REMOTE SENSING

MODULE OF

REMOTE SENSING DATA ANALYSIS

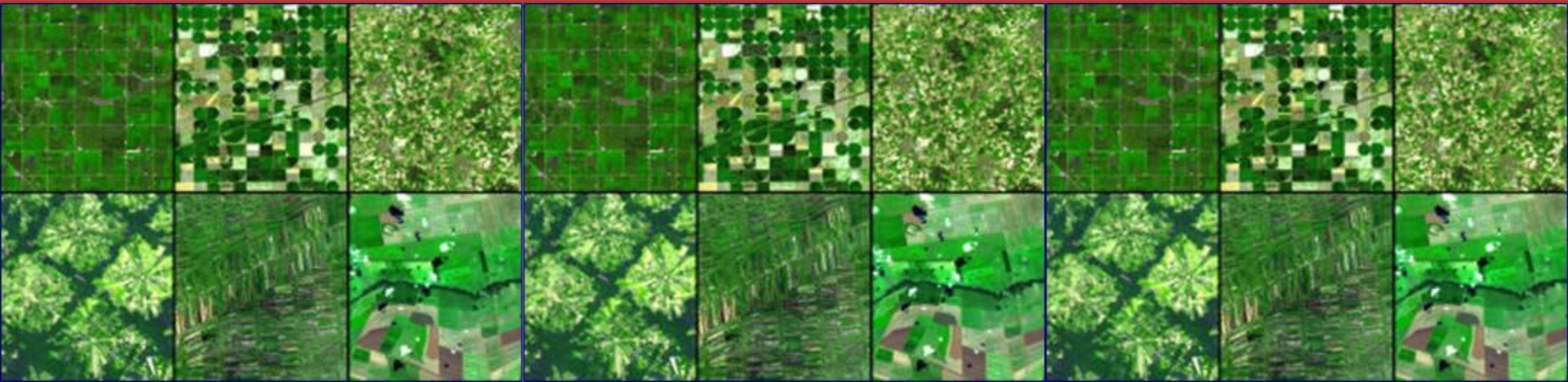
(6 CFU)

A.Y. 2013/14

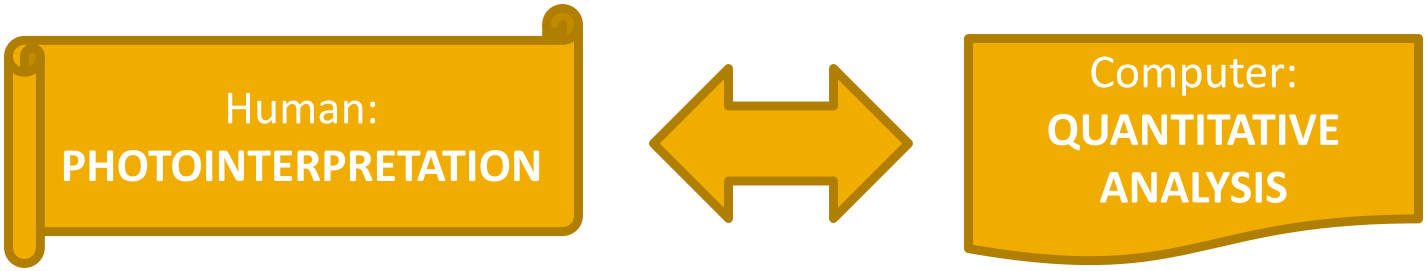
MASTER OF SCIENCE IN COMMUNICATION TECHNOLOGIES AND MULTIMEDIA

PROF. ALBERTO SIGNORONI

THE INTERPRETATION OF DIGITAL IMAGE DATA



Approaches to Interpretation



- **Photointerpretation**, involving direct human interaction and therefore high level decisions, is good for spatial (shape, size, orientation and texture) assessment but poor in quantitative accuracy
  - Roads, coastlines and river systems, fracture patterns, and lineaments generally, are usually readily identified by their spatial disposition.
  - Temporal data, such as the change in a particular object or cover type in an image from one date to another can often be used by the photointerpreter as, for example, in discriminating deciduous or ephemeral vegetation from perennial types.
  - Spectral clues are utilised in photointerpretation based upon the analyst’s foreknowledge of, and experience with, the spectral reflectance characteristics of typical ground cover types, and how those characteristics are sampled by the sensor on the satellite or aircraft used to acquire the image data.
- **Quantitative analysis**, requiring little human interaction, has poor spatial ability but high quantitative accuracy.
  - Its high accuracy comes from the ability of a computer, if required, to process every pixel in a given image and to take account of the full range of spectral, spatial and radiometric detail present.
  - Its poor spatial properties come from the relative difficulty with which decisions about shape, size, orientation and texture can be made using standard sequential computing techniques.

Approaches to Interpretation

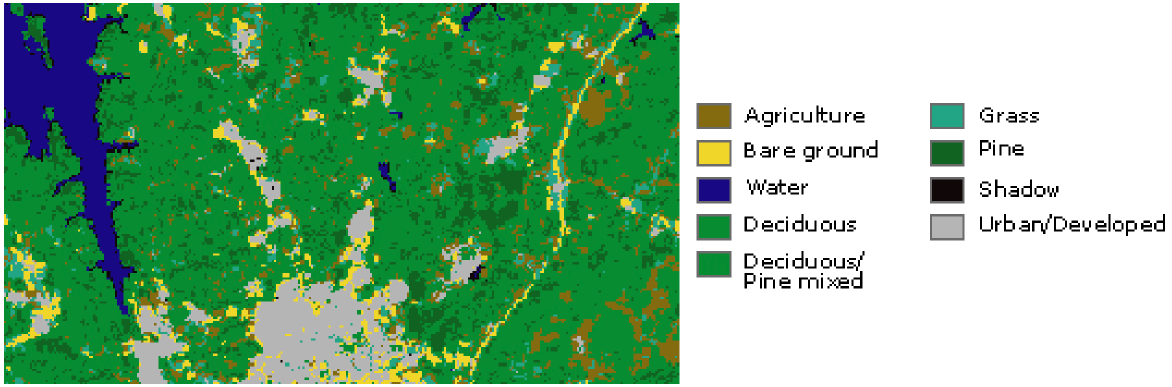
- Those two approaches to image interpretation have their own roles and often these are **complementary**:
  - photointerpretation is aided substantially if a degree of digital image processing is applied to the image data beforehand, while
  - quantitative analysis depends for its success on information provided at key stages by an analyst.This information very often is drawn from photointerpretation.

Photointerpretation (by a human analyst/interpreter)	Quantitative analysis (by computer)
On a scale large relative to pixel size	At individual pixel level
Inaccurate area estimates	Accurate area estimates possible
Only limited multispectral analysis	Can perform true multispectral (multidimensional) analysis
Can assimilate only a limited number of distinct brightness levels (say 16 levels in each feature)	Can make use quantitatively of all available brightness levels in all features (e.g. 256, 1024, 4096)
Shape determination is easy	Shape determination involves complex software decisions
Spatial information is easy to use in a qualitative sense	Limited techniques available for making use of spatial data

Approaches to Interpretation

- In computer-based quantitative analysis the attributes of each pixel (such as the spectral bands available) are examined in order to give the pixel a **label** identifying it as belonging to a particular *class of pixels of interest* to the user.

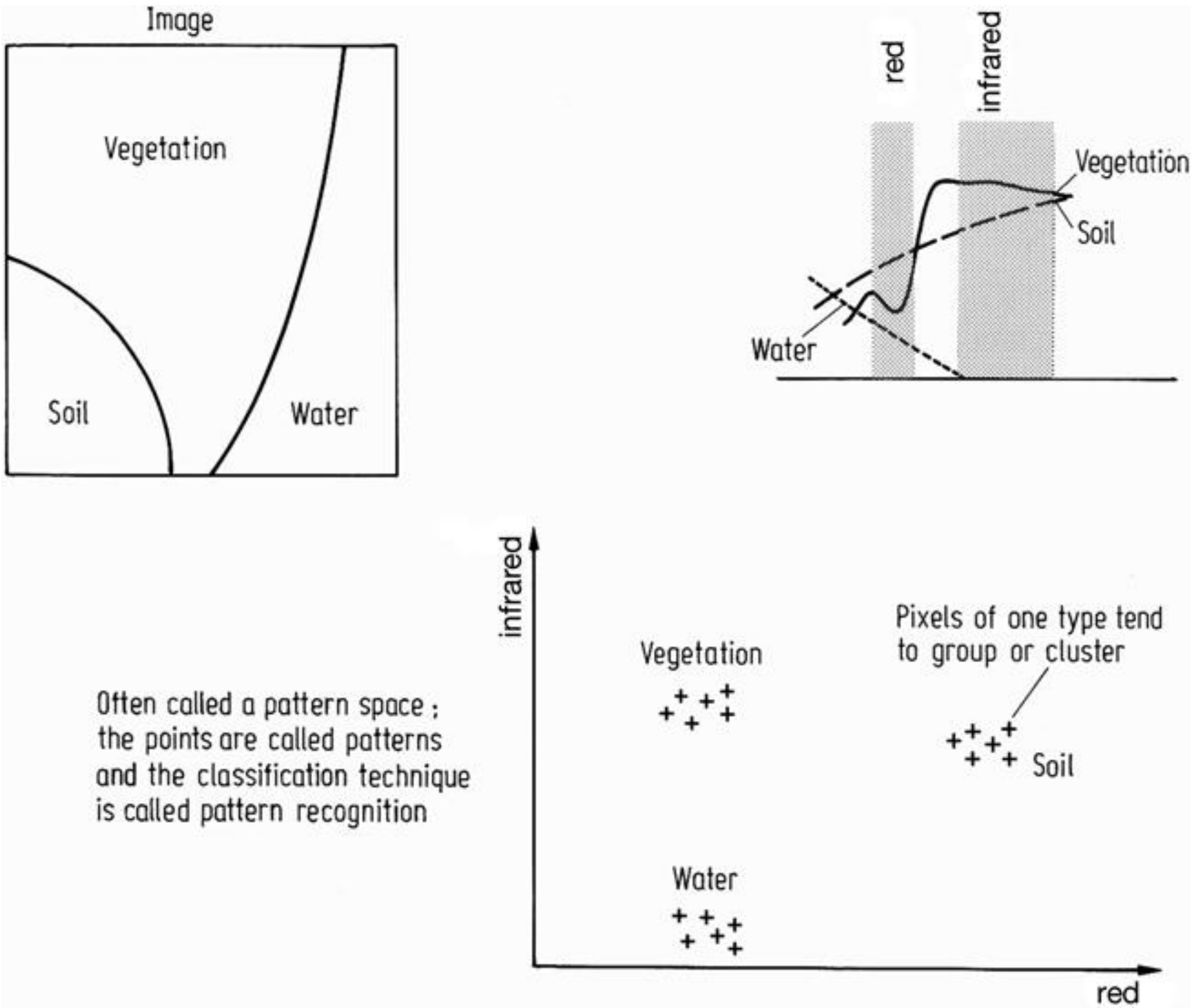
- As a result, the process is often also called ***classification***.



- The human analyst is unable to apply photointerpretation at pixel level (unless only small groups of pixels are of interest) and to discriminate to the limit of the radiometric resolution generally available (only three or so of the complete set of spectral components of an image can be used readily, by color composite images).
- When a computer can be used for analysis (data are in digital format), it could conceivably do so at the pixel level and could examine and identify as many pixels as required. In addition, it should be possible for computer analysis of remote sensing image data to take full account of the multidimensional aspect of the data including its full radiometric resolution.

Multispectral Space and Spectral Classes

- The most effective means by which multispectral data can be represented in order to formulate algorithms for quantitative analysis is to plot them in a **pattern space**, or multispectral vector space, with as many dimensions as there are spectral components.

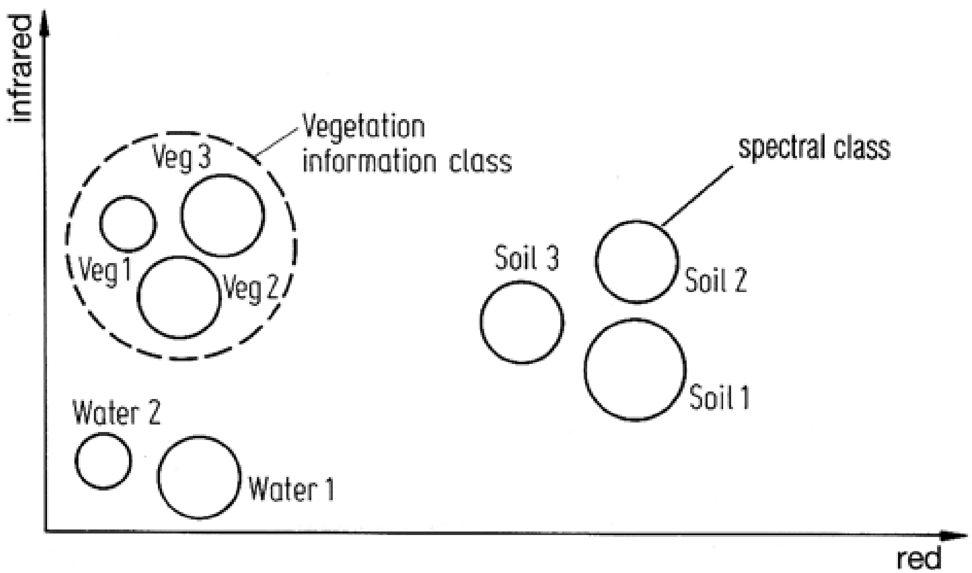


Spectral reflectance characteristics of different ground cover types

In pattern space, each pixel of an image plots as a point with co-ordinates given by the brightness value of the pixel in each component.

Multispectral Space and Spectral Classes

- Provided the spectral bands have been designed to provide good *discrimination* it is expected that pixels would form *groups* in multispectral space *corresponding to various ground cover types*, the sizes and shapes of the groups being dependent upon varieties of cover type, systematic noise and topographic effects.
  - The groups or clusters of pixel points are referred to as **information classes** since they are the actual classes of data which a computer will need to be able to recognize.
- In practice *the information class groupings may not be single clusters* as depicted in the previous Figure. Instead it is not unusual to find *several clusters for the same region of soil*, for the same apparent type of vegetation and so on for other cover types in a scene.
  - These are not only as a result of specific differences in types of cover but also result from differences in moisture content, soil types underlying vegetation and topographic influences.
  - Consequently, a multispectral space is more likely to appear as shown in Figure in which each information class is seen to be composed of several so called **spectral classes**.

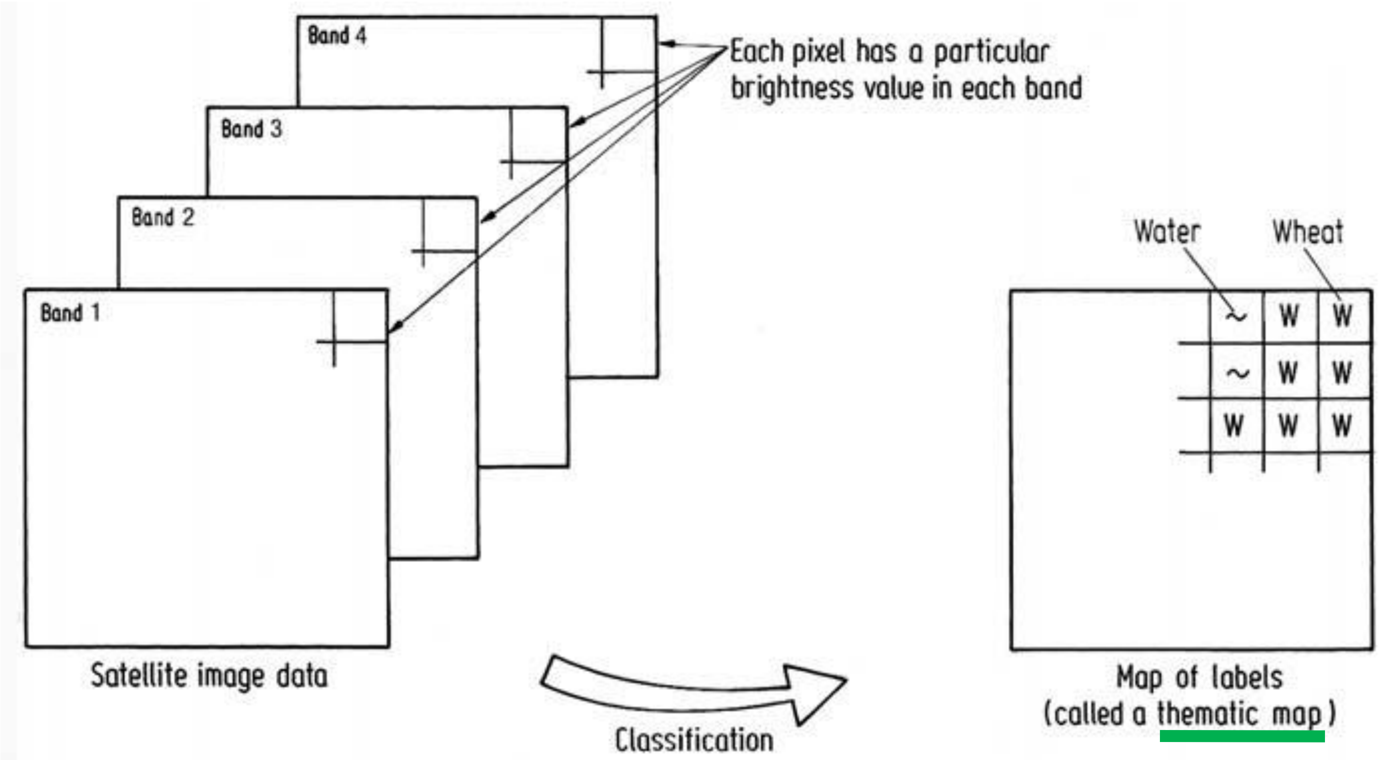


Multispectral Space and Spectral Classes

- In many cases the information classes of interest do not form distinct clusters or groups of clusters but rather are part of a *continuum of data* in the multispectral space.
  - This happens for example when, in a land systems exercise, there is a gradation of canopy closure with position so that satellite or aircraft sensors might see a gradual variation in the mixture of canopy and understory. The information classes here might correspond to nominated *percentage mixtures* rather than to sets of well defined subclasses as depicted in Figure.
  - It is necessary in situations such as this to determine appropriate sets of spectral classes that represent the information classes.
- In quantitative analysis it is the spectral classes that a computer will be asked to work with since they are the “natural" groupings or clusters in the data.
  - After quantitative analysis is complete *the analyst simply associates* all the relevant spectral classes with the one appropriate information class.
  - In the context of the most commonly adopted approach to classification, based on statistical models, spectral classes will be seen to be unimodal probability distributions and information classes as possible multimodal distributions.
  - The latter need to be resolved into sets of single modes for convenience and accuracy in analysis.

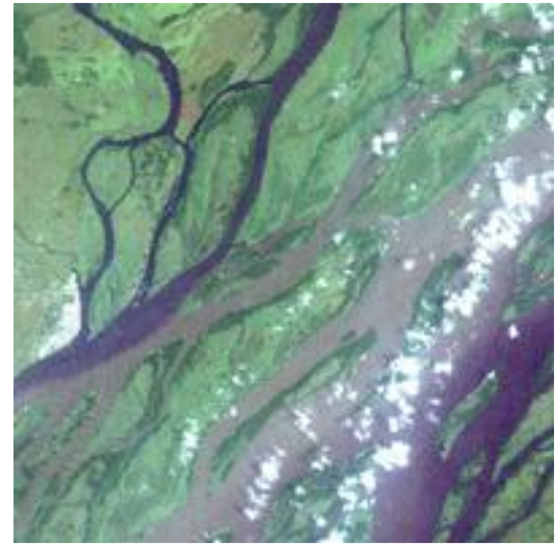
Quantitative Analysis by Pattern Recognition

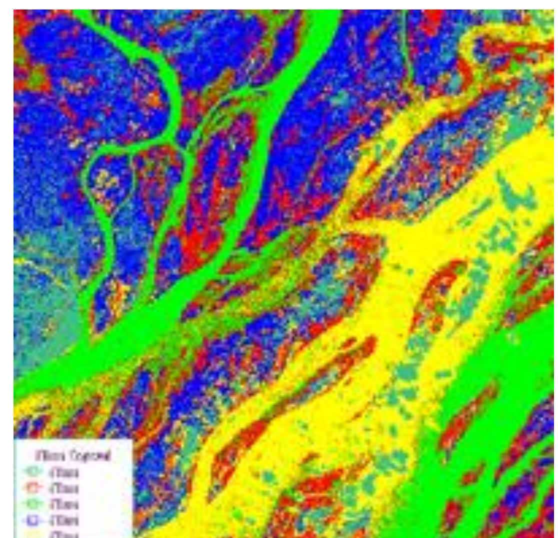
- Recognition that image data exists in sets of spectral classes, and identification of those classes as corresponding to specific ground cover types, is carried out using the techniques of mathematical **pattern recognition** or *pattern classification* and their more recent *machine learning* variants.
  - The patterns are the multiband *pixel vectors* themselves, which contain the sets of brightness values *for* the pixels arranged in column form $\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}$



Unsupervised Classification

- Unsupervised classification is a means by which pixels in an image are assigned to spectral classes <u>without the user having foreknowledge of the existence or names of those classes</u>.
- It is performed most often using **clustering methods**.
  - These procedures can be used to determine the number and location of the spectral classes into which the data falls and to determine the spectral class of each pixel.
  - The analyst then identifies those classes afterwards by associating a sample of pixels in each class with available reference data, which could include maps and information from ground visits.
  - Clustering procedures are generally *computationally expensive* yet they are central to the analysis of remote sensing imagery.
- While the information classes for a particular exercise are known, the analyst is usually totally *unaware* of the spectral classes, or sub-classes as they are sometimes called.
- Unsupervised classification is therefore useful for *determining the spectral class composition of the data prior to detailed analysis by the methods of supervised classification.*





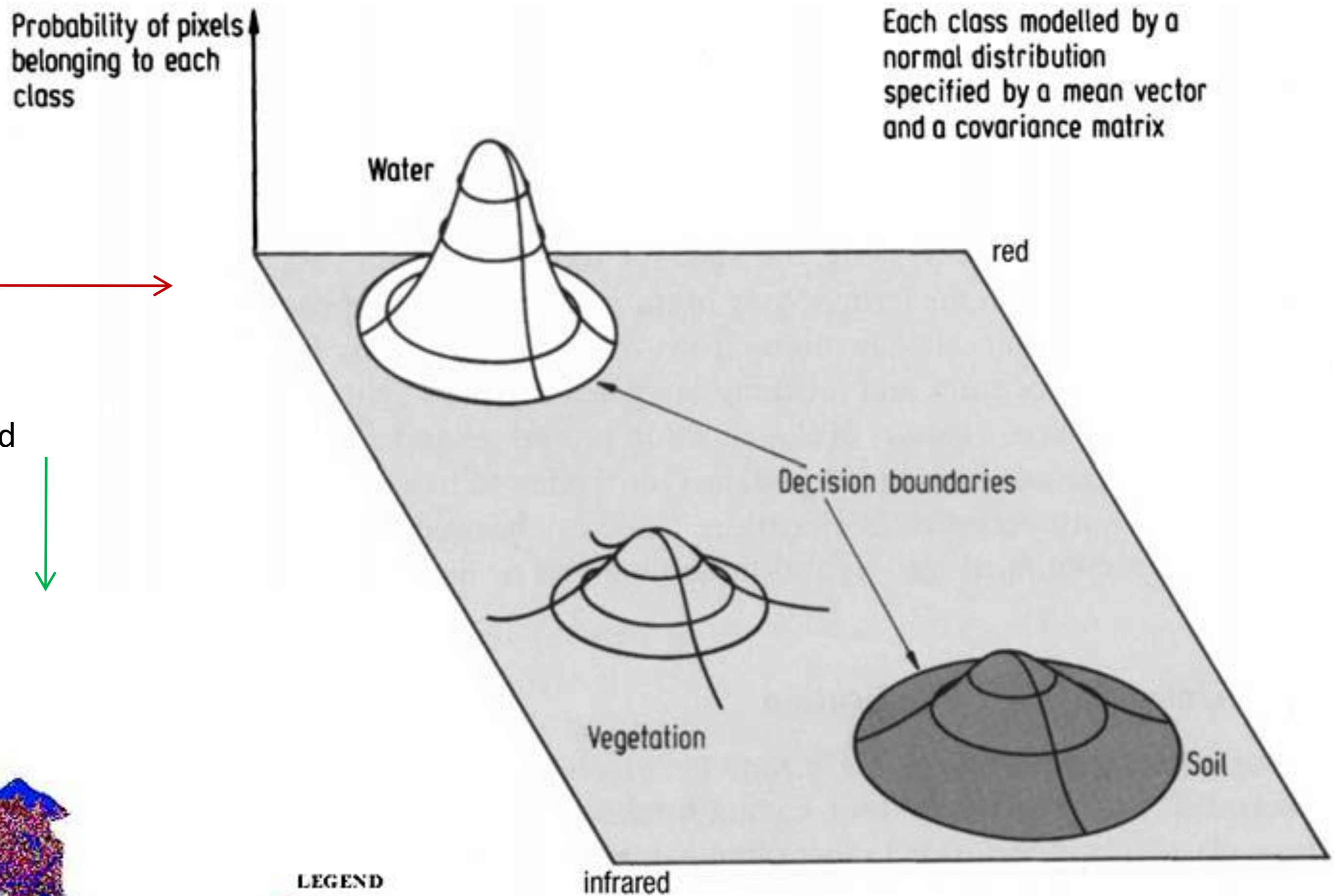
Supervised Classification

- A range of supervised classification procedures is possible.
  - **Statistical** methodologies have been the mainstay of quantitative analysis since the 1970s.
  - Other methods are based on non-statistical, **geometric** techniques that seek to place separating surfaces between the classes.
  - We will see both statistical and geometric supervised classification in detail.

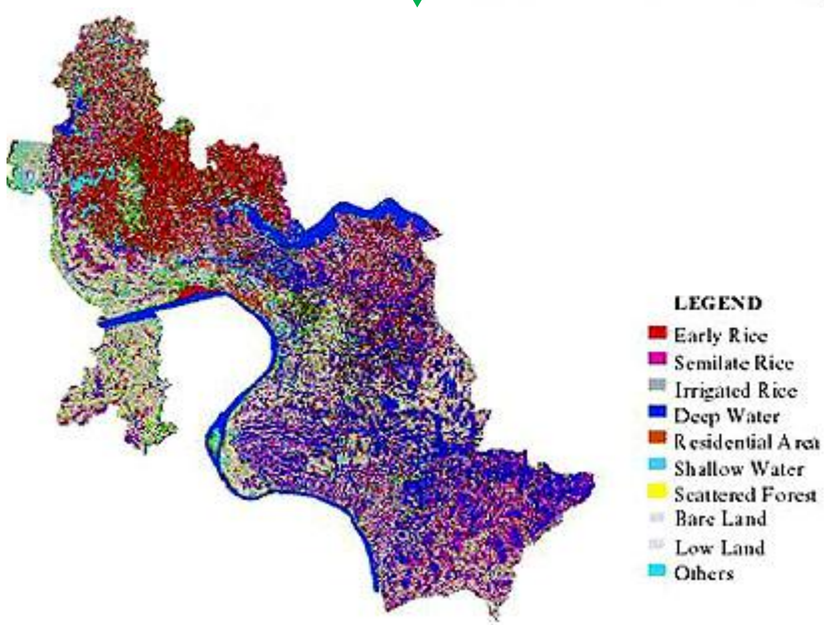
- Here follows a brief introduction to the concepts involved in <u>statistical classification</u>.
  - An important assumption in statistical supervised classification usually adopted in remote sensing is that *each spectral class can be described by a probability (multivariable) distribution in multispectral space*.
    - This is not unreasonable since it would be imagined that most pixels in a distinct cluster or spectral class would lie towards the centre and would decrease in density for positions away from the class centre, thereby *resembling* a probability distribution.
    - The distribution found to be of most value is the <u>normal or Gaussian distribution</u>. It gives rise to *tractable mathematical descriptions of the supervised classification process*, and is *robust* in the sense that *classification accuracy is not overly sensitive to violations of the assumptions* that the classes are normal.
  - A two dimensional multispectral space with the spectral classes so modeled is depicted in the next Figure.
    - The <u>decision boundaries</u> shown in the figure represent those points in multispectral space where a pixel has *equal chance of belonging to two classes*. The boundaries therefore partition the space into regions associated with each class.
    - *This is not new to telecommunications students*, isn't it?

Supervised Classification

The idea of statistical (parametric) modeling for supervised learning



Example of supervised classification result



Supervised Classification

- Supervised classification consists therefore of three broad steps.

1. First a set of training pixels is selected for each spectral class.
   - This may be done using information from ground surveys, aerial photography, topographic maps or any other source of reference data.
2. The second step is the learning phase.
   - In the case of statistical methods, <u>this lead to the parameters</u> of the statistical model which is assumed to be valid.
3. The third step is the classification phase, in which the relative likelihoods for each pixel in the image are computed and the pixel labeled according to the *highest likelihood*.

- The view of supervised classification adopted here has been based upon an assumption that the classes can be modeled by probability distributions and, as a consequence, are described by the parameters of those distributions.
  - As a result it is also referred to as a **parametric supervised method**.
- Other supervised techniques also exist, in which neither distribution models nor parameters are relevant.
  - These are referred to as **non-parametric methods**.
  - More recently, *neural networks* and *support vector machine* non-parametric classification methods have been shown to offer promise in remote sensing applications.