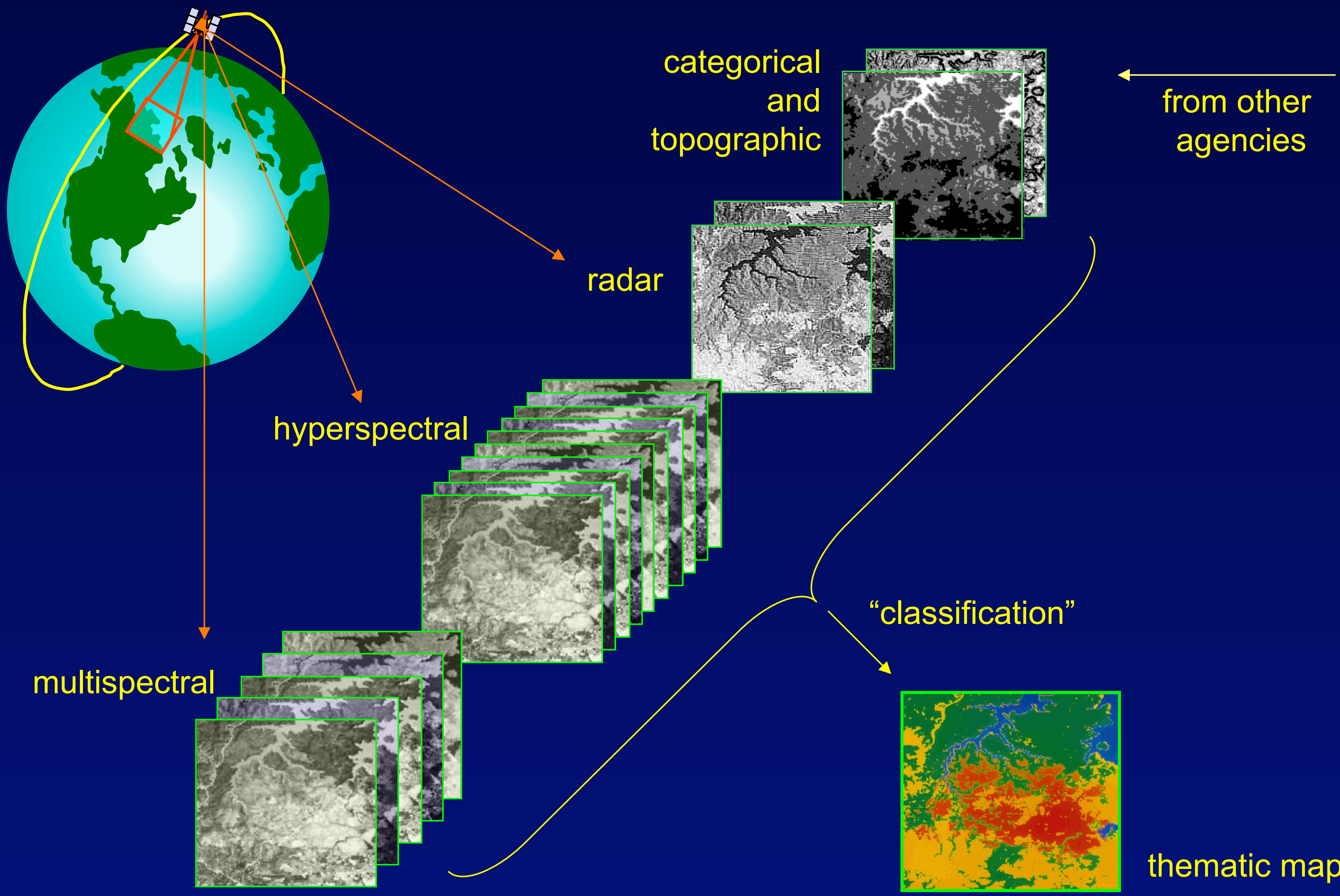
THE AUSTRALIAN NATIONAL UNIVERSITY

Thematic Mapping from Mixed Spatial Data Types

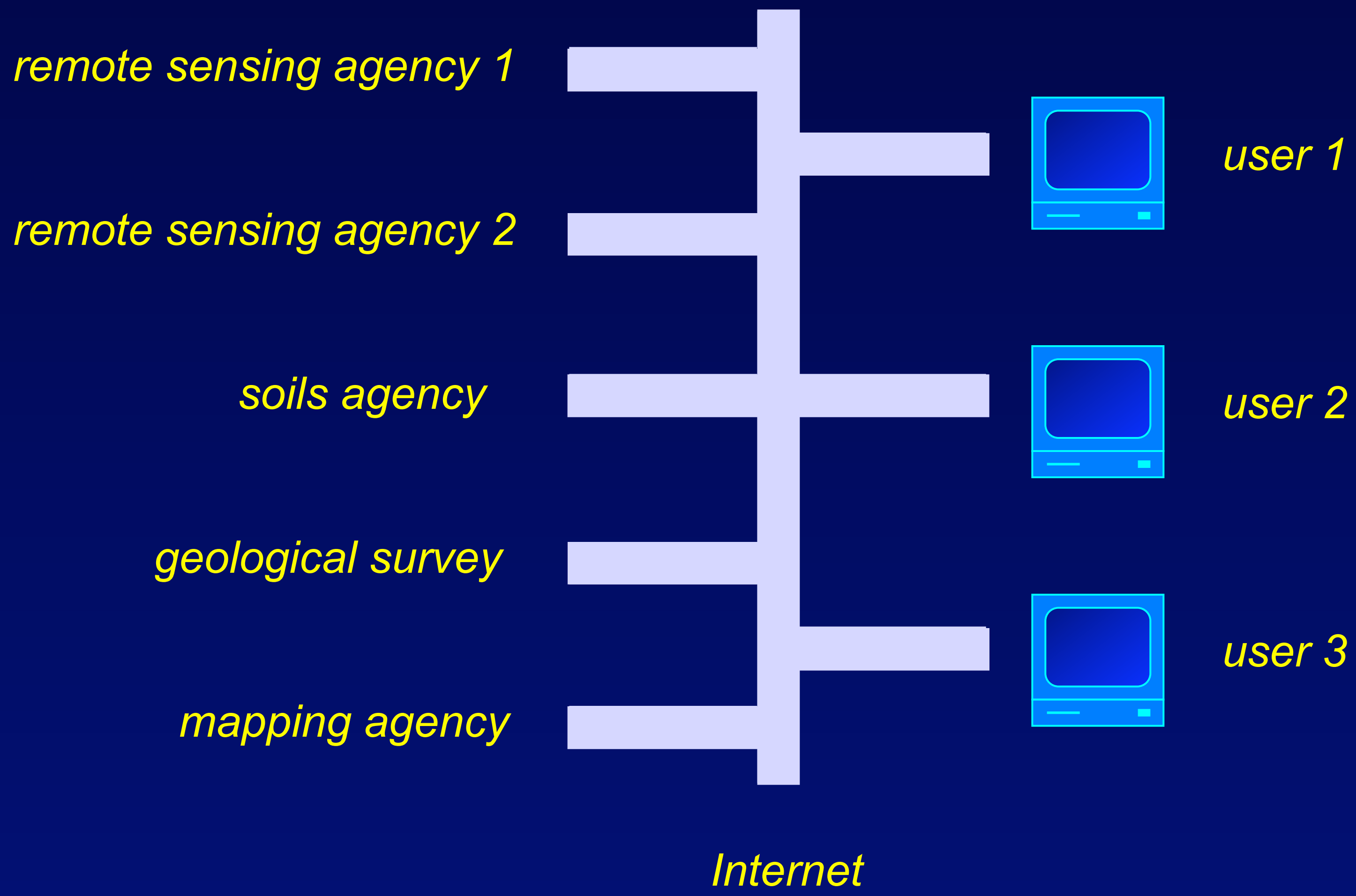
John Richards

Research School of Information Sciences and Engineering
The Australian National University

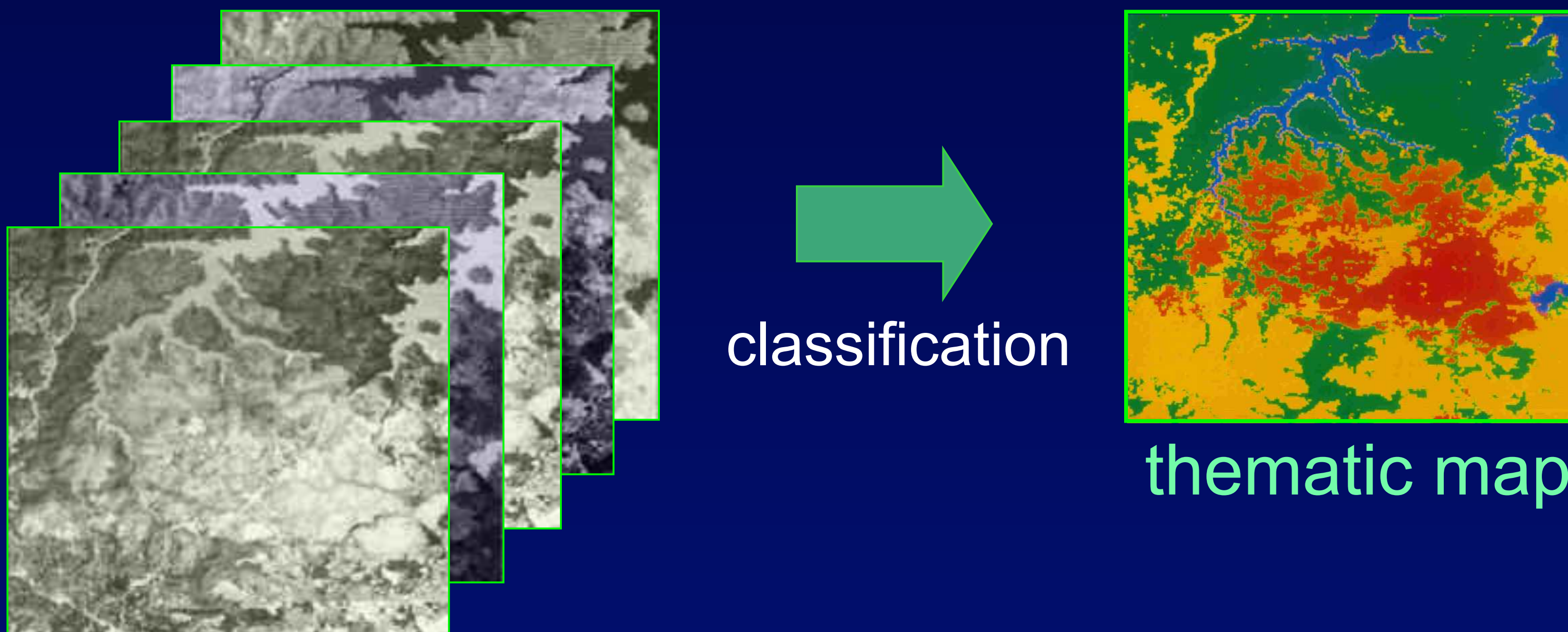
The essential problem

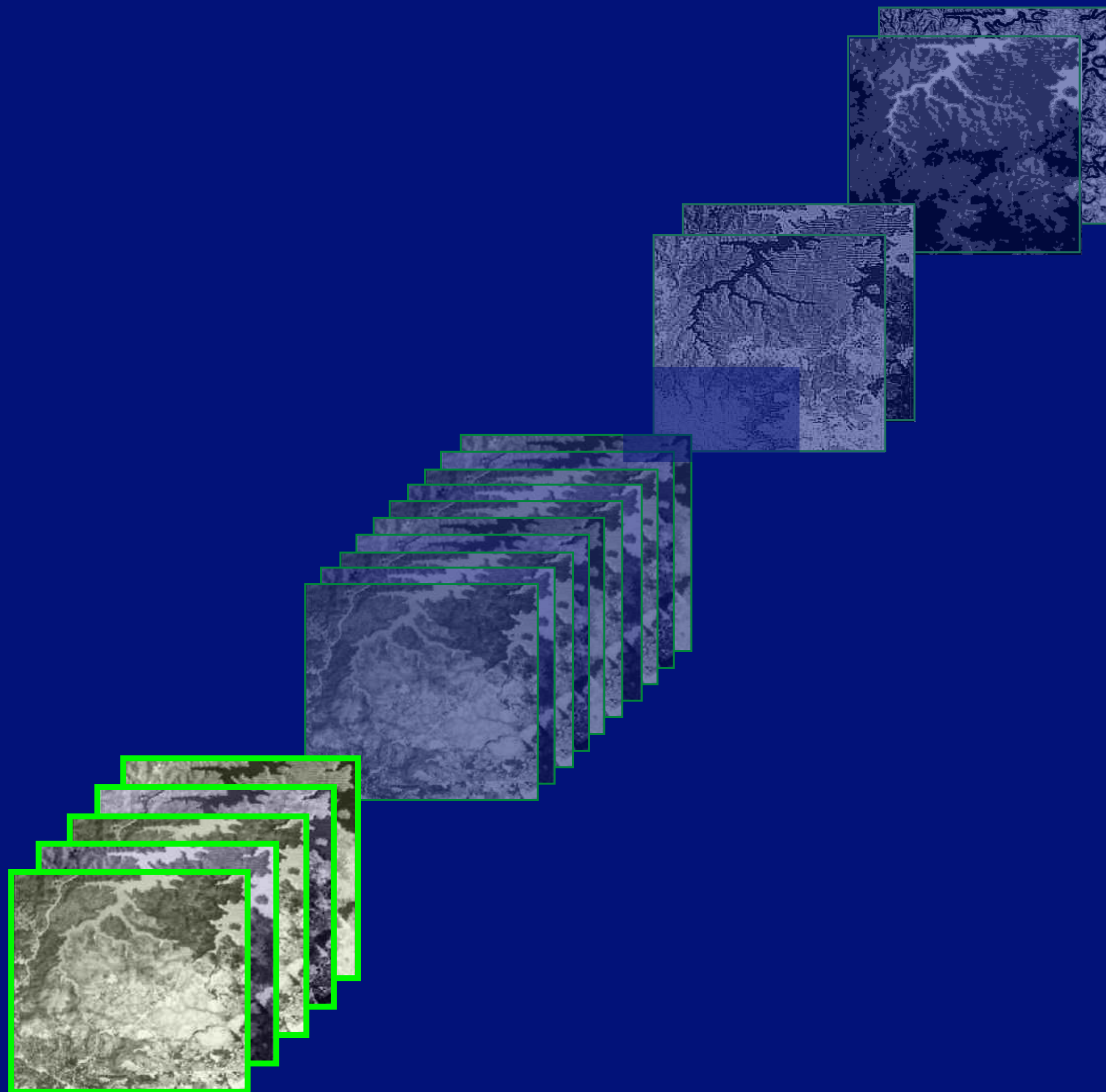


Consider the possibility of a distributed GIS



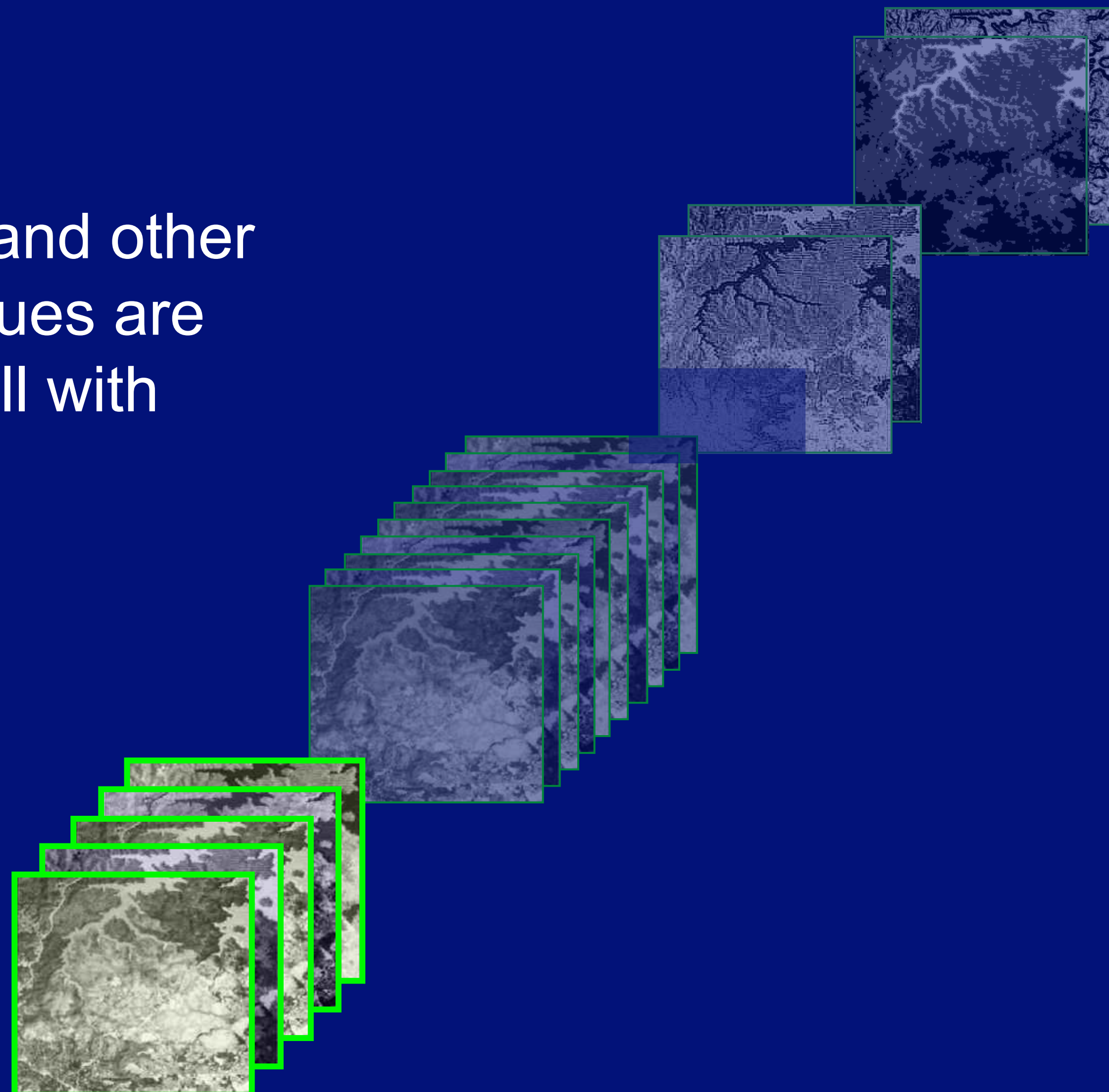
Classification is a mapping from data to labels (symbols)



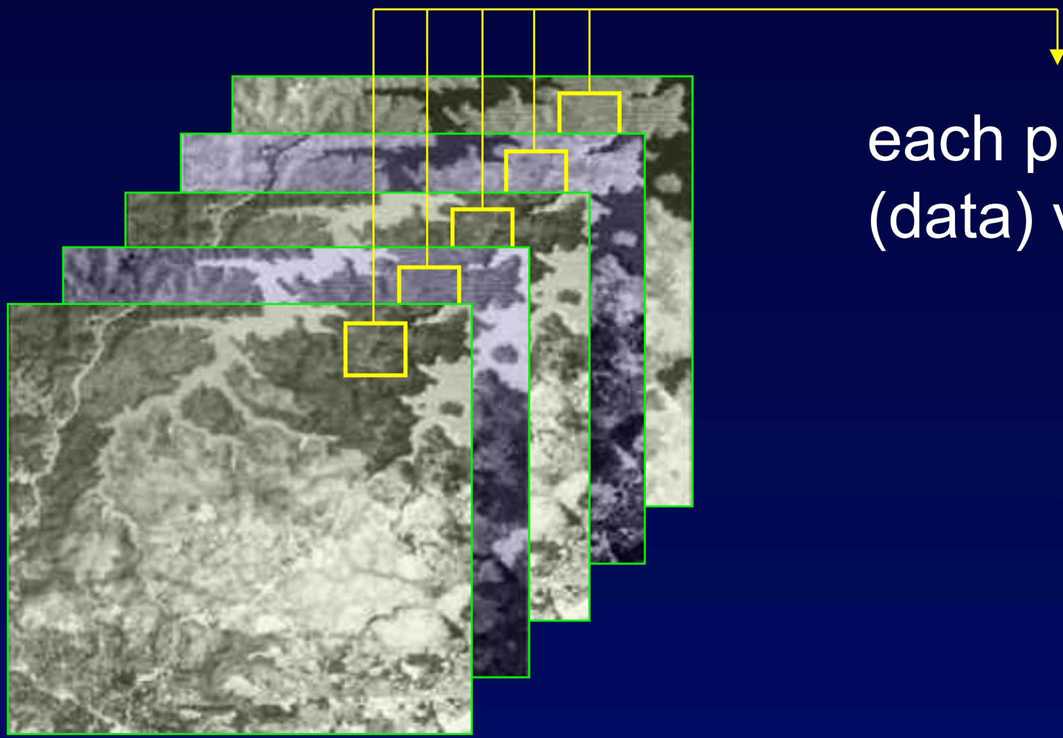


multispectral data

statistical, neural and other numerical techniques are known to work well with multispectral data



multispectral data



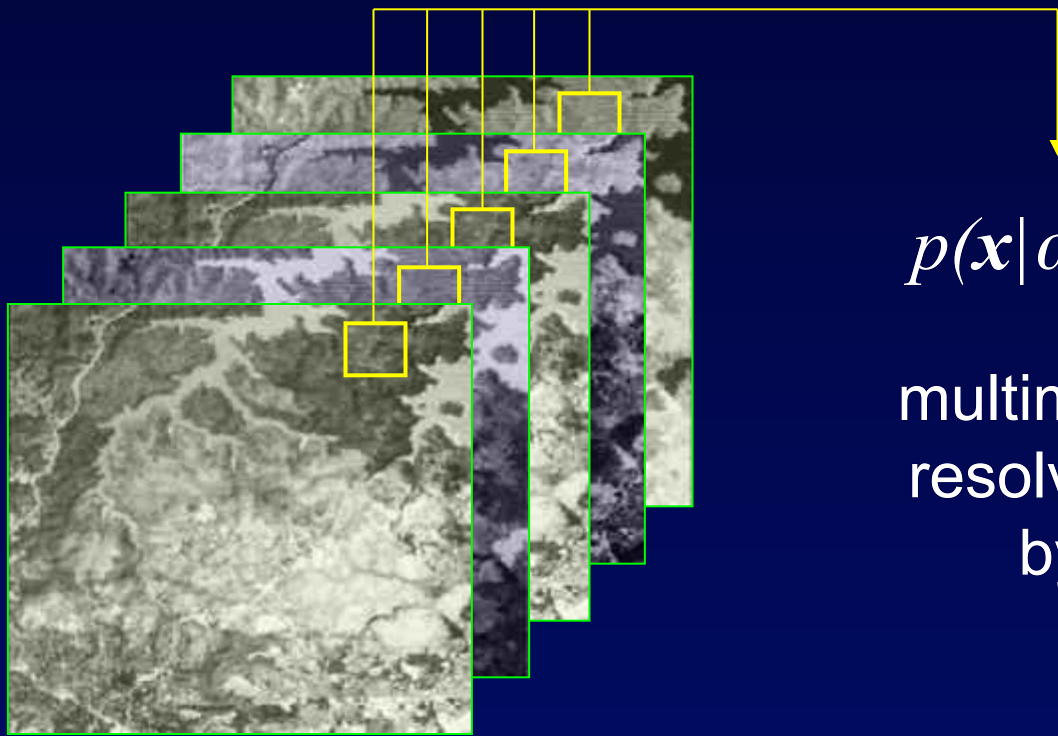
each pixel is described by a (data) vector of measurements

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \end{bmatrix}$$

The pixels in each class are assumed to be distributed in a multidimensional normal fashion (Gaussian distribution)

$$p(\boldsymbol{x}|\omega) \sim N\{\, \boldsymbol{m}, \Sigma \,\}$$

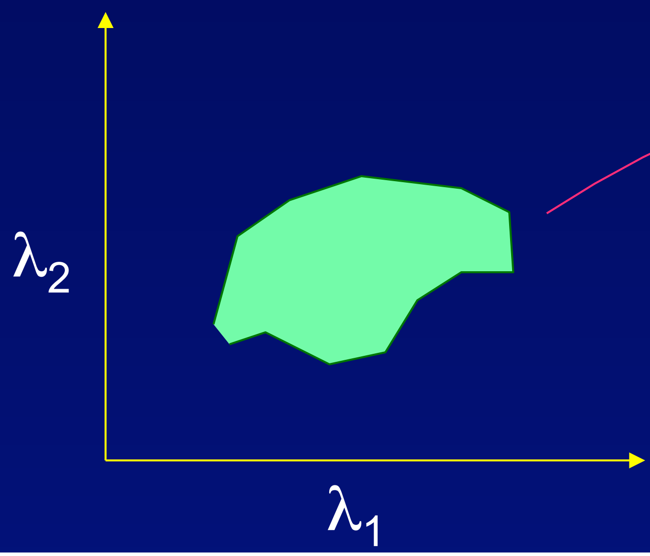
$\boldsymbol{m}$ and Σ are "learned" through training

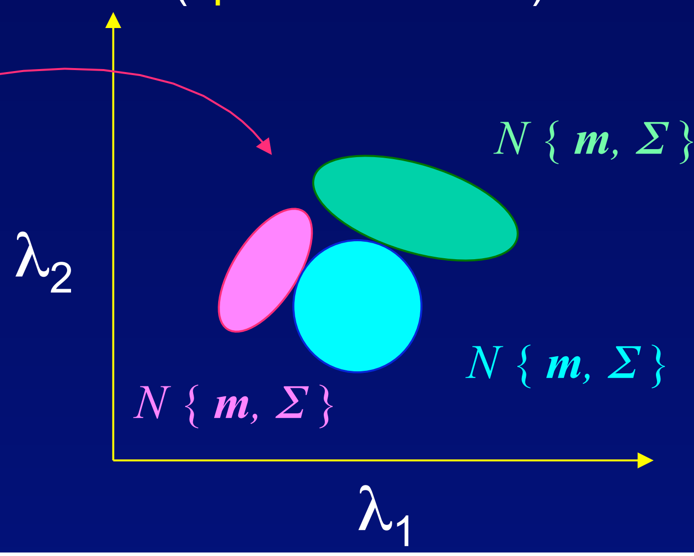


$$p(\boldsymbol{x}|\omega) \sim N\,\{\,\boldsymbol{m},\,\boldsymbol{\Sigma}\,\}$$

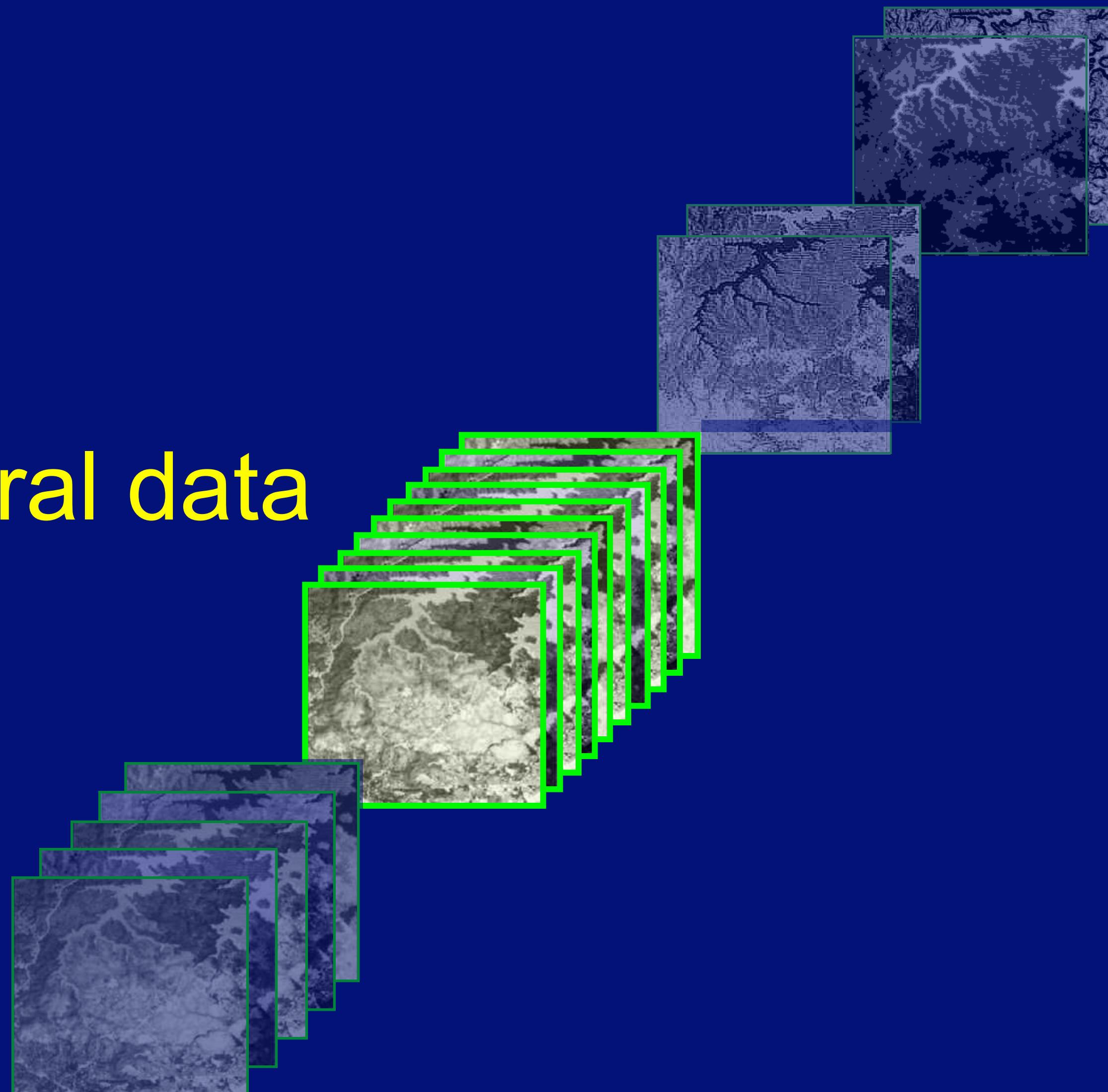
multimodality can be resolved (modelled) by clustering

resolved into clusters (spectral classes)





hyperspectral data



For hyperspectral data there are two approaches

- standard classification methods
  (both parametric and non-parametric)

  requires careful treatment of the class covariance matrix

- interpretation by an expert
  (usually involves library searching)

  requires spectral libraries and expert knowledge to be available

Hyperspectral data analysis: the classification approach

There is a problem with estimating the elements of the class conditional covariance matrix Σ from the available training data with hyperspectral data sets, which for c channels (or bands) is symmetric of size c x c

$$\Sigma = \begin{bmatrix} v_{11} & v_{12} & v_{13} & v_{14} & v_{15} \\ & v_{22} & v_{23} & v_{24} & v_{25} \\ & & v_{33} & v_{34} & v_{35} \\ & & & v_{44} & v_{45} \\ & & & & v_{55} \end{bmatrix}$$

$(c = 5)$

$\frac{1}{2}c(c+1)$ independent elements

need $c(c+1)$ independent samples to avoid singularity

each pixel vector contains c spectral samples

Therefore we need at least (c+1) training pixels per class – say 10(c+1)

Hyperspectral data analysis: the classification approach

Handling the ill-conditioned covariance matrix:

Regularisation (approximation)

Simplification (truncation)

Regularisation

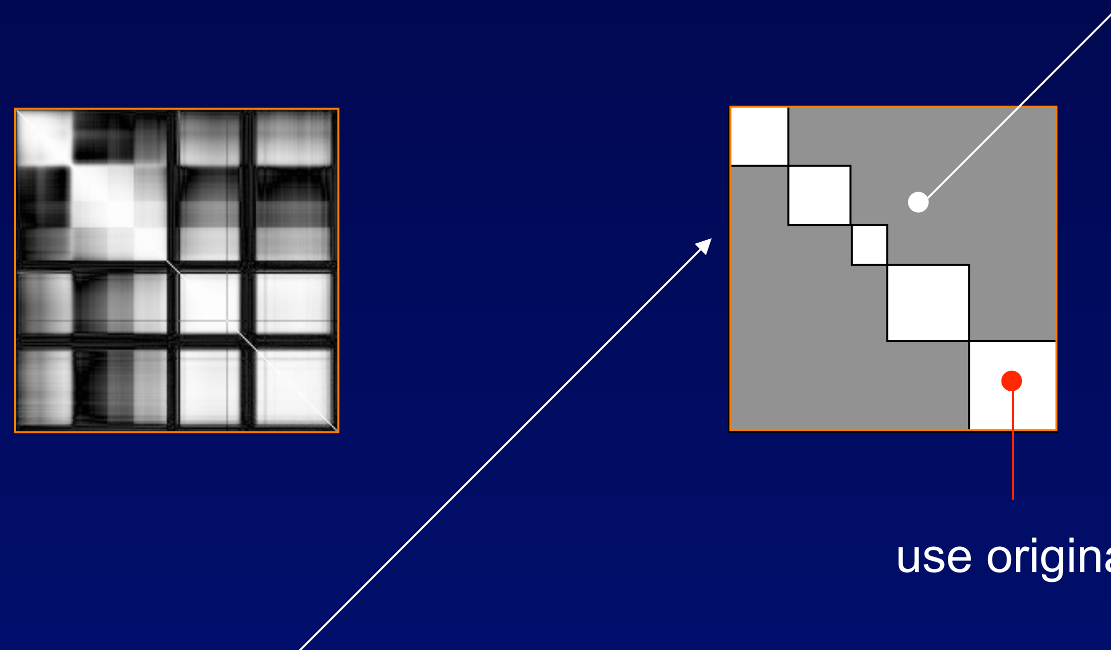
$$\hat{\Sigma}_i = \alpha\, \Sigma_{i(est)} + (1-\alpha)\, B$$

a simple, more reliable estimate, such as the global covariance or even a diagonal version of the class conditional covariance

Hyperspectral data analysis: the classification approach

Simplification

assume zero correlation
and thus covariance



use original covariances

The sub-matrices are independent. Therefore:

- The discriminant function is the sum of the discriminants of the blocks
- The number of training samples is determined by the largest sub-matrix.

For hyperspectral data there are two approaches

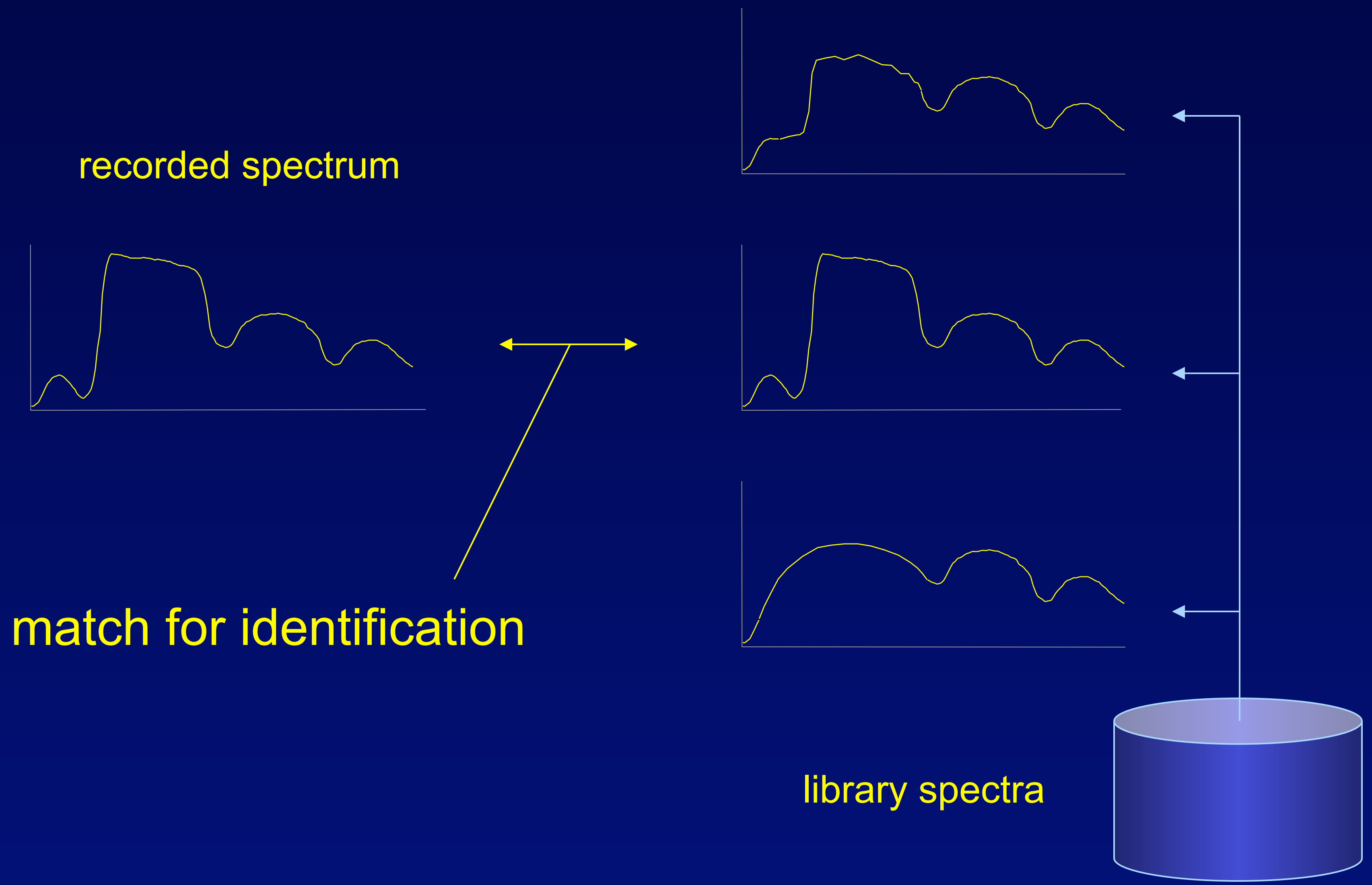
- standard classification methods
  (both parametric and non-parametric)

  requires careful treatment of the class covariance matrix

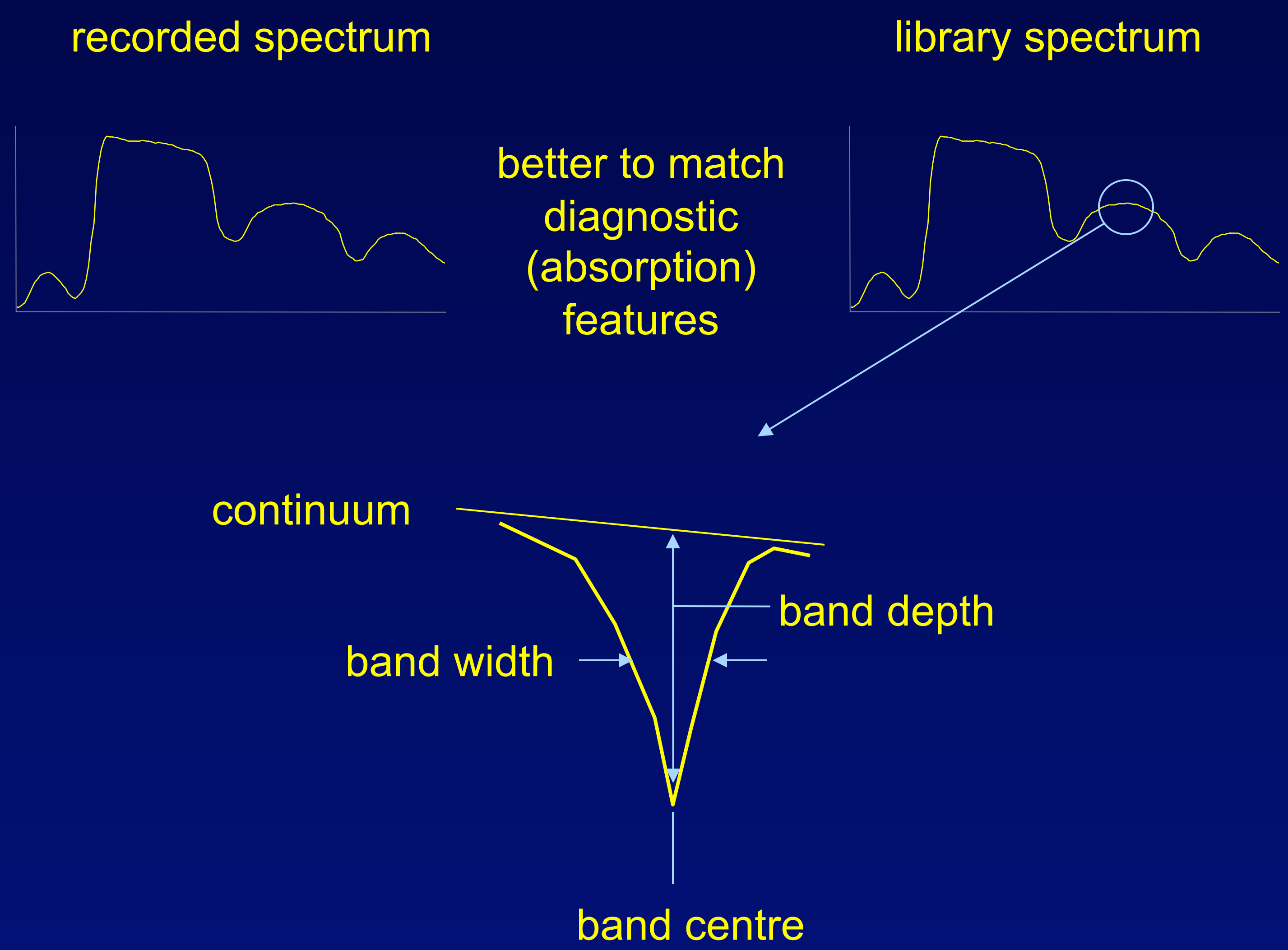
- interpretation by an expert
  (usually involves library searching)

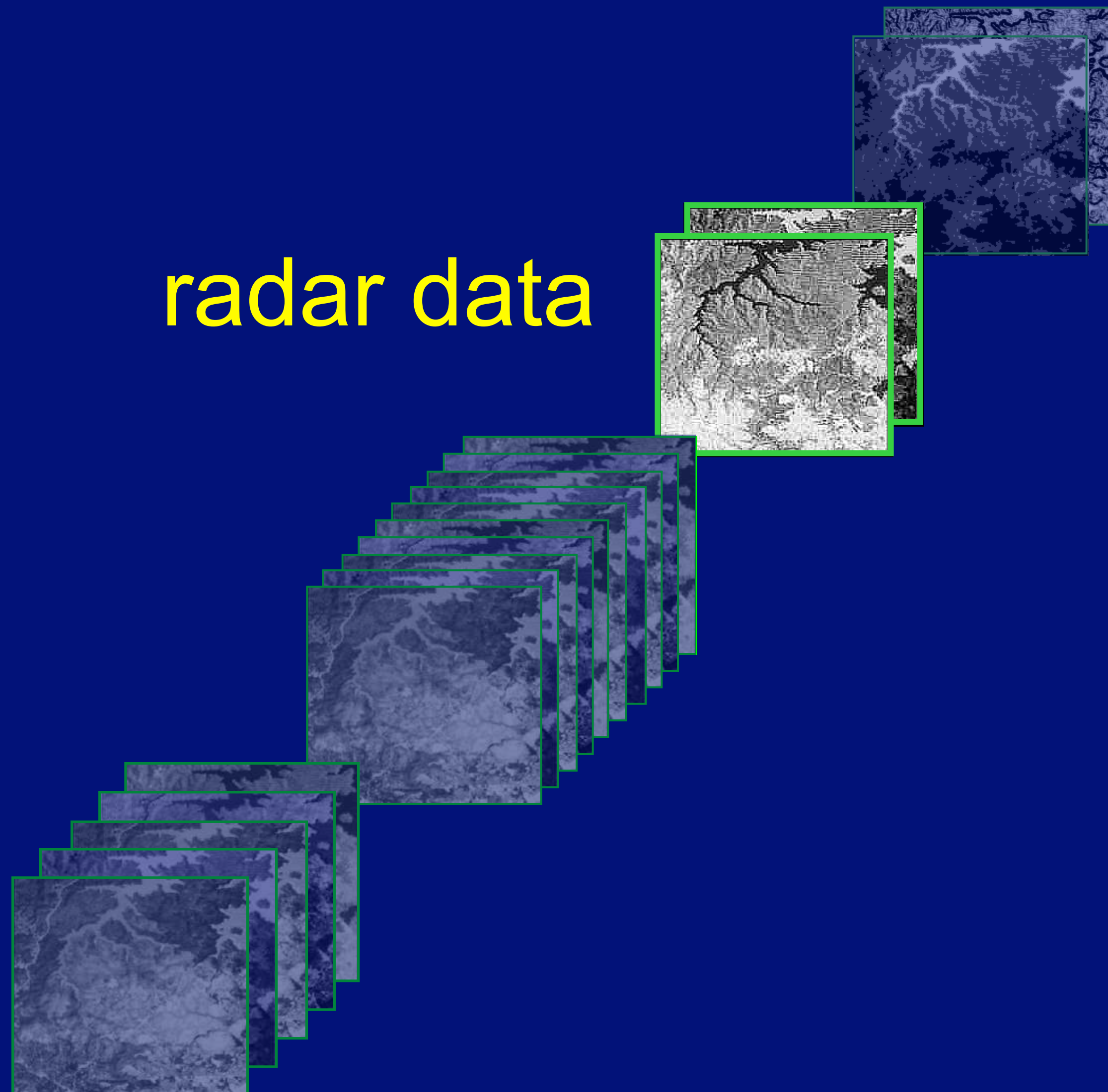
  requires spectral libraries and expert knowledge to be available

Hyperspectral data analysis: interpretation by an expert



Spectroscopic interpretation



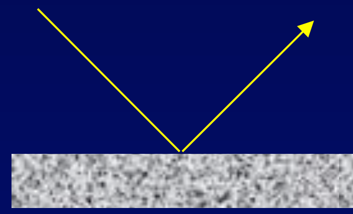


radar data

Considerations when interpreting radar data

signal to noise ratio is poor because of speckle

scattering is coherent



$\rho = |\rho| \angle\phi$

scattering is more than from surface elements

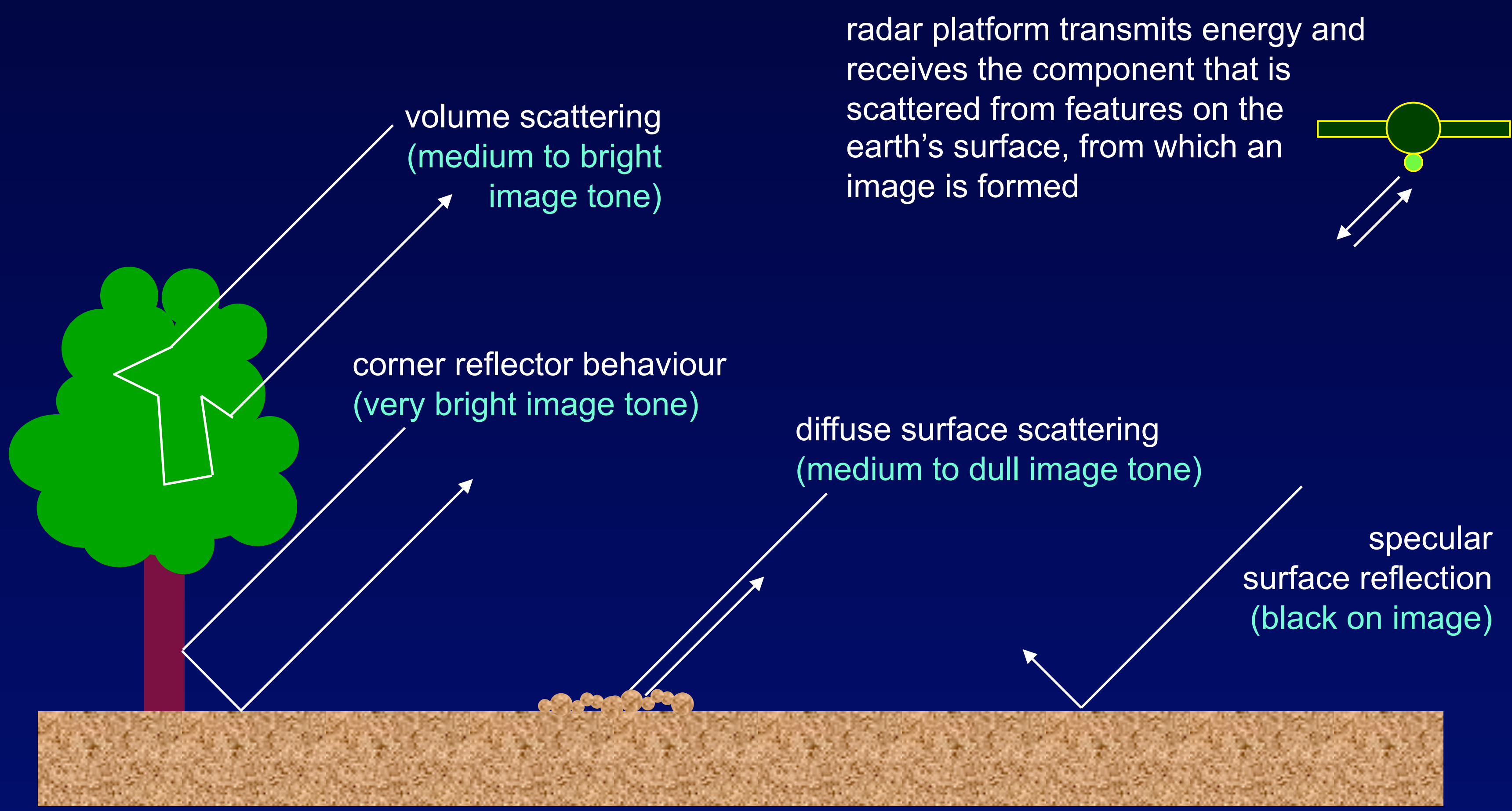


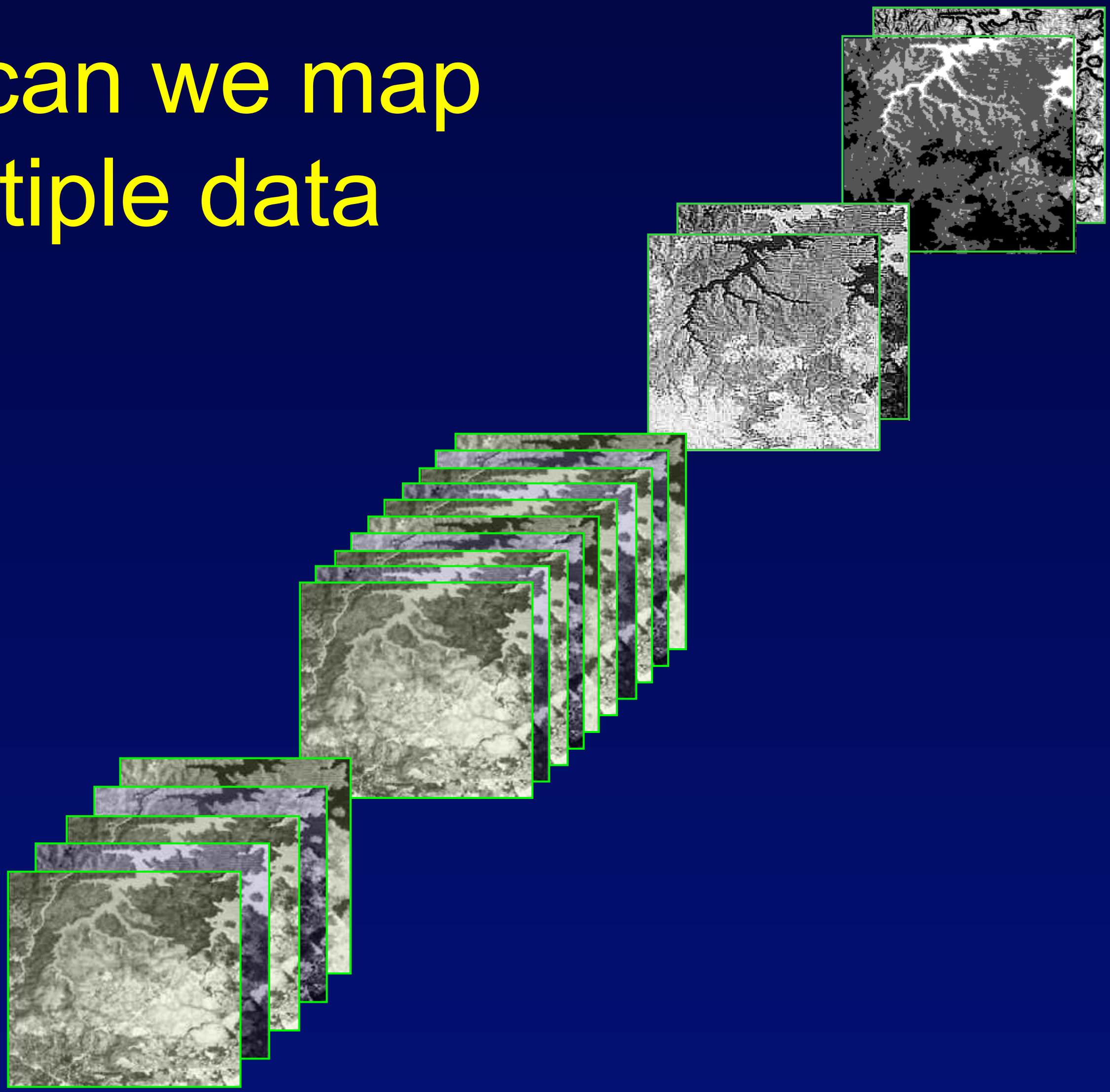
image tone is a function of moisture content, incidence angle, wavelength and polarisation

Particular data types have preferred methods for analysis

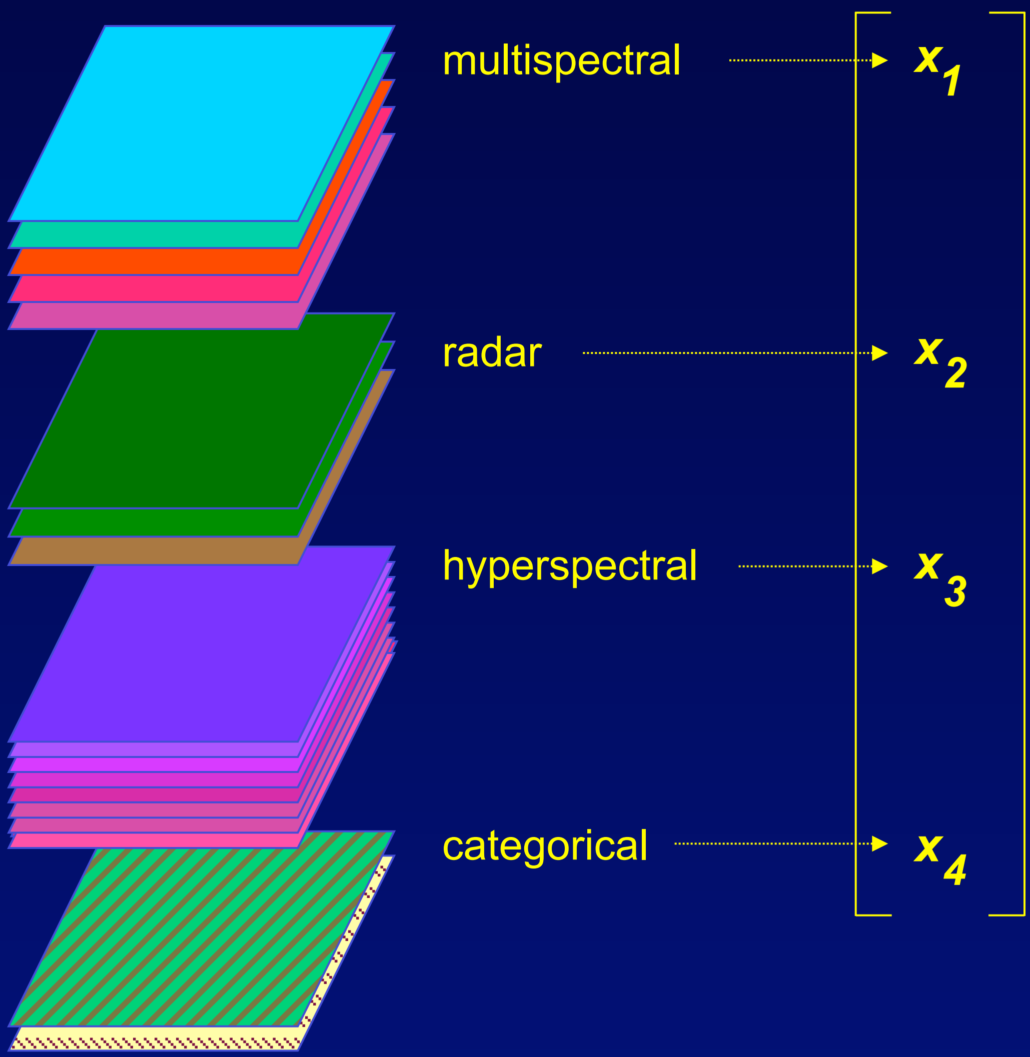
multispectral	statistical and neural methods
hyperspectral	library searching methods or approximate statistical
radar	backscatter modelling or sometimes statistical

The techniques are not easily transferable between data types if good results are expected

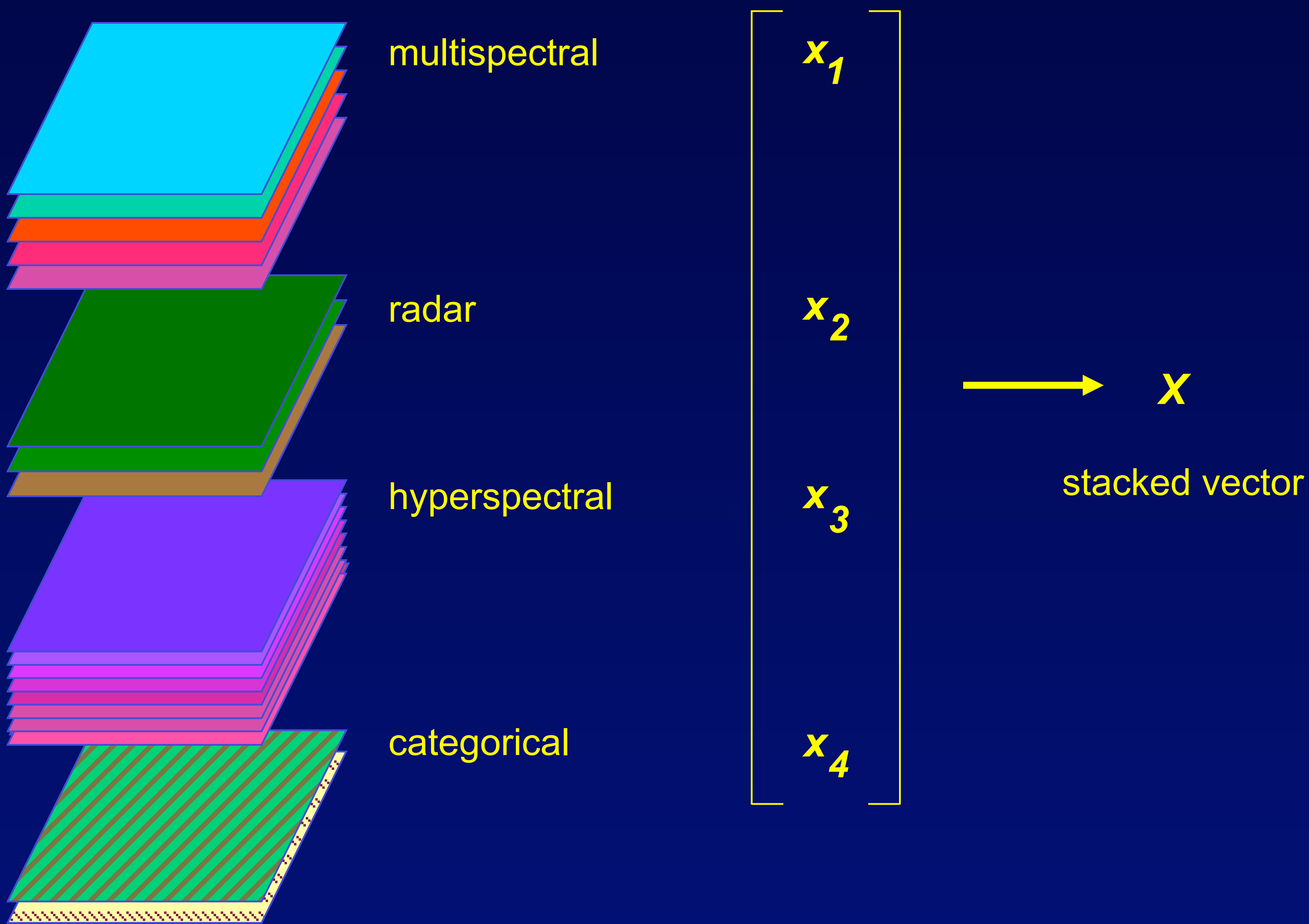
So how can we map from multiple data types?



Sometimes a "stacked" data vector is formed



Sometimes a “stacked” data vector is formed



Multisource statistical classification has been proposed

stacked vector

$$\boldsymbol{X} = [\, \boldsymbol{x}_1^t \, . \, . \, . \, \boldsymbol{x}_n^t \,]^t$$

decision rule

$$\boldsymbol{X} \in \omega_j \quad \textit{if} \quad p(\omega_j \,|\, \boldsymbol{X}) > p(\omega_k \,|\, \boldsymbol{X}) \qquad \textit{for all } k \neq j$$

source independence is assumed, leading to

$$p(\omega_j \,|\, \boldsymbol{X}) = p(\omega_j \,|\, \boldsymbol{x}_1)\, p(\omega_j \,|\, \boldsymbol{x}_2) \, . \, . \, . \, p(\omega_j \,|\, \boldsymbol{x}_n)\; p(\omega_j)^{1-n}$$

There are problems with the “stacked” vector approach

The different data types:

- May have different statistical properties
- May have very different dynamic ranges
- May have very different noise properties
- May not be available simultaneously (???)

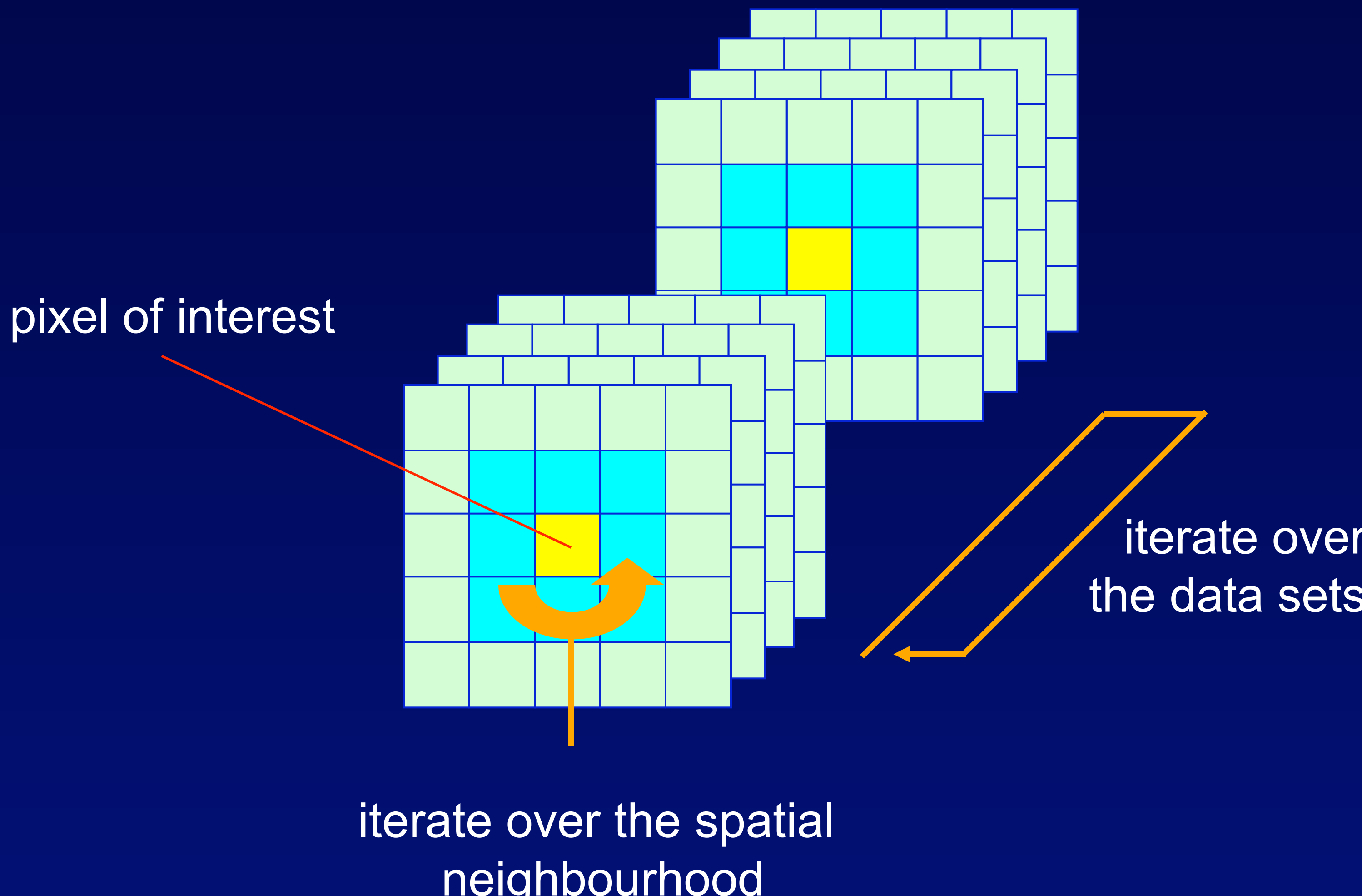
Use the prior probability term in the maximum likelihood rule

$$\boldsymbol{x} \in \omega_j \quad \text{if} \quad p(\omega_j \mid \boldsymbol{x}) > p(\omega_k \mid \boldsymbol{x}) \qquad \text{for all } k \neq j$$

applying Bayes theorem leads to

$$\boldsymbol{x} \in \omega_j \quad \text{if} \quad p(\boldsymbol{x} \mid \omega_j)\, p(\omega_j) > p(\boldsymbol{x} \mid \omega_k)\, p(\omega_k) \quad \text{for all } k \neq j$$

Use label relaxation



Evidential classification is a useful fusion method

a measure of evidence (called mass) is formed for each labelling possibility for a given pixel from a given data source

for example, for data source $\boldsymbol{x}_1$

$$m_1(<\omega_1, \omega_2, \omega_3, \theta>) = <0.40, 0.20, 0.20, 0.20>$$

while for data source $\boldsymbol{x}_2$

$$m_2(<\omega_1, \omega_2, \omega_3, \theta>) = <0.20, 0.45, 0.30, 0.05>$$

labelling propositions can also be uncertain, in that mass can also be assigned to expressions such as $\omega_2 \vee \omega_3$

this suggests that there is residual doubt in our mind as to which of those two classes the pixel is in

Dempster and Shafer

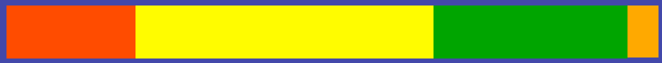
Evidential classification is a useful fusion method

for example, for data source $\boldsymbol{x}_1$

$m_1(<\omega_1,\omega_2,\omega_3,\theta>) \longrightarrow$

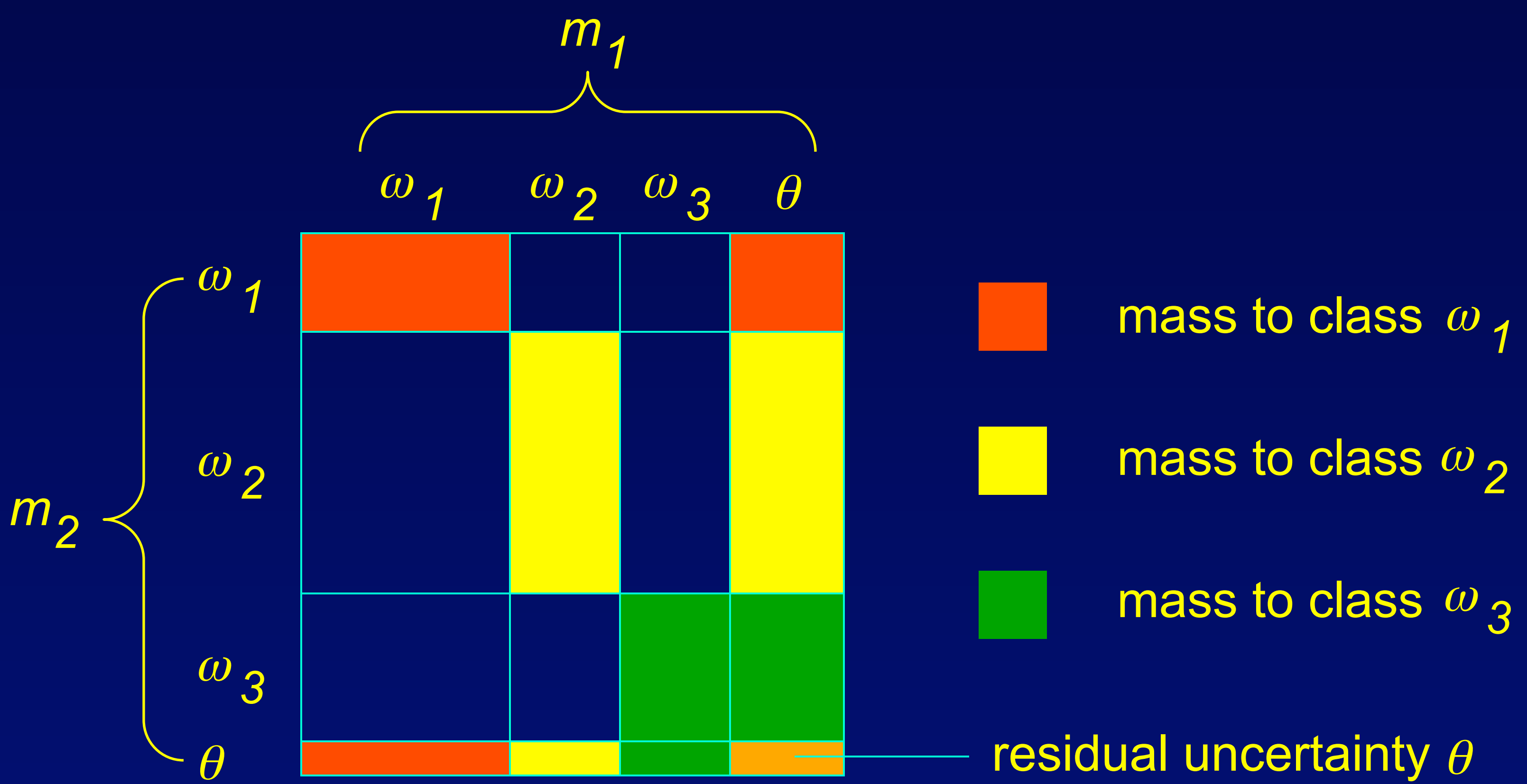
while for data source $\boldsymbol{x}_2$

$m_2(<\omega_1,\omega_2,\omega_3,\theta>) \longrightarrow$



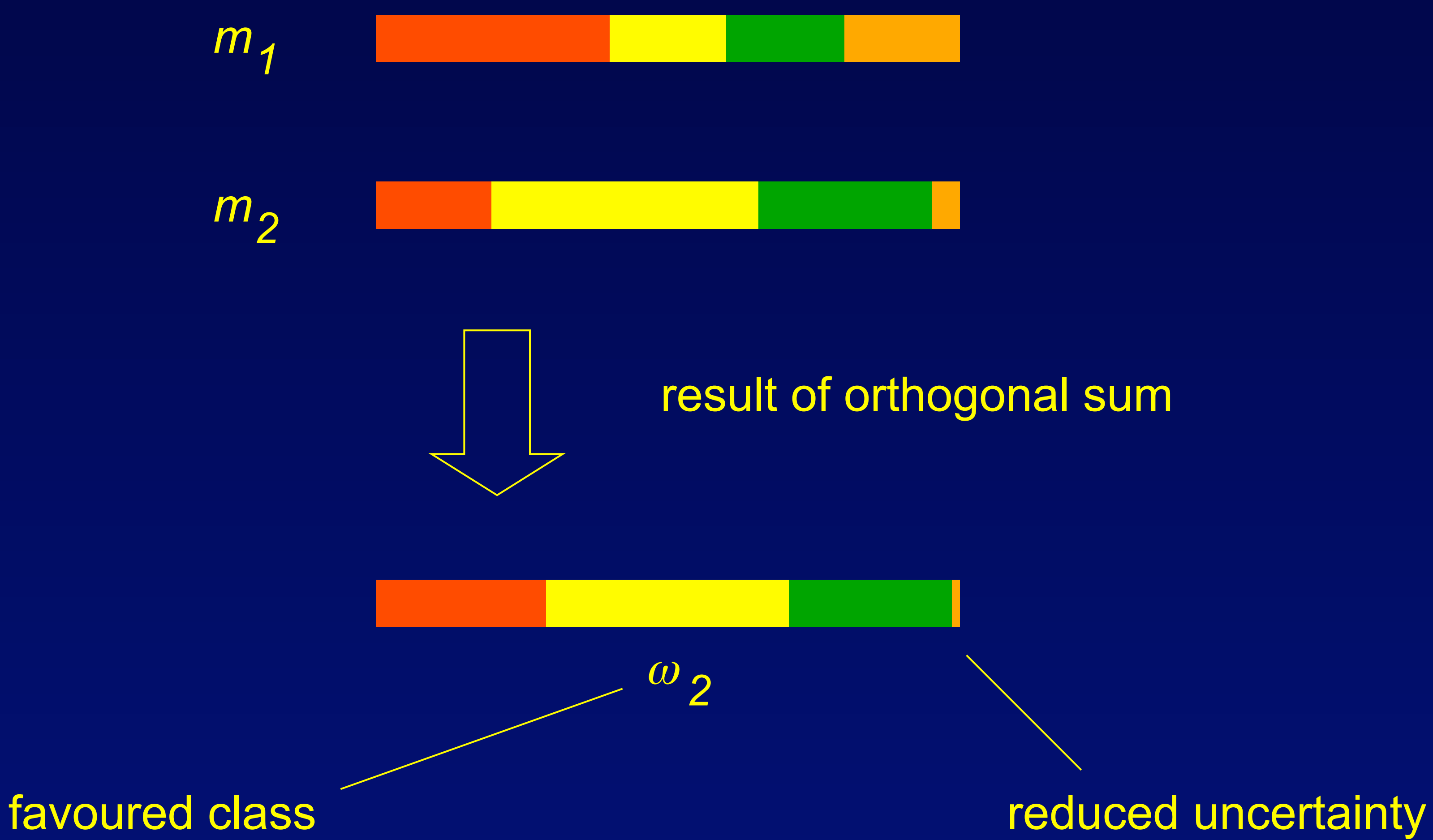
Evidential classification - cont

the measures of evidence are fused using the “orthogonal sum”

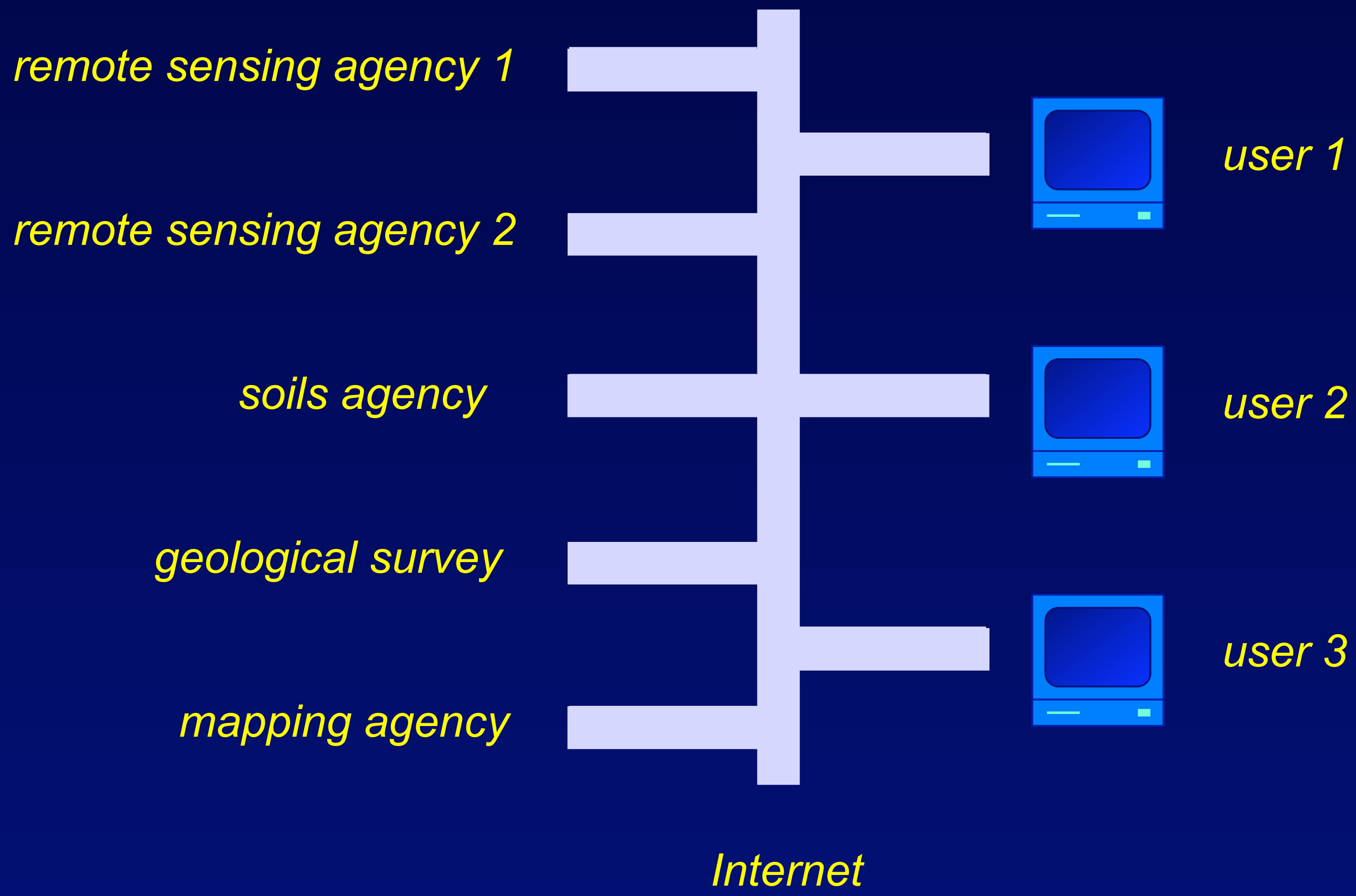


the operation can be applied repetitively to handle more data sources

Evidential classification - cont



Recall we want to operationalise a distributed GIS



What should a GIS thematic mapping task do?

- Efficiently handle data from different sensors
- Allow categorical data to be “fused” with numerical data
- Allow each data source to be analysed separately
- Allow quality of the data sources to be considered
- Allow for classes from the combined data to be different from the classes found with the data sources separately

Classes can be data specific

multispectral	wheat, pine trees, clear water, sand, clay, snow, alfalfa, grassland, etc
radar	volume scatterer, smooth surface, rough surface, moist, dry, strong reflector
hyperspectral	geochemical and biochemical composition
maps	soil types, geological units, land ownership, town planning, topography

How do the techniques match up?

	stacked vector	evidential	statistical	use priors	relaxation labelling
data from different sensors	x	x	x	x	x
incorporate categorical data		x			
analyse data sources separately		x	x	x	x
handle data source quality	x	x	x		x
combined data classes different					

Consider a human photointerpreter

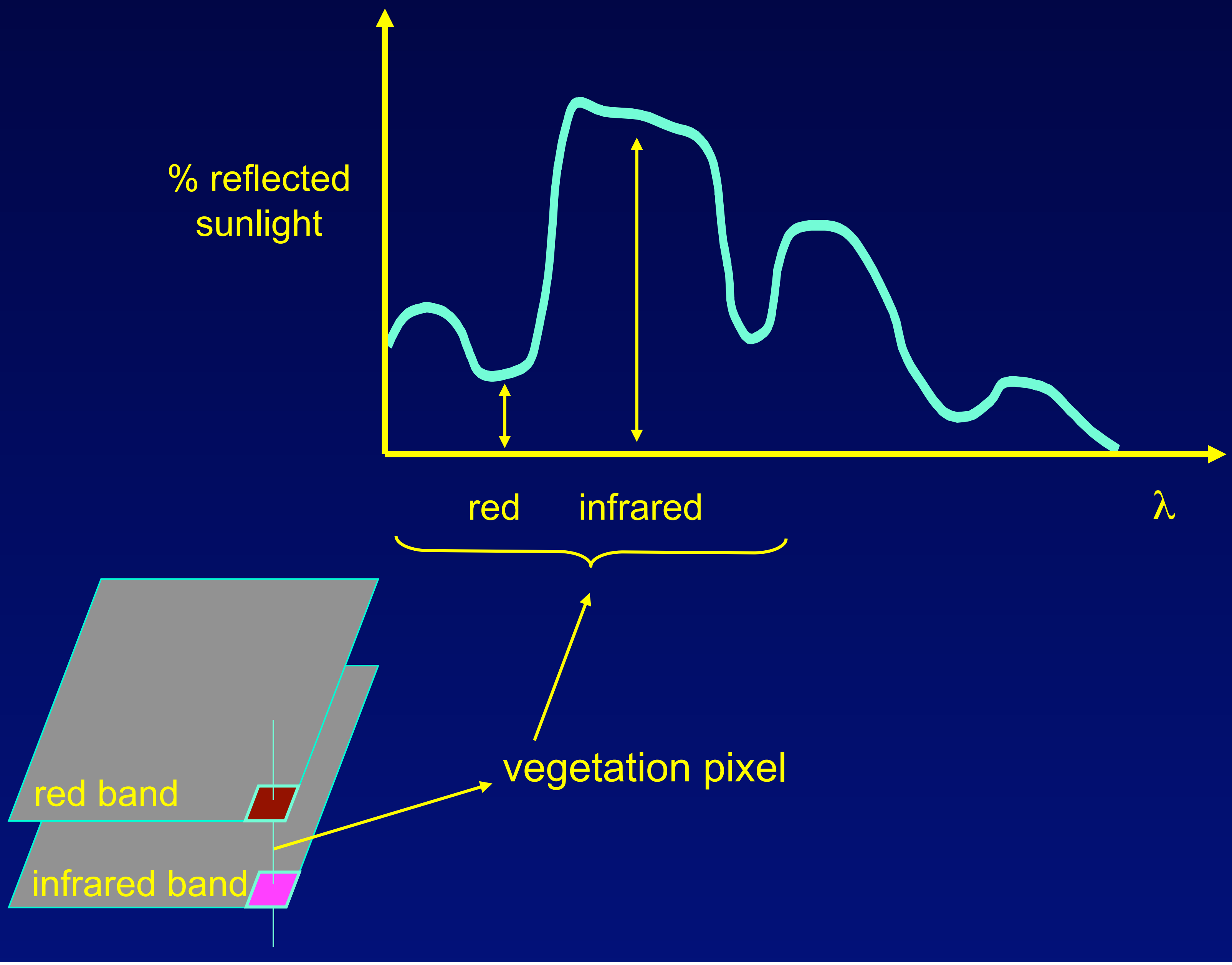
Although relatively poor in a quantitative sense, a human photointerpreter can cope with each of the issues

- data from different sensors
- incorporate categorical data
- analyse data sources separately
- handle data source quality
- combined data classes different

How does a human photointerpreter handle mixed data?

- Uses expert knowledge
- Tends to reason in labels
- And can handle relative data source quality (subjectively)

How does a human photointerpreter handle mixed data?



or, expressed another way

if (infrared/red) is high **then** probably vegetation

This is a *production rule*, used in an *expert system*

The general form of a production rule is:

if condition **then** inference

Rules can be used

to go from data to labels (ie for basic thematic mapping):

if (infrared/red) is high **then** probably vegetation

if radar tone (DN) is dark **then** smooth cover type

Rules can be used

to go from data to labels (ie for basic thematic mapping):

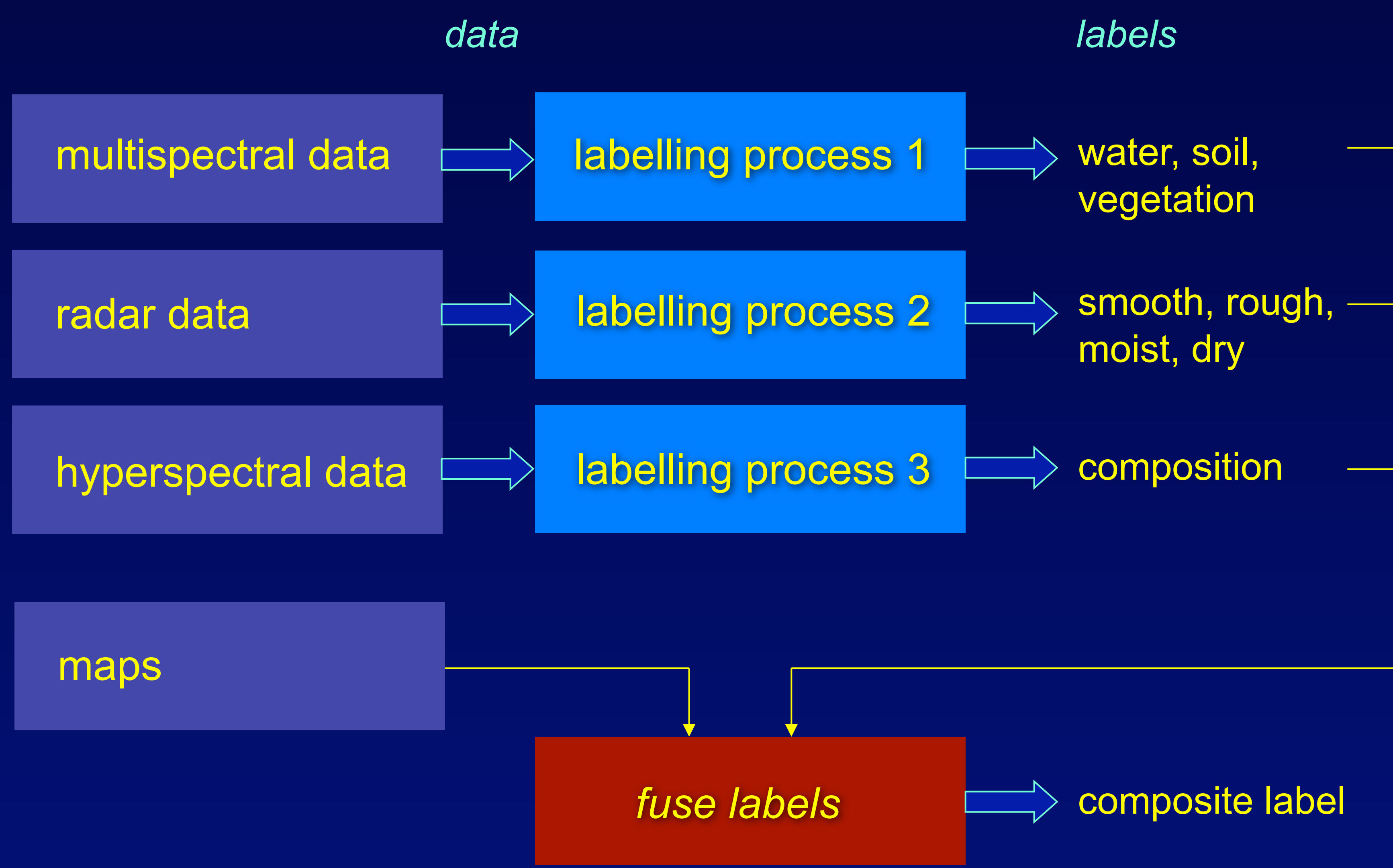
if (infrared/red) is high **then** probably vegetation

if radar tone (DN) is dark **then** smooth cover type

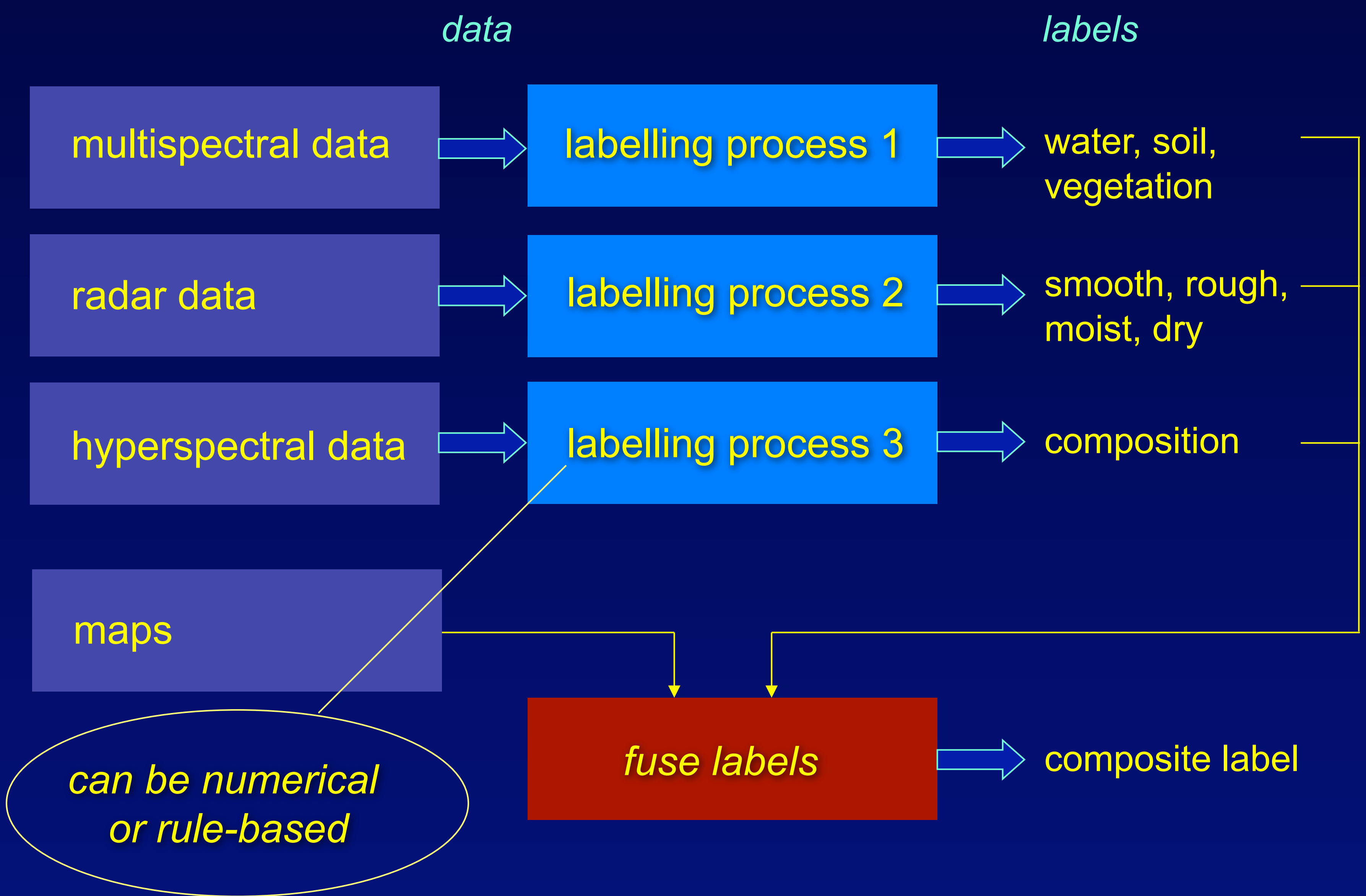
to process labels, in which case compound rules are used:

if vegetation **and** smooth **then** probably grassland

Should fuse at the label level and not the data level



Should fuse at the label level and not the data level



Thus, how should a GIS thematic mapping task work?

Analyse data from each source using most appropriate algorithm and information classes for that data type. In this manner it is not even necessary to have each data set available simultaneously.

Data from each source has been mapped into labels - ie a common “language”.

Using expert rules (or other symbolic process) fuse at the label level. Final labels can be quite different from those relevant to each data source on its own. Data quality can be built into the rules.

So what's the outstanding challenge?

To refine an expert system or machine learning approach to thematic mapping that can be used in a production sense

So what's the outstanding challenge? It is to operationalise.