classifying), and others

In practice, we may test classification accuracy in four ways:

1) field checks at selected points (usually non-rigorous and subjective), chosen either at random or along a grid;

2) estimate (non-rigorous) the agreement of the theme or class identity between a class map and reference maps, determined usually by overlaying one on the other(s);

3) statistical analysis (rigorous) of numerical data developed in sampling, measuring, and processing data, using tests, such as root mean square, standard error, analysis of variance, correlation coefficients, linear or multiple regression analysis, and Chi-square testing .

4) confusion matrix calculations (rigorous).

With the class identities in the photo as the standard, we arranged the number of pixels correctly assigned to each class and those misassigned to other classes in the confusion matrix , listing errors of commission, omission, and overall accuracies.

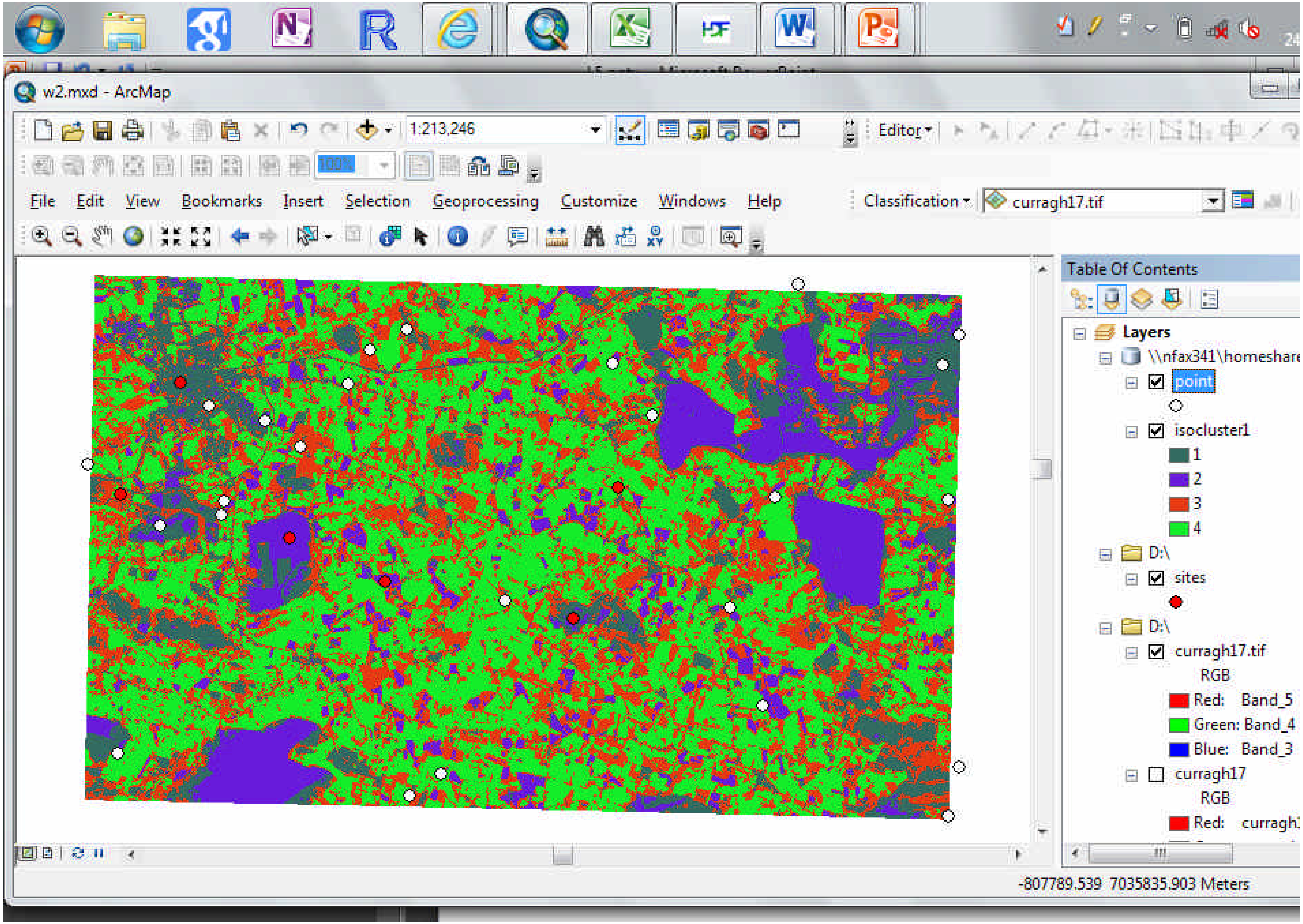
The producer's accuracy relates to the probability that a reference sample (photo-interpreted land cover class in this project) will be correctly mapped and measures the errors of omission (1 - producer's accuracy).

In contrast, the user's accuracy indicates the probability that a sample from land cover map actually matches what it is from the reference data (photo-interpreted land cover class in this project) and measures the error of commission (1- use's accuracy).

Errors of *commission* An error of commission results when a pixel is committed to an incorrect class

.

Errors of *omission* An error of omission results when a pixel is incorrectly classified into another category. The pixel is omitted from its correct class.


w2.mxd - ArcMap
1:213,246
Editor
File
Edit
View
Bookmarks
Insert
Selection
Geoprocessing
Customize
Windows
Help
Classification
curragh17.tif
Table Of Contents
Layers
\\nfax341\homeshare
point
isocluster1
1
2
3
4
D:\
sites
D:\
curragh17.tif
RGB
Red: Band_5
Green: Band_4
Blue: Band_3
curragh17
RGB
Red: curragh1
-807789.539 7035835.903 Meters

	urban	grass	natural	water	forestry	*map*	
urban	**12**	5	1	6	7		31
grass	1	**34**	7		2		44
natural	2	9	**23**		6		40
water				**14**	2		16
forestry				4	**20**		24
ground							
	15	48	31	24	37		**155**

Urban Comission (15-12)/15= 25%
So users accuracy is 75%

Urban Ommission: (31-12)/31= 61%
So Producers accuracy is: 39%

Total mapp accuray is (12+34+23+14+20)/155 = 66%