Remote Sensing and Land Classification

Rory Hutson

Remote Sensing Group

Plymouth Marine Laboratory

Introduction

- Remote sensing background theory
  - Electromagnetic spectrum
  - Radiative transfer basics
- Land classification
  - Supervised and unsupervised classification
  - Ground truth and validation
- Interactive session
  - Experiment with classification techniques in ENVI.

Remote sensing background

- Why is this important?
- Electromagnetic spectrum
- Radiative transfer processes
- Black-body radiation
- Atmospheric window regions
- Spectral patterns / band ratios

Why RT is important?

- Need to know what to expect from RS measurements.
- How different surfaces affect the readings
- How the atmosphere affects readings
- What type of radiation is emitted from different sources e.g. sun, earth, atmosphere
- What different sensors can tell us.

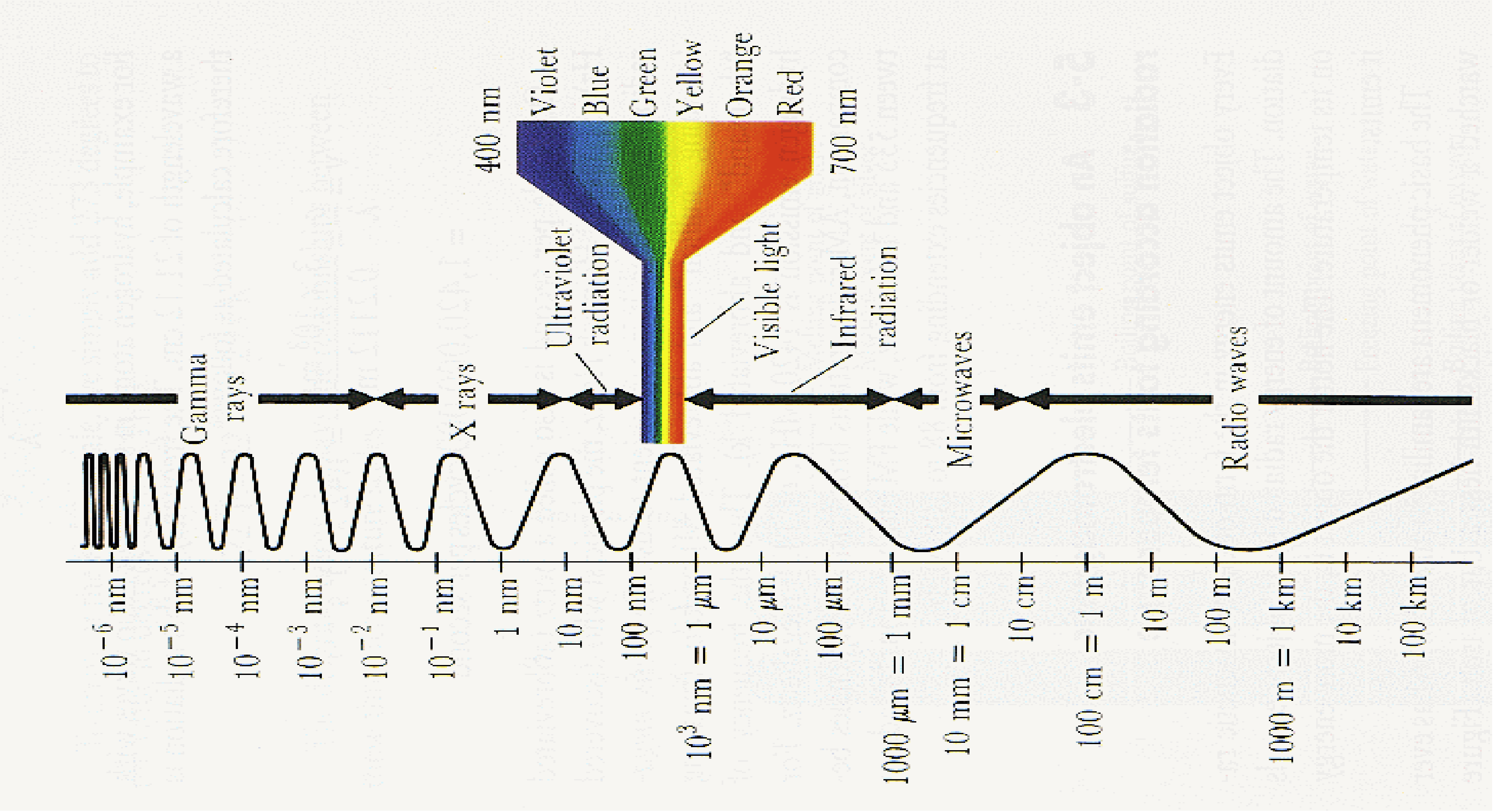
Electromagnetic spectrum

- EM spectrum is split into a few regions.
- Wavelength (λ) increases from left to right
- Frequency (f) decreases from left to right
- Linked by speed of light c:

$$c = \lambda * f$$

- We only see a small portion of the EM spectrum in the wavelength 400-700nm

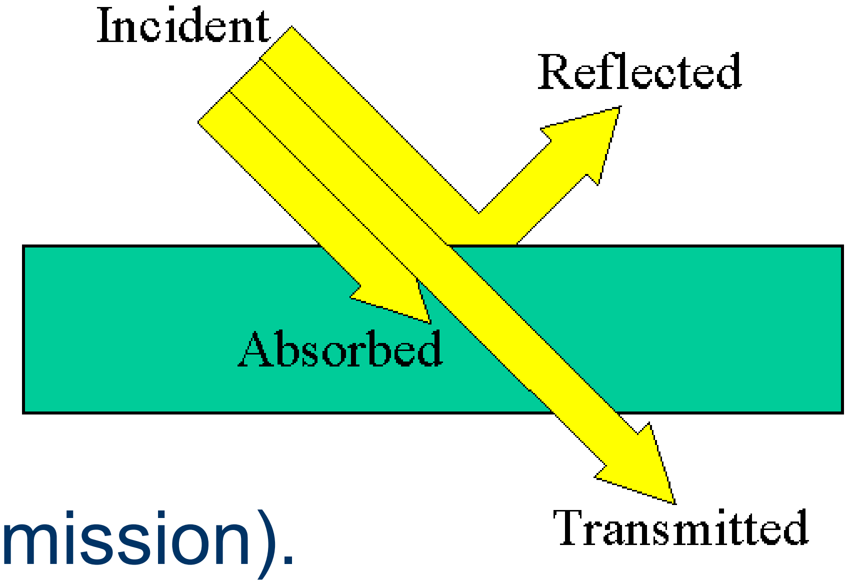
Diagram of EM spectrum



Radiative transfer basics

EM radiation is subject to four main processes;

- Reflection.
- Transmission.
- Scattering.
- Absorption (and re-emission).



Sources of RS radiance.

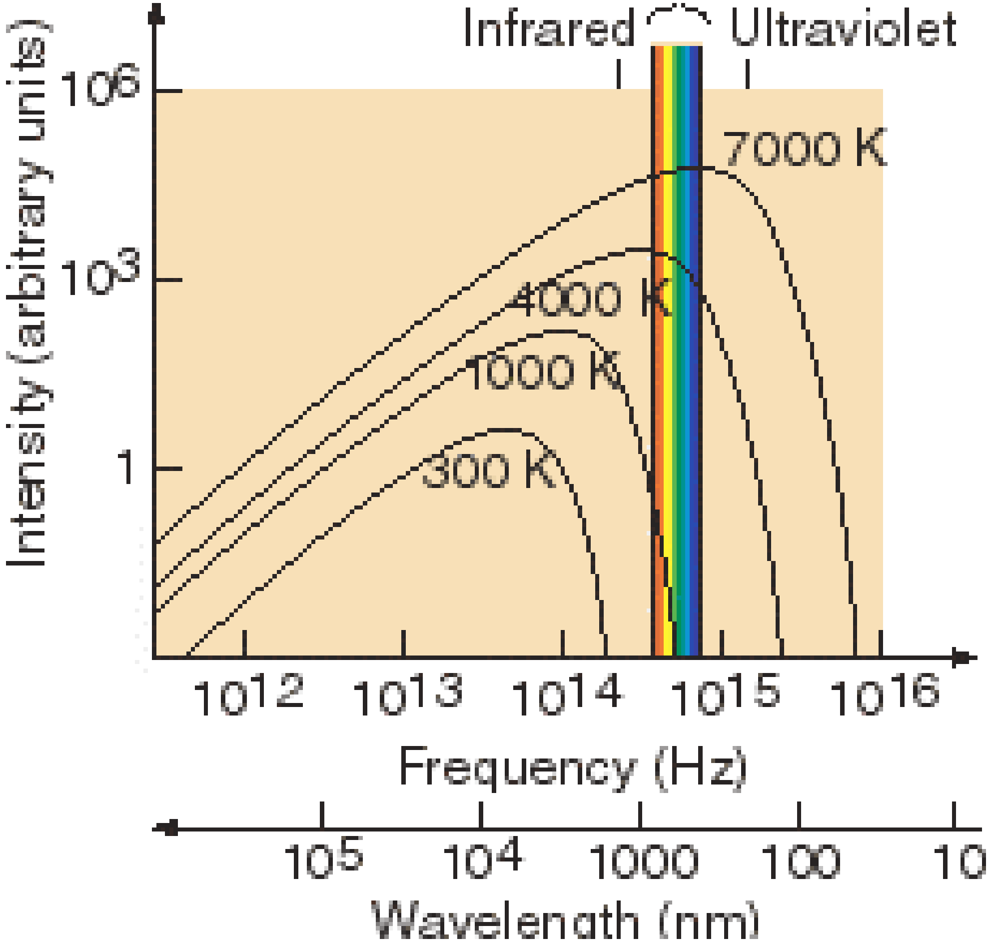
Remotely Sensed radiance values measured at a sensor come from various sources:

- Surface reflected
- Surface emitted
- Cloud top reflected
- Cloud emitted
- Atmospheric back scatter

Black body radiation

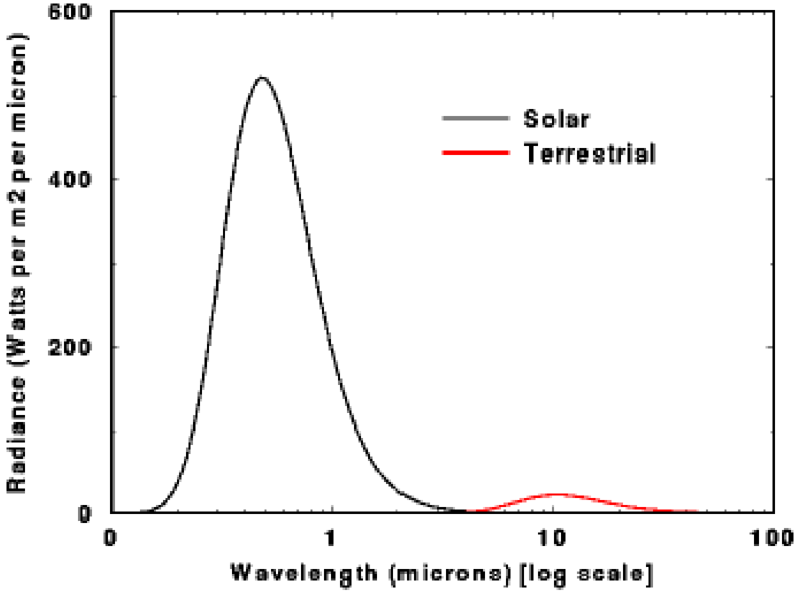
- The amount of radiation from an object will depend on its temperature and emissivity.
- Natural objects emissivity will vary with wavelength.
- Black-bodies are perfect radiators so emit maximum amount radiation, defined by Planck curves.
- As the object heats up the amount of radiation it emits increases.
- The wavelength at the peak of the Planck curve decreases as objects get hotter.

Planck Curves

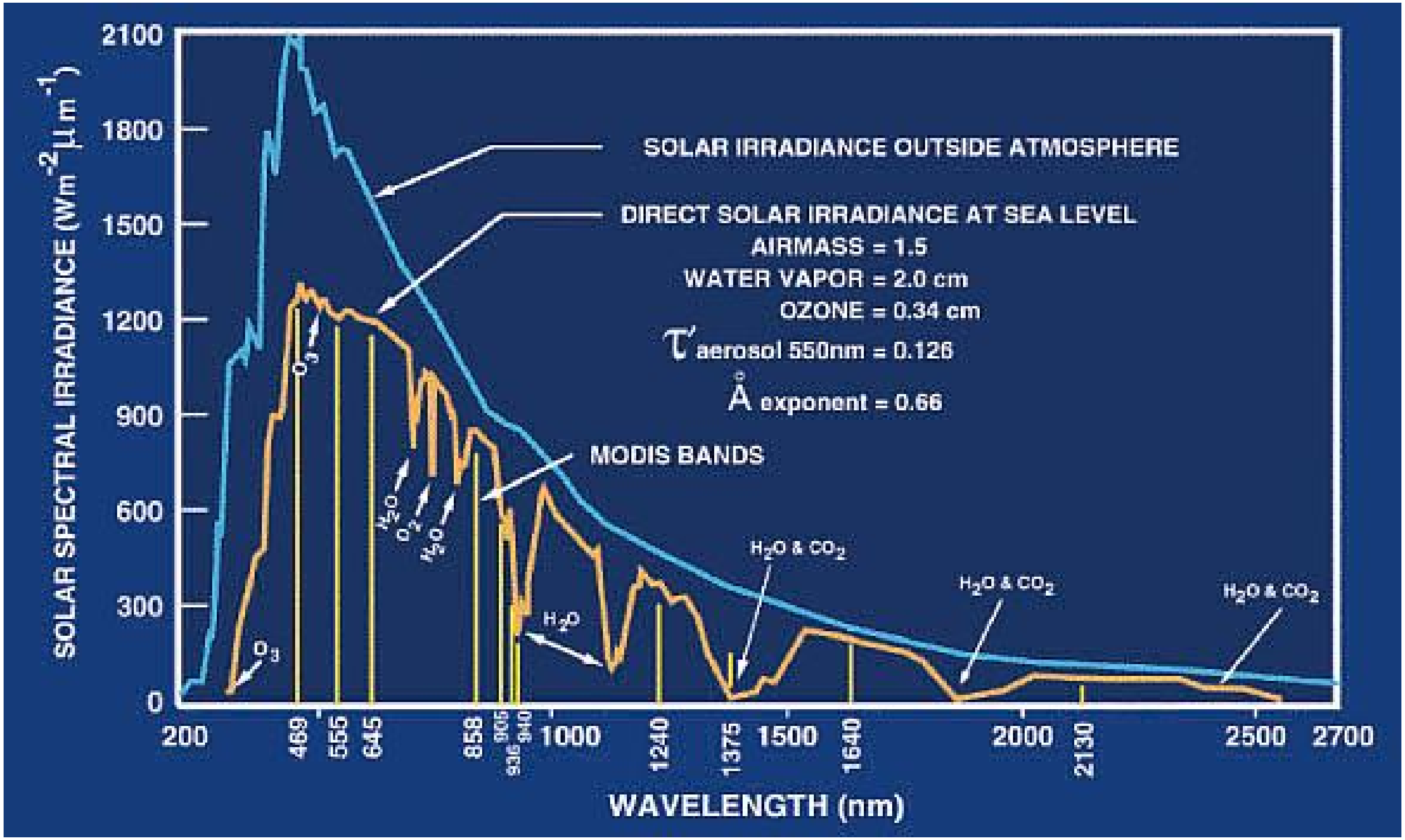


Solar and terrestrial radiation

- The Sun acts like a BB at 6000K
  - This BB spectra peaks in the visible wavelengths.
- The Earth acts like a BB at around 300K.
  - BB spectral peak around 10-12µm (thermal IR)



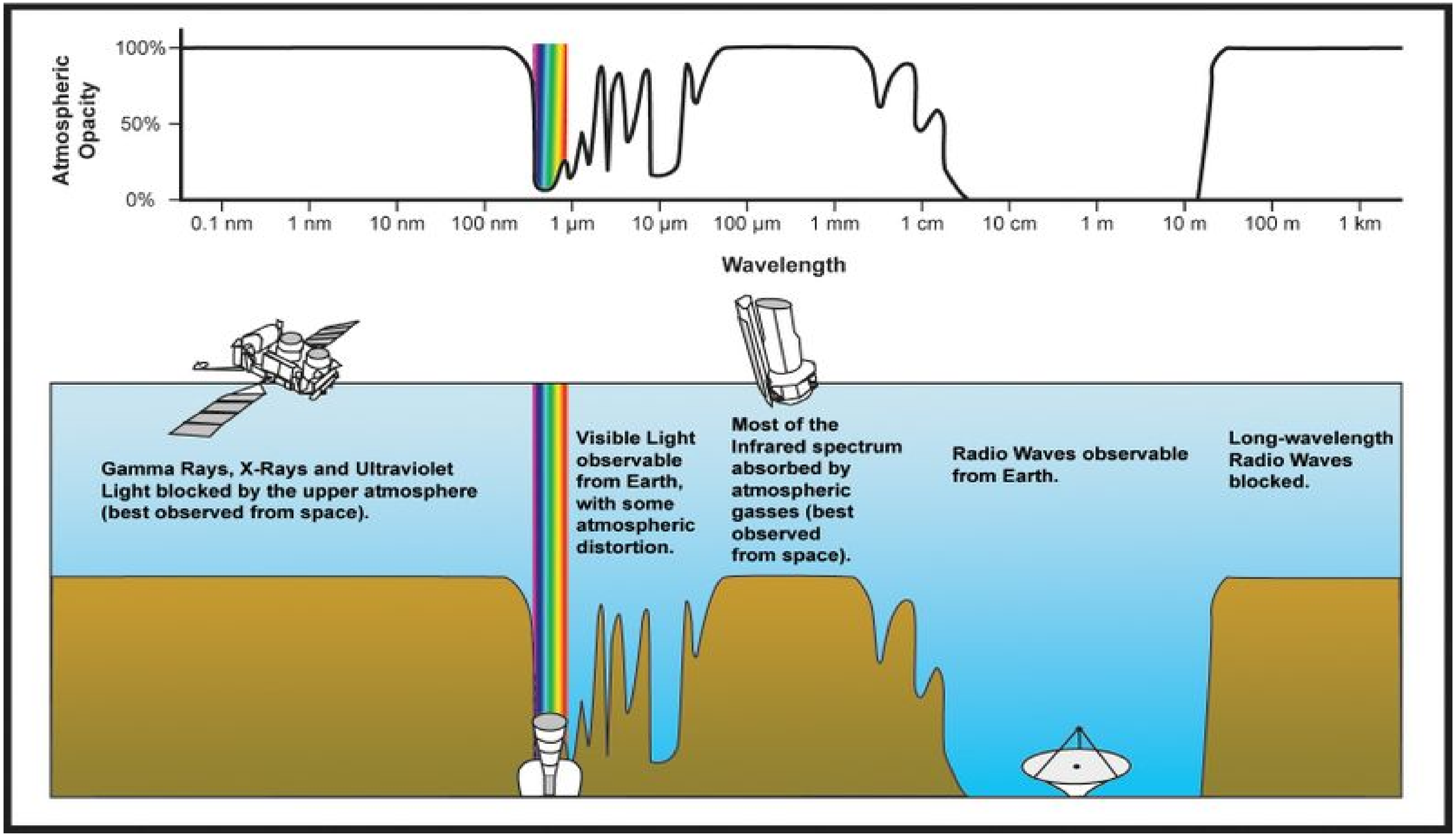
Example of solar spectra



Atmospheric absorption

- Radiance values measured by a sensor will be affected by the atmosphere.
- This happens more at certain wavelengths.
- Main absorbers are gasses like water vapour and carbon dioxide.
- Some wavelength regions are affected less by atmospheric absorption. These are called window regions and can be used by RS sensors.

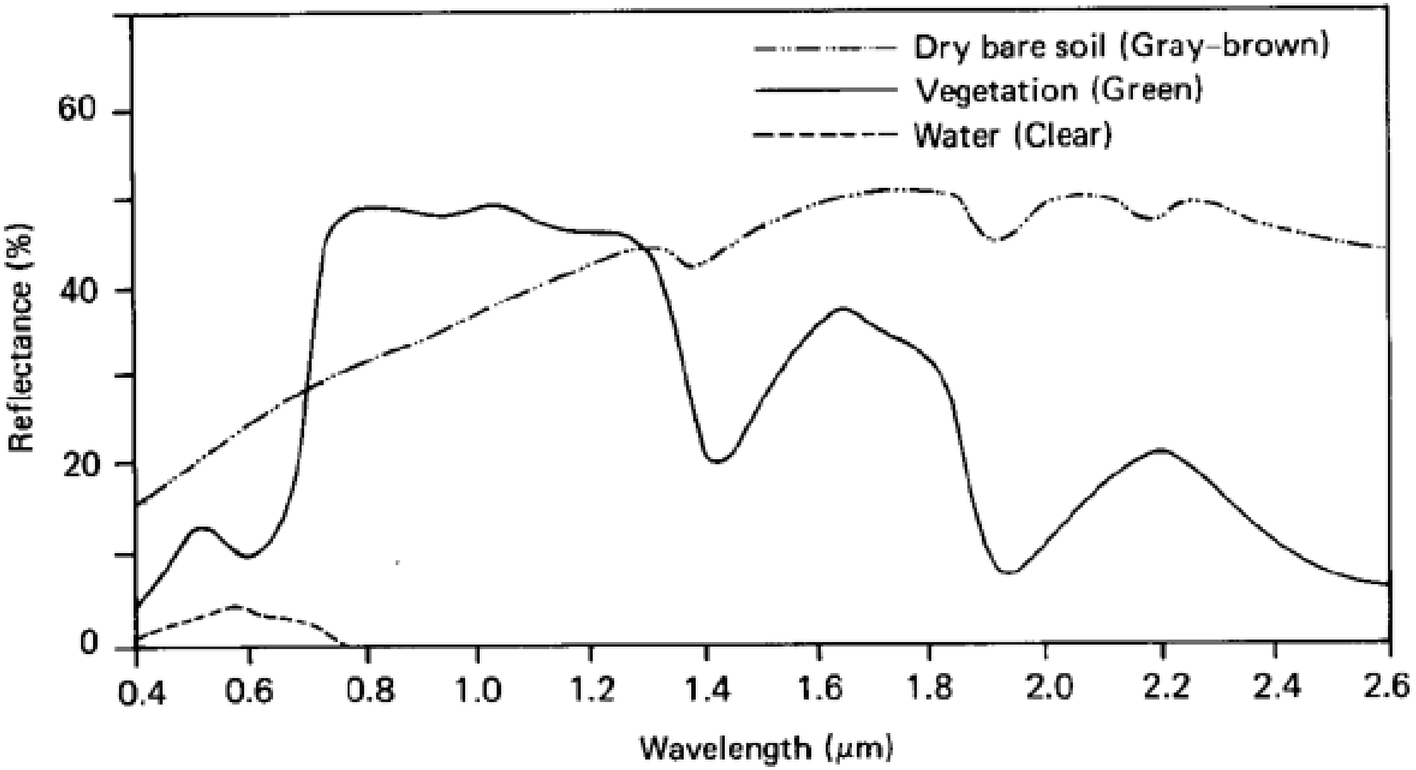
Atmospheric window regions



Spectral signatures (patterns)

- The emissivity of natural objects varies with wavelength.
- At some wavelengths they may act like a BB.
- At others they may have low emissivity
- A spectral pattern can be generated for an object by measuring reflectance or emissivity at different wavelengths.
- These patterns can then be used to distinguish between surface types.
- The spectral pattern for an surface may change i.e. due to solar zenith angle or moisture content.

Example spectral signatures



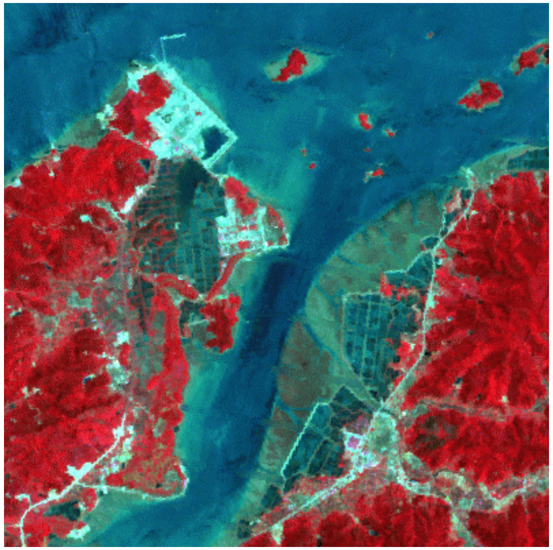
Vegetation indices

- Plants absorb radiation in the blue and red regions of the EM spectrum.
- But in the near-IR they are very reflective.
- This change in reflectance is distinctive and can be used to identify vegetation cover.
- Band ratios using the red and near-IR channels are often used. Using a ratio also helps to remove variations due to solar zenith angle.
- Common one is NDVI = (NIR — VIS)/(NIR + VIS)

Example Landsat data

Near-IR

Red



Combined (432)*

* False colour combining bands 4 (nearIR), 3 (red) and 2 (green).

RS Theory Summary

- Previous section looked aspects of radiative transfer relevant to RS.
- Satellites observe the EM spectrum through atmospheric window regions. so they can measure reflected solar, and emitted terrestrial radiation.
- Given that different surfaces have generally distinct spectral patterns we can try to identify these from RS data.

Land classification

- Aims to label each pixel in a scene to specific land cover types.
- Pixels can then be either correctly classified, incorrectly classified or unclassified.
- Two main type of classification
    - Unsupervised
    - Supervised

Unsupervised classification

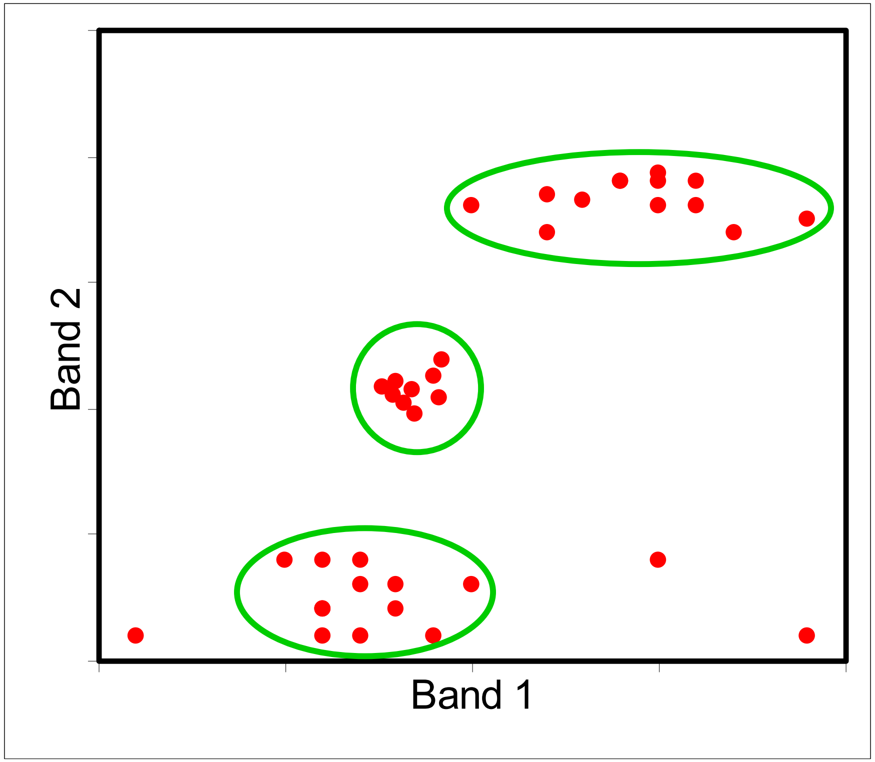
- No previous knowledge assumed about data.
- Tries to spectrally separate the pixels.
- User has controls over:
  - Number of classes
  - Number of iterations
  - Convergence thresholds
- Two main algorithms: Isodata and k-means

Example Landsat bands

Near-IR band

Red band

Example spectral plot

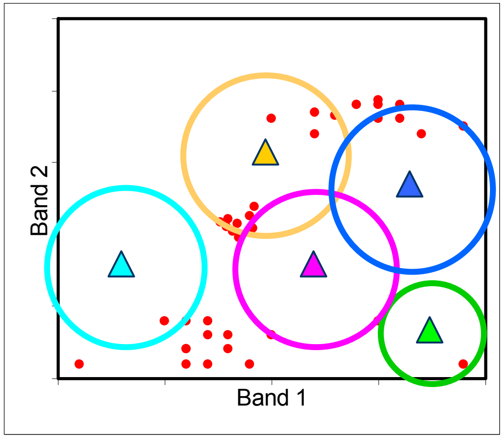


- Two bands of data.
- Each pixel marks a location in this 2d spectral space
- Our eye’s can split the data into clusters.
- Some points do not fit clusters.

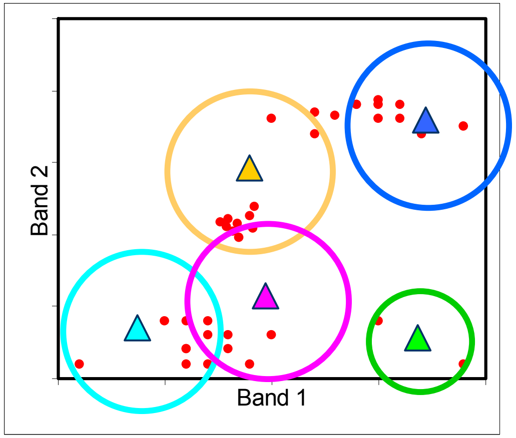
K-means (unsupervised)

1. A set number of cluster centres are positioned randomly through the spectral space.
2. Pixels are assigned to their nearest cluster.
3. The mean location is re-calculated for each cluster.
4. Repeat 2 and 3 until movement of cluster centres is below threshold.
5. Assign class types to spectral clusters.

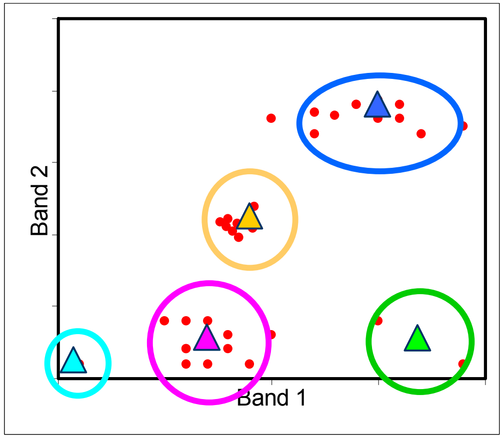
Example k-means



1. First iteration. The cluster centres are set at random. Pixels will be assigned to the nearest centre.



2. Second iteration. The centres move to the mean-centre of all pixels in this cluster.

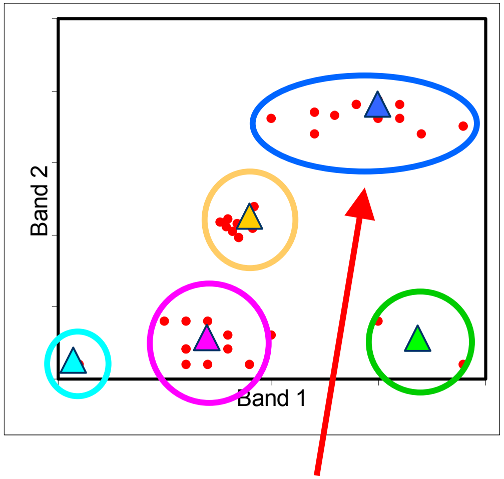


3. N-th iteration. The centres have stabilised.

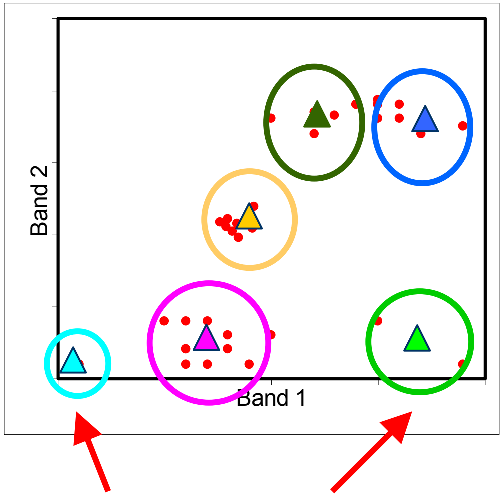
ISODATA (unsupervised)

- Extends k-means. Also calculate standard deviation for clusters.
- After stage 3 we can either:
  - Combine clusters if centres are close.
  - Split clusters with large standard deviation in any dimension.
  - Delete clusters that are to small.
- Then reclassify each pixel and repeat.
- Stop on max iterations or convergence limit.
- Assign class types to spectral clusters.

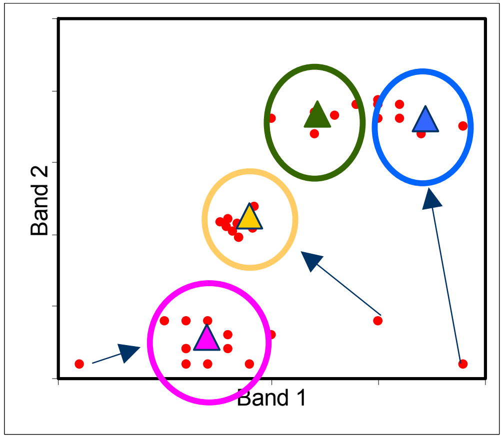
Example ISODATA



1. Data is clustered but blue cluster is very stretched in band 1.



2.Cyan and green clusters only have 2 or less pixels. So they will be removed.



3. Either assign outliers to nearest cluster, or mark as unclassified.

Supervised classification

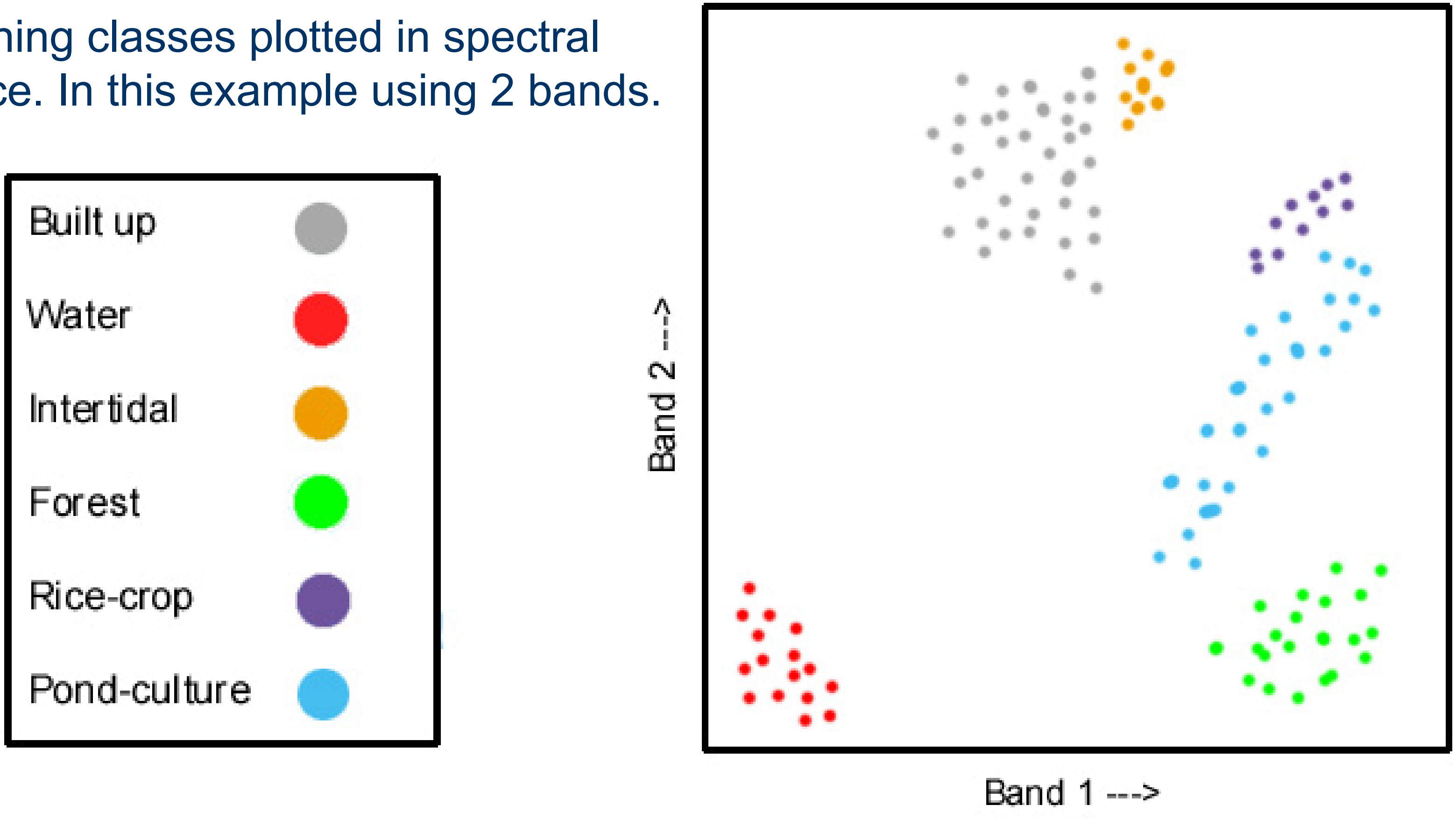
- Start with knowledge of class types.
- Classes are chosen at start
- Training regions are created for each class
- Ground truth used to verify the training regions.
- Quite a few algorithms. Here we will look at:
  - Parallelepiped
  - Maximum likelihood

Parallelepiped (supervised)

- For each training region determine the range of values observed in each band.
- These ranges form a spectral box (or parallelepiped) which is used to classify this class type.
- Assign new image pixels to the parallelepiped which it fits into best.
- Pixels outside all boxes can be unclassified or assigned to the closest one.
- Problems with classes that exhibit high correlation between bands. This creates long ‘diagonal’ data-sets that don’t fit well into a box.

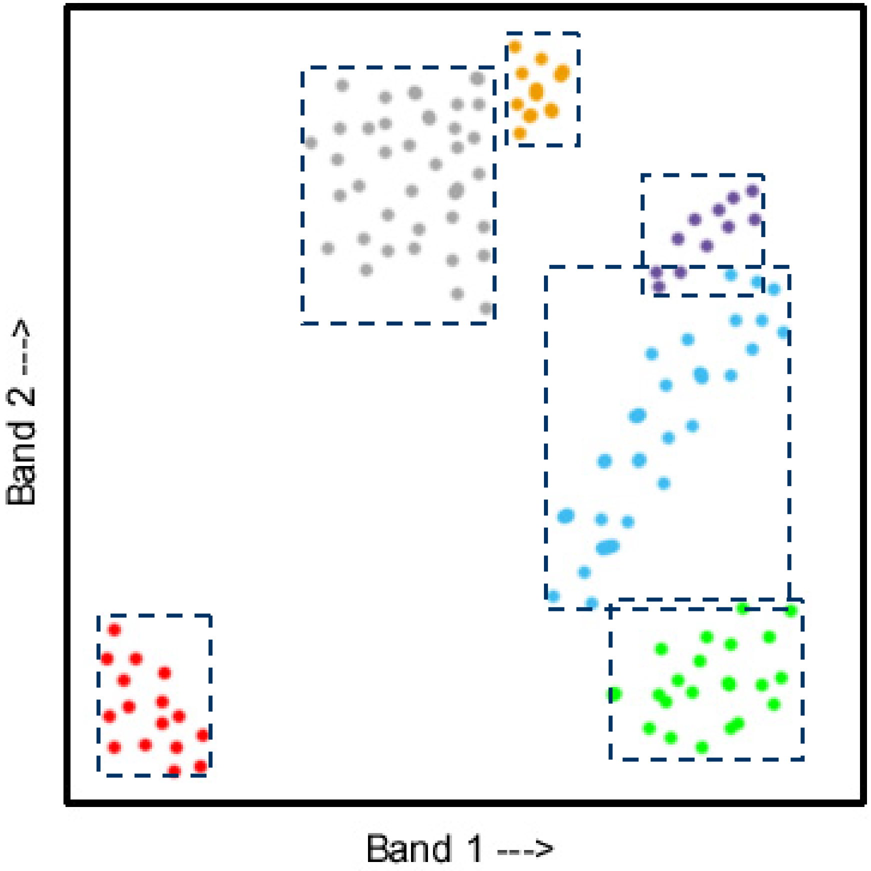
Parallelepiped example

Training classes plotted in spectral space. In this example using 2 bands.



Parallelepiped example continued

- Each class type defines a spectral box
- Note that some boxes overlap even though the classes are spatially separable.
- This is due to band correlation in some classes.
- Can be overcome by customising boxes.

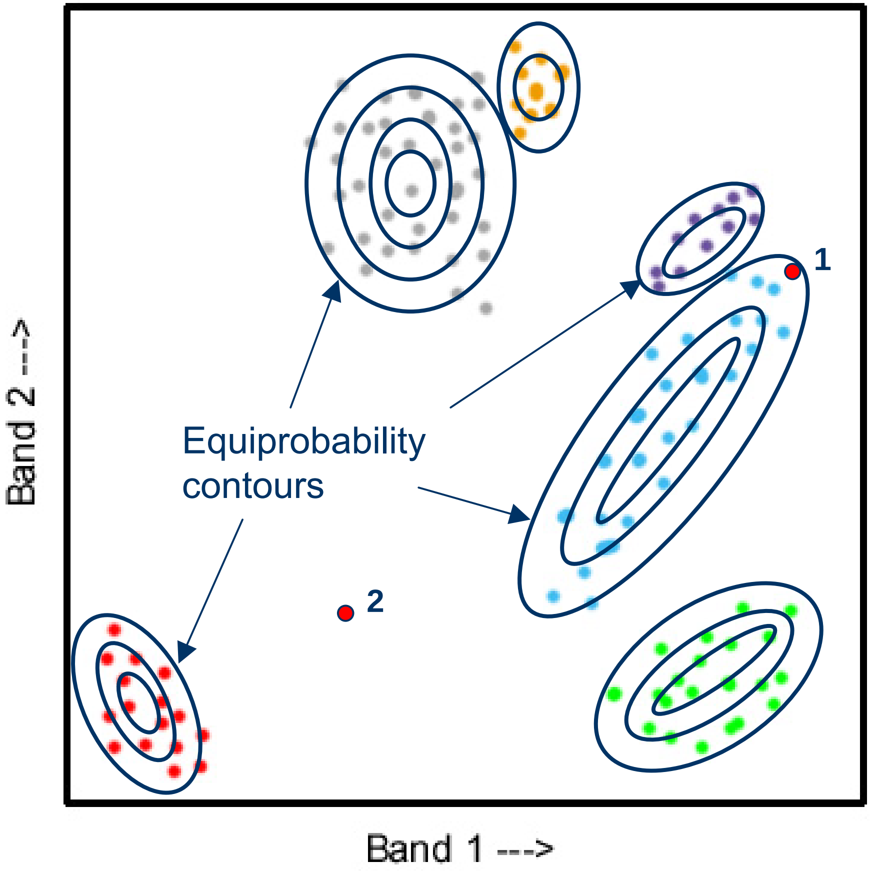


Maximum likelihood (supervised)

- For each training class the spectral variance and covariance is calculated.
- The class can then be statistically modelled with a mean vector and covariance matrix.
- This assumes the class is normally distributed. Which is generally okay for natural surfaces.
- Unidentified pixels can then be given a probability of being in any one class.
- Assign the new pixel to the class with the highest probability – or unclassified if all probabilities low.

Maximum likelihood example

- Normal probability distributions are fitted to each training class.
- The lines in the diagram show regions of equal probability.
- Point 1 would be assigned to class ‘pond culture’ as this is most probable.
- Point 2 would generally be unclassified as the probabilities of fitting into one for the classes would be below threshold.



Ground truth

- Ideally the training regions need to be based on ground observation.
- They should be large enough to capture all the spectral variability in the class type.
  - E.g. different types of forest, shallow water and deep ocean etc.
- Do not need to get too detailed otherwise classes will not be spectrally separable.

Post classification

- Can check non-training regions with more ground truth if available.
- Calculate classification statistics.
  - **Confusion Matrix**: Columns show ground truth, rows show how many pixels are assigned to each class.
  - **Overall accuracy**: Total correct pixels/total pixels
  - **Commission errors**: Incorrect pixels assigned to a class
  - **Omission errors**: Pixels in class that are assigned a different class
- Visually check to see if any major errors or unwanted features.

Classification Summary

- Covered the basic issues of land classification.
- Looked at different types of supervised and unsupervised classification.
- Details of ground truth to set up ROIs.

Interactive session - Overview

- Look at dataset in ENVI.
- Experiment with unsupervised classification.
- Setup some training ROIs
- Experiment with supervised classification
- Examine results and see what was most successful.

Example dataset

- Sub-set of Landsat ETM+ scene of Huangdun Bay from 2005.
- Only has 5 bands here:
  - Red, green, blue, near-IR, thermal-IR
- Look at individual bands.
- Try combining e.g. RGB, near-IR R G etc.
- Get a feel for what is picked out in different bands.

Unsupervised classification

- How many clusters should we aim for?
- What happens if we have too few or too many?
- Try running with different settings.
- What class types would you assign to the spectral classes?
- What settings give best results.

Supervised classification

- Need to set up training regions.
  - What classes do we want?
  - Will these represent the data?
- Define ROIs in ENVI.
  - Does each classes ROI fully represent the data?
  - Do we need more than one ROI per class?
- Run the classifier.
  - Start with maximum likelihood and try others later.

Supervised classification results

- How well has it performed:
  - Do the classified zones look sensible?
  - Do we need more or less classes?
- Generate performance statistics. This will show how well the classifier has worked on the training sets.
- To test the whole image would really need ground truth from different regions.