Introduction to Remote Sensing

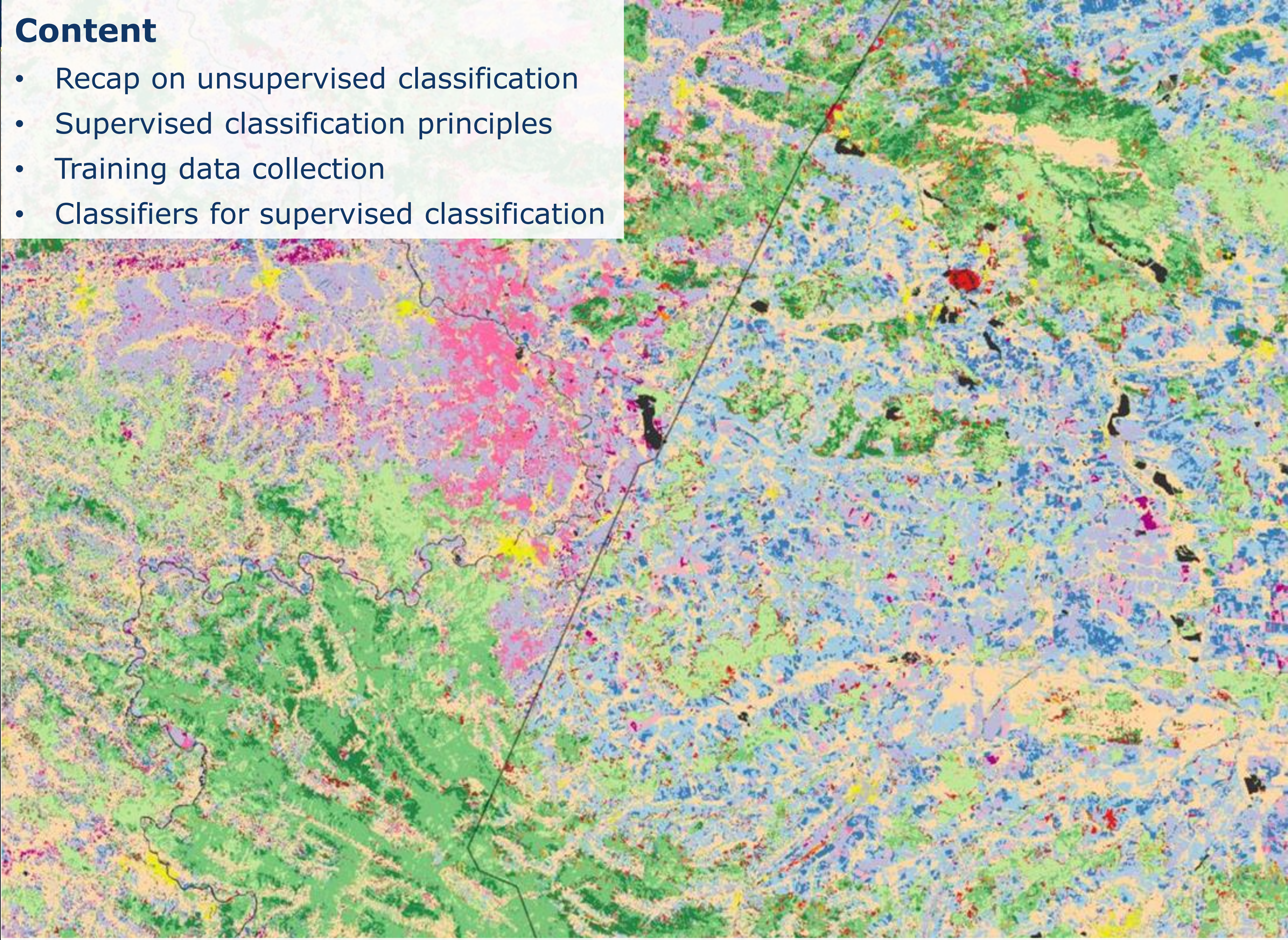
Supervised classification

Patrick Hostert

patrick.hostert@geo.hu-berlin.de
http://www.hu-geomatics.de
Tel.: (030) 2093 – 6905
RUD 16, 2‘226

Content

- Recap on unsupervised classification
- Supervised classification principles
- Training data collection
- Classifiers for supervised classification



Forest and agricultural dynamics between socialist and recent times at the Polish-Ukrainian border. Classified from Landsat time series composites

Recapitulation of last week‵s topics

(1) Read
Horning, N. (2004). Land cover classification methods. (available on Moodle)

(2) Pose questions!

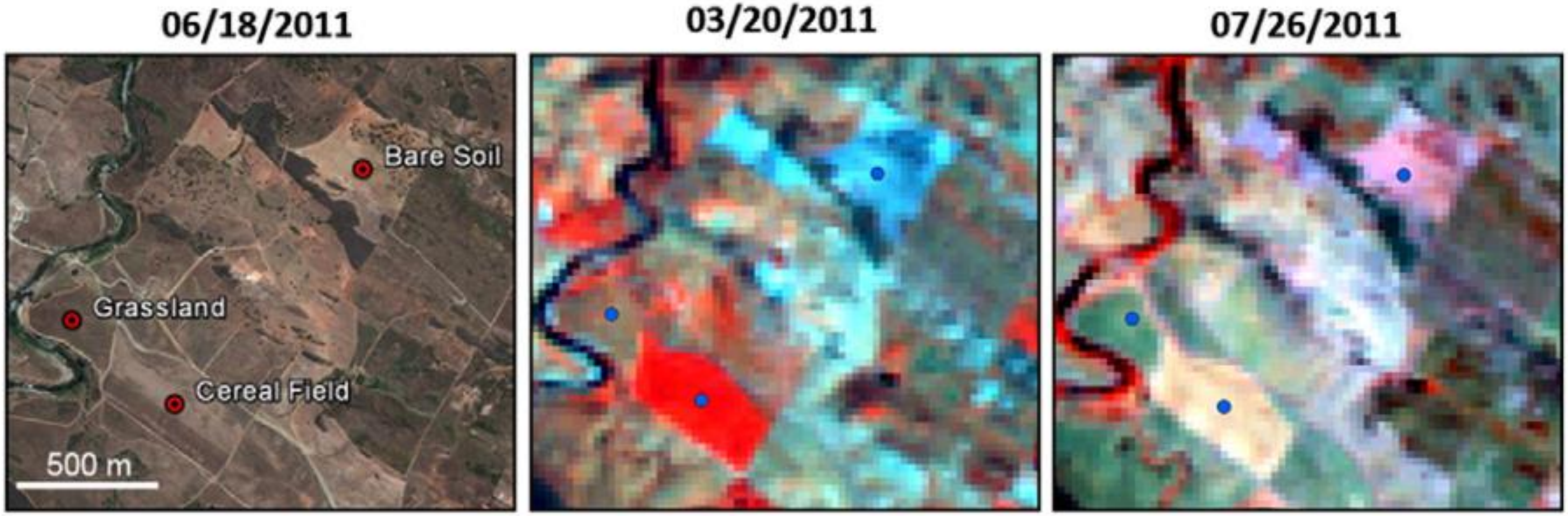
(3) Why is a decision tree a good entry point for understanding image classification?

Supervised image classification principles

Unsupervised classification	Supervised classification
Computational intensive (but nowadays insignificant), adequate for quick ‘snapshot‘: which classes are spectrally well discriminable?	Computational less intensive
Previous knowledge not required (but desirable!)	Knowledge about study area should exist (ground truth, GPS)
Definition of number of classes is critical: If to high – classes not discriminable, if to low – unnecessary merging of classes	Selection of training areas critical: risk of being not representative, not enough, extreme overlap of classes in the feature space
Results are objective (class assignment subjective!)	Results depend on selected training areas selected by the user
Applicable to arbitrary data sets; however, class assignment individually for each data set!	For each new EO-data: possibly new training sites necessary (illumination- and atmospheric conditions, land cover dynamics…)
‚New‘, unknown classes in the study area can be identified	Only already defined classes (based on training sites) can be identified
Accuracy often insufficient	Approved method
Both approaches can be performed with prevalent remote sensing software packages	

Supervised classification principles

- The procedures in supervised classification schemes are always the same, by and large regardless of the respective classifier to be used
- define the objectives of your image classification – what‘s the problem and what does that mean class-wise? Consider regional surface characteristics (incl. phenological behavior depending on image acquisition dates)



Spectral-temporal behavior of agric. lands in S-Portugal in WorldView-2 and Landsat imagery (Senf, C., Leitão, P.J., Pflugmacher, D., van der Linden, S., & Hostert, P. (2015). Mapping land cover in complex Mediterranean landscapes using Landsat: Improved classification accuracies from integrating multi-seasonal and synthetic imagery. *Remote Sensing of Environment, 156, 527-536*

- Fix the number of thematic classes (e.g. land cover / land use classes) and decide in how many spectral classes you may need for each thematic class

Supervised classification principles

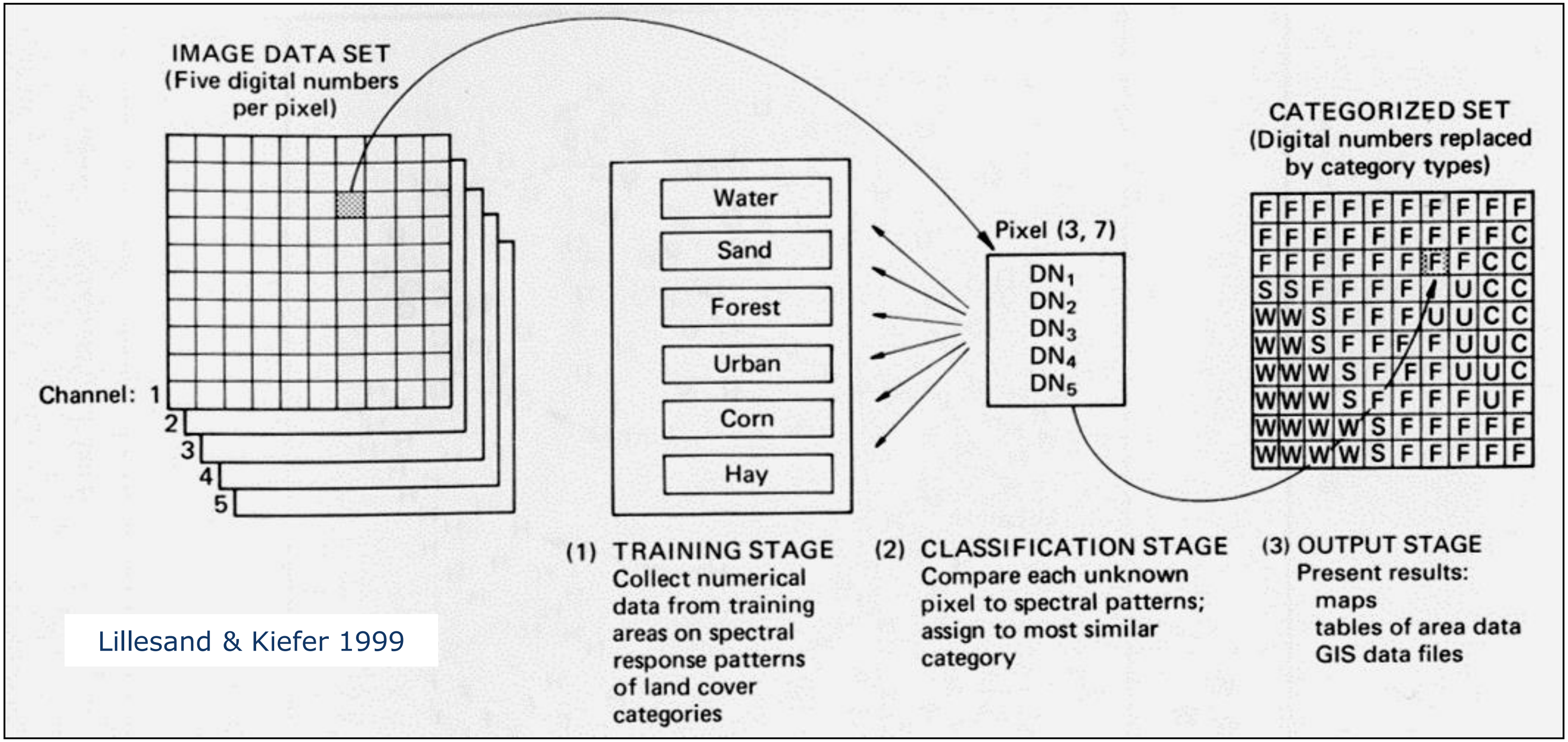
- Define your training data for each spectral class
- Your digitized reference areas will be used to extract spectral signatures
- These class-wise spectral signatures are the training data that inform the chosen classification algorithm
- The classifier compares each of the spectral class signatures from your training data with each pixel in the image and decides on the class assignment
- You finally need to validate your results with independent validation data

Collecting training and validation data during field work in Pará, Amazonia, 2011. Confirmation of deforestation patterns in Landsat imagery in the field.

Collecting training and validation data from on-screen digitizing in Google Earth data

Supervised classification

- In supervised classification, you guide the image processing software to decide how to classify spectral features into classes
- you first extract values from the image that represent classes of interest
- the classifier compares these *training signatures* with every image pixel and assigns the pixel to the best-fitting class



F	F	F	F	F	F	F	F	F	F
F	F	F	F	F	F	F	F	F	C
F	F	F	F	F	F	F	F	C	C
S	S	F	F	F	F		U	C	C
W	W	S	F	F	F	U	U	C	C
W	W	S	F	F	F	F	U	U	C
W	W	W	S	F	F	F	U	U	C
W	W	W	S	F	F	F	F	U	F
W	W	W	W	S	F	F	F	F	F
W	W	W	W	S	F	F	F	F	F

(1) TRAINING STAGE
Collect numerical data from training areas on spectral response patterns of land cover categories

(2) CLASSIFICATION STAGE
Compare each unknown pixel to spectral patterns; assign to most similar category

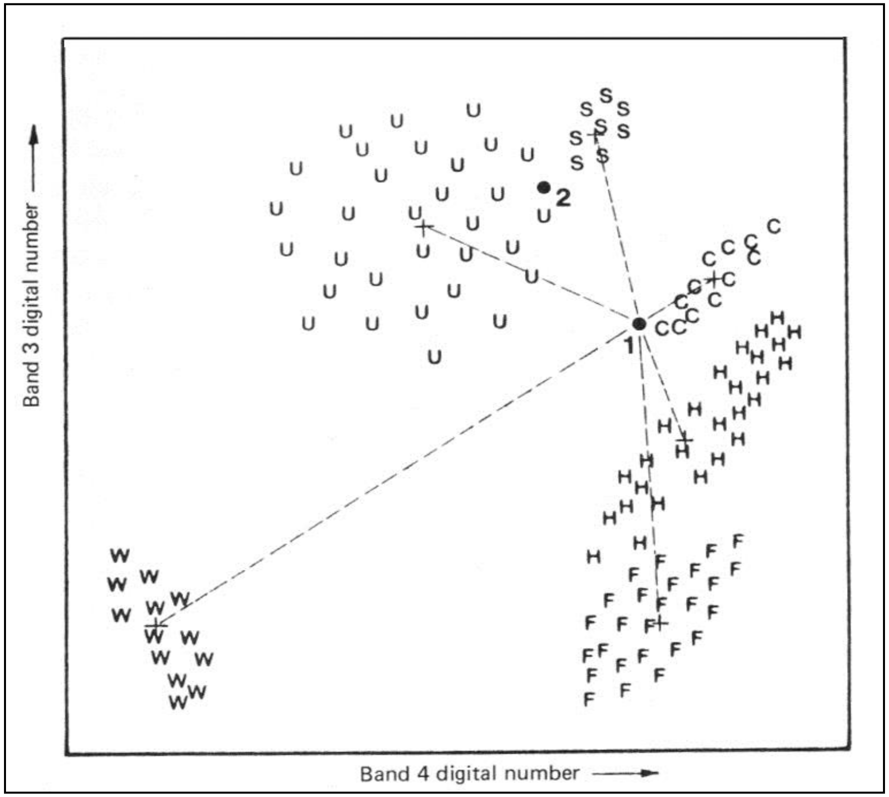
(3) OUTPUT STAGE
Present results:
maps
tables of area data
GIS data files

Lillesand & Kiefer 1999

Supervised classification: parametric vs. non-parametric classifiers

- **Parametric** classifiers rely on assumptions concerning the underlying class distribution and fit a statistical model to the training data (accordingly also termed: statistical classifiers)
- The model is then applied to classify the imagery

- Underlying assumption: we can statistically describe a prototype class distribution (e.g. of class "urban – U") in a given spectral feature space
- the classification algorithm then uses a specific (set of) rule(s) to assign each image pixel (e.g. pixel 2) to the appropriate class
- Common parameters used to describe classes are e.g. class mean or class standard deviation

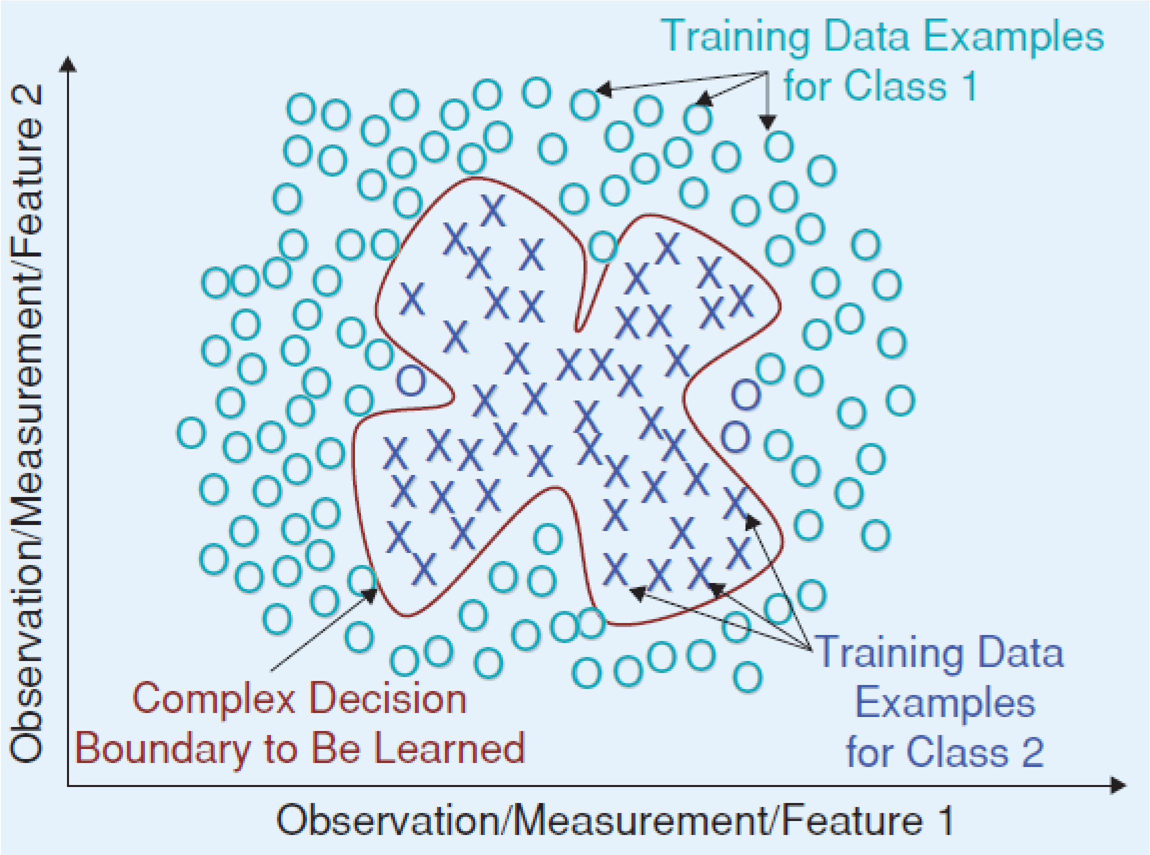


Lillesand & Kiefer 1999

Supervised classification: parametric vs. non-parametric classifiers

- **Nonparametric** methods, like classification and regression trees, use other means than statistical distributions to determine the class assignment

- Such methods are specifically useful when describing complex classes
- "Complex" may entail class distributions in spectral feature space that are hard to describe statistically…
- …or classes where one class of interest includes many spectral sub-classes (e.g. in urban environments)

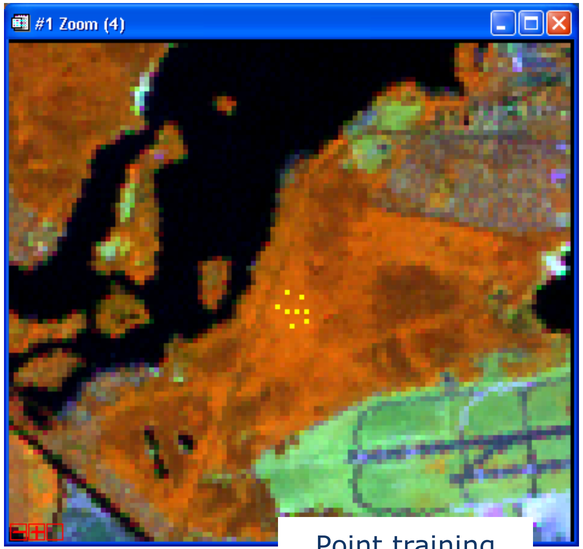

Training Data Examples for Class 1
Complex Decision Boundary to Be Learned
Training Data Examples for Class 2
Observation/Measurement/Feature 2
Observation/Measurement/Feature 1

Polikar 2006

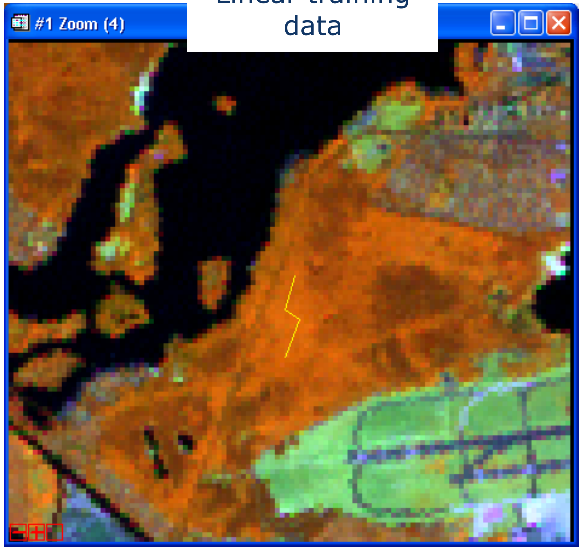
Training data collection

Reference signatures (training signatures) from areas of interest

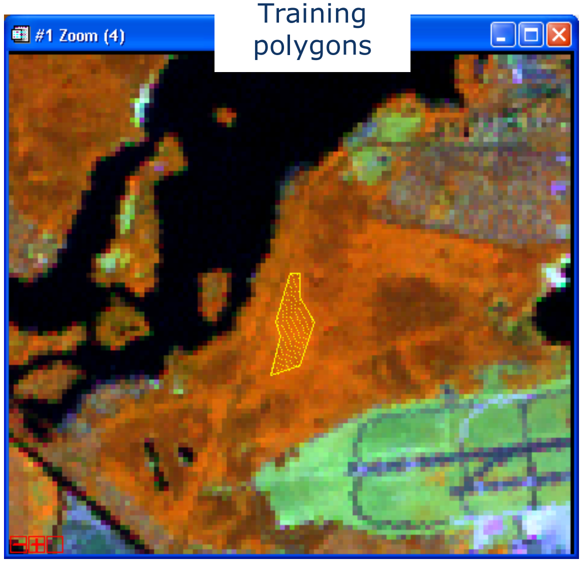
- Reference signatures should cover all relevant spectral classes in spectral feature space
- Spatially, we differentiate training data derived from digitized points, lines (not common) or polygons
- Parametric (statistical) classifiers usually need spectral signatures derived based on polygons, as they e.g. need information on class means or standard deviation
- Nonparametric classifiers can usually be trained based on training signatures from individual pixels



Point training data



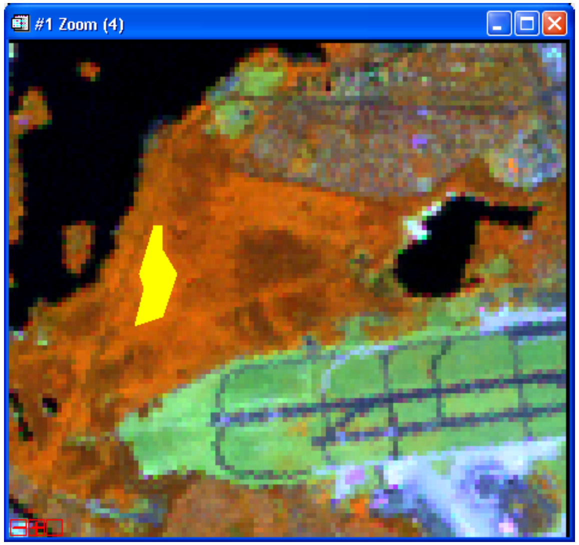
Linear training data



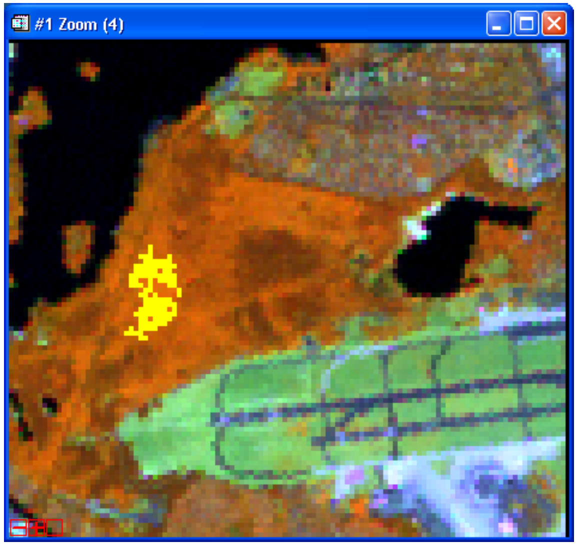
Training polygons

Reference signatures (training signatures) from areas of interest

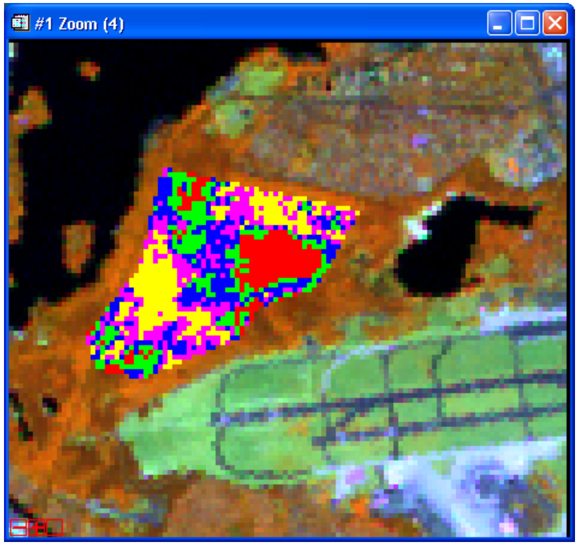
- Different methods exist to derive a training signature from polygons
- Example: digitizing a sample area-of-interest (often abbreviated „AOI") to derive a spectral signature for deciduous forest



interactive digitizing of homogeneous areas on-screen (alternative: import from maps or field mapping)



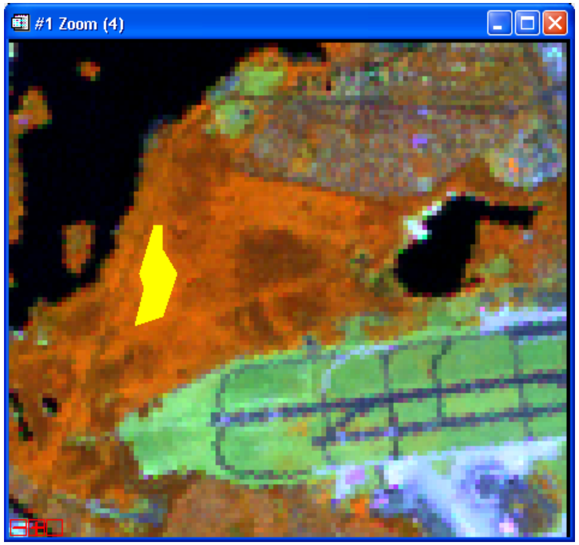
definition of seed-clusters and max. standard deviation for neighborhood similarities; then growing neighborhoods



digitizing of heterogeneous initial polygons; then clustering; then selection of appropriate training areas for supervised classification

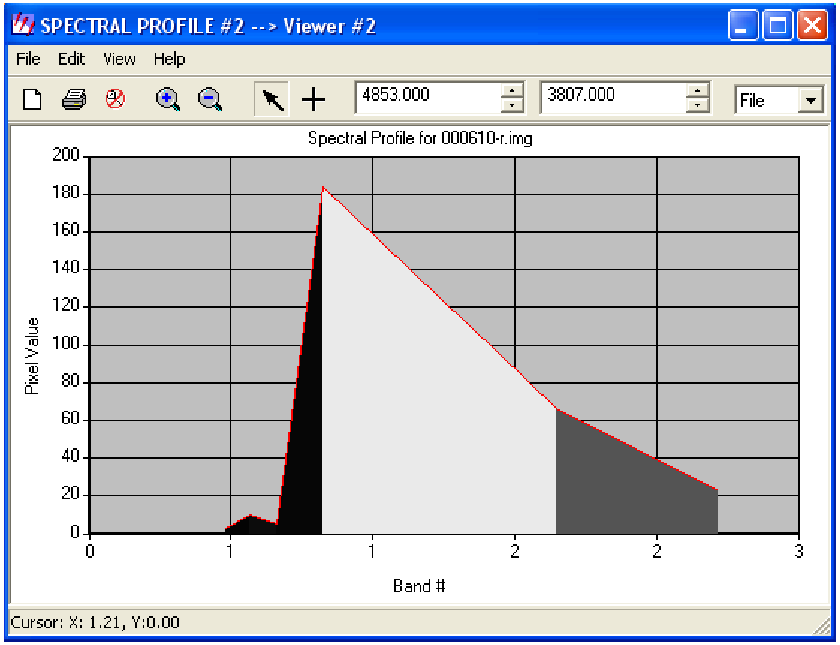
Reference signatures (training signatures)

- Reference signatures are then used to define reference classes in multispectral feature space
- In case of a polygon-based training strategy, the values of all pixels in the AOI are extracted from the underlying image to statistically describe the spectral class



Area of interest (left, yellow polygon) that defines a deciduous forest area in Landsat TM/ETM+ 4-5-3 (RGB) image on-screen

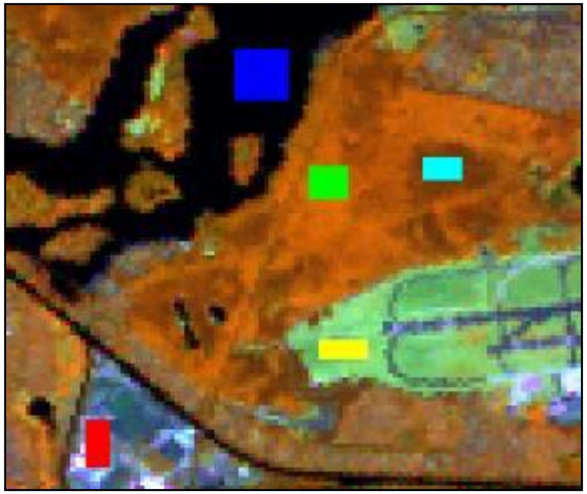
Mean spectrum in Landsat TM/ETM+ bands 1-5 and 7 derived from extracting all pixels in the area of interest (right)



- The characteristics of the reference signature are determined by the characteristics of the underlying image (e.g. 6 bands from Landsat-TM data excl. the thermal band)

Reference signatures (training signatures)

- Each signature extracted from a digitized AOI is then used to extract the statistics needed for a chosen classifier
- Spectral signatures contain information from each spectral band in the input image‘s spectral feature space



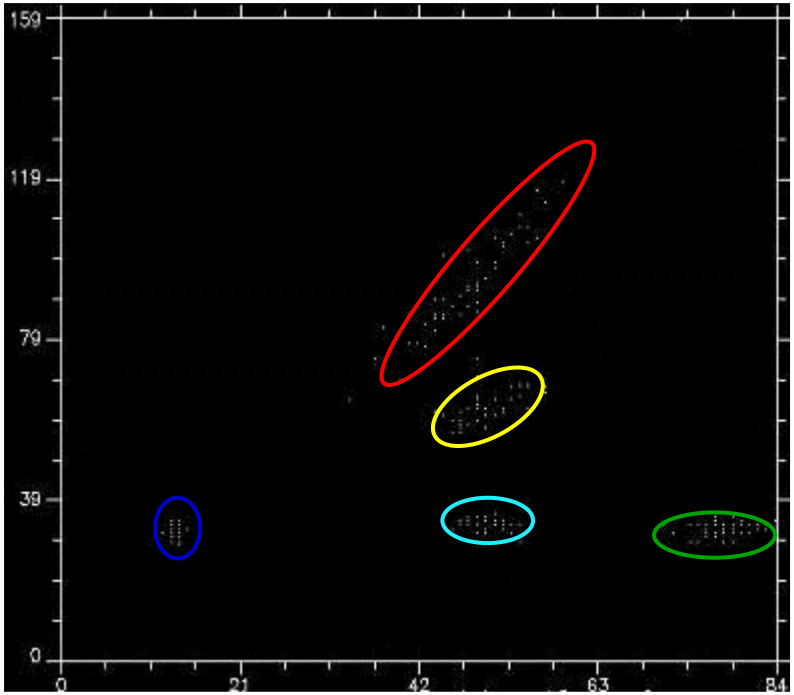
Blue: water

Red: built-up

Green: deciduous forest

Cyan: coniferous forest

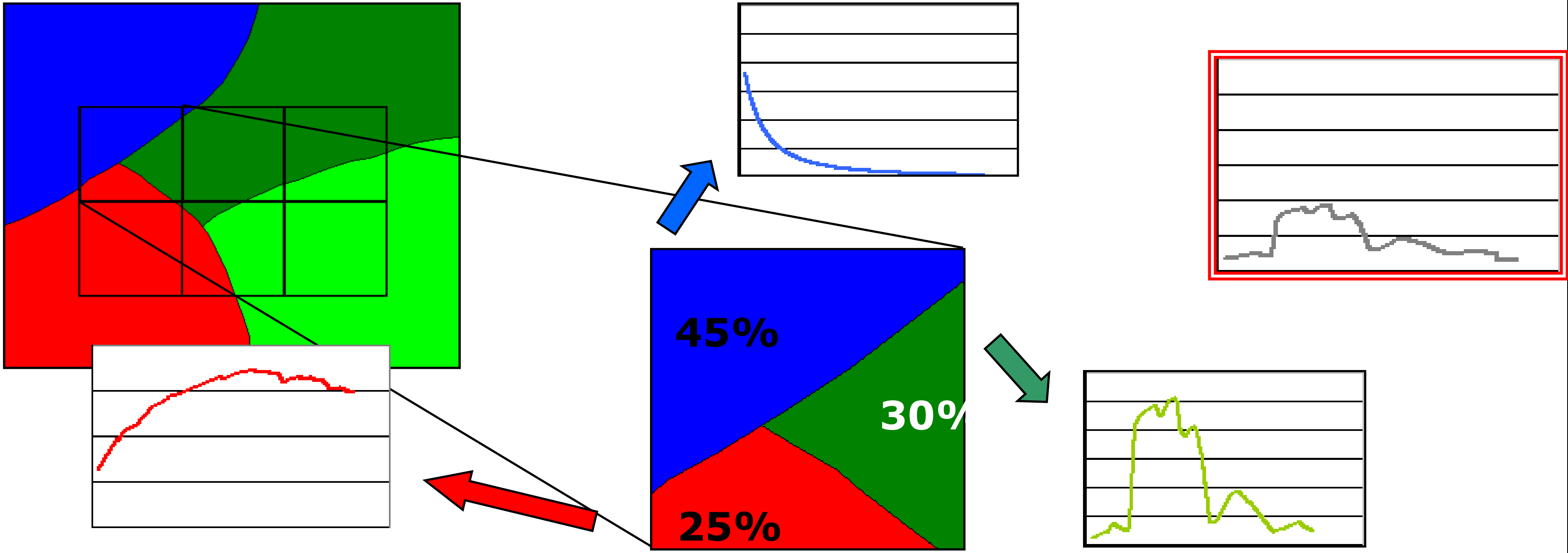
Yellow: grass



Exemplary signatures from an area around Tegel Airport in Berlin transferred into a 2-dimensional scatterplot (Landsat red band versus near infrared band)

The mixed-pixel problem

- Pixel size and image objects often don't match, i.e. with coarser pixels, the number of so called "mixed pixels" increases
- Mixed pixels lead to mixed spectral signatures that are hard to classify correctly



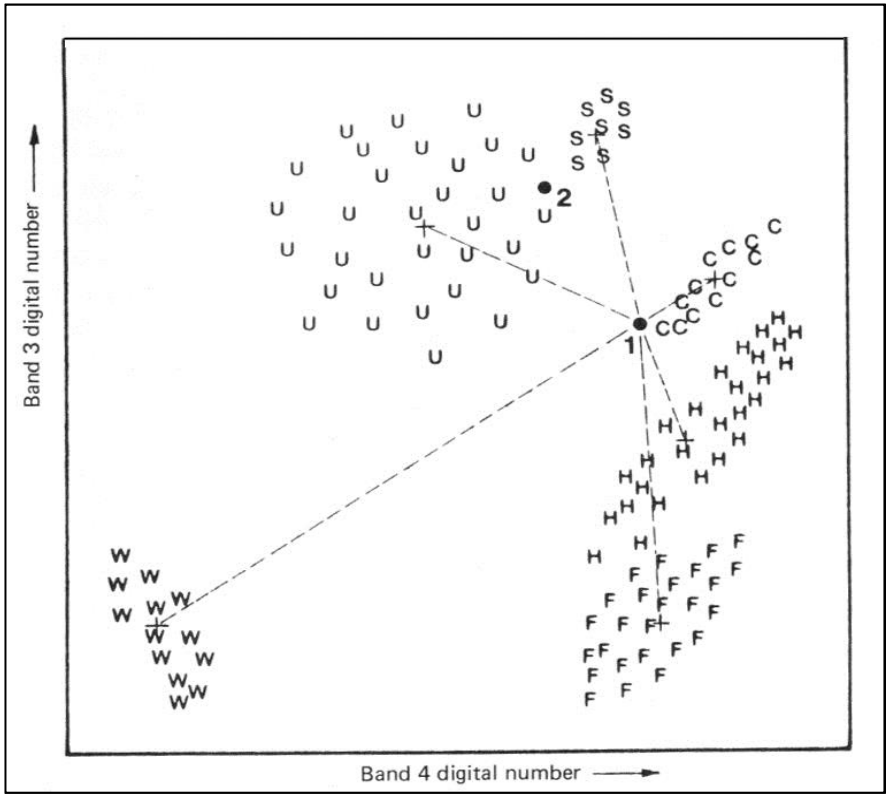
- Urban environments are a typical example of fine scale features resulting in many mixed pixels

Classifiers for supervised image classification:

Parametric (or statistical) classifiers

A simple parametric classifier: Minimum Distance

- the Minimum Distance Classifier (or Minimum Distance to Means Classifier) only considers cluster means in multispectral feature space



- each image pixel is assigned to a class solely on its spectral distance to a training cluster mean value

- euclidian distance in spectral feature space is calculated to define that distance:

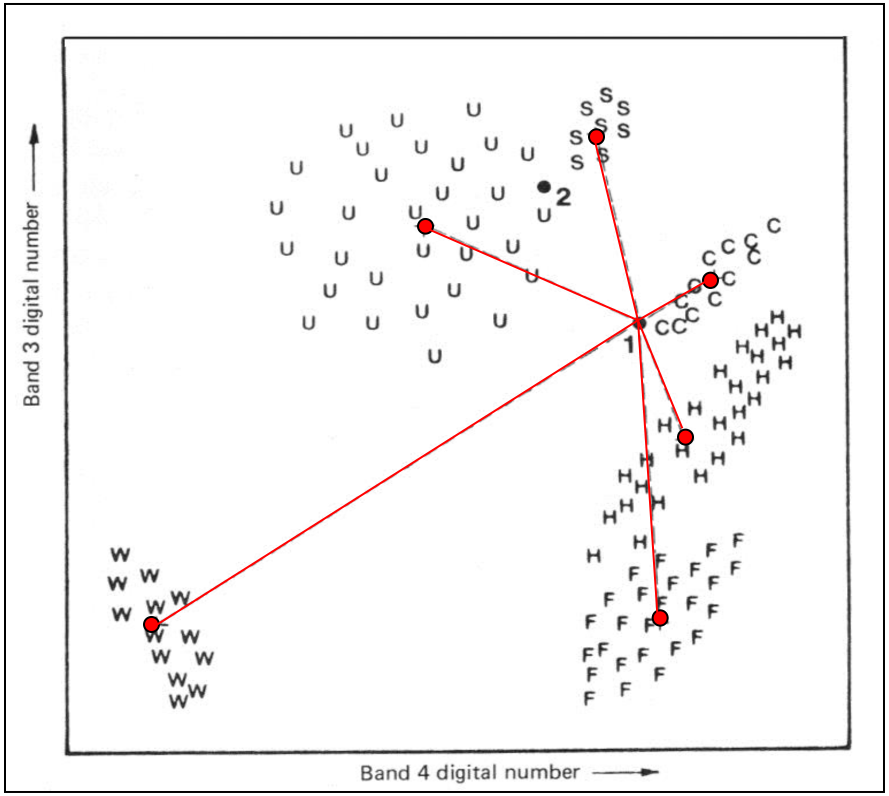
$$d = \sqrt{\sum_{b=1}^{n} (x_b - \mu_b)^2}$$

n: number of bands

Lillesand & Kiefer 1999

Parametric: Minimum Distance Classifier

- Minimum-Distance based image classification hence requires the following steps:



- calculating spectral mean m_i for each reference class k_i

- distance calculation d_i of a pixel x to all m_i

$$d_i = d(\mu_i, x)$$

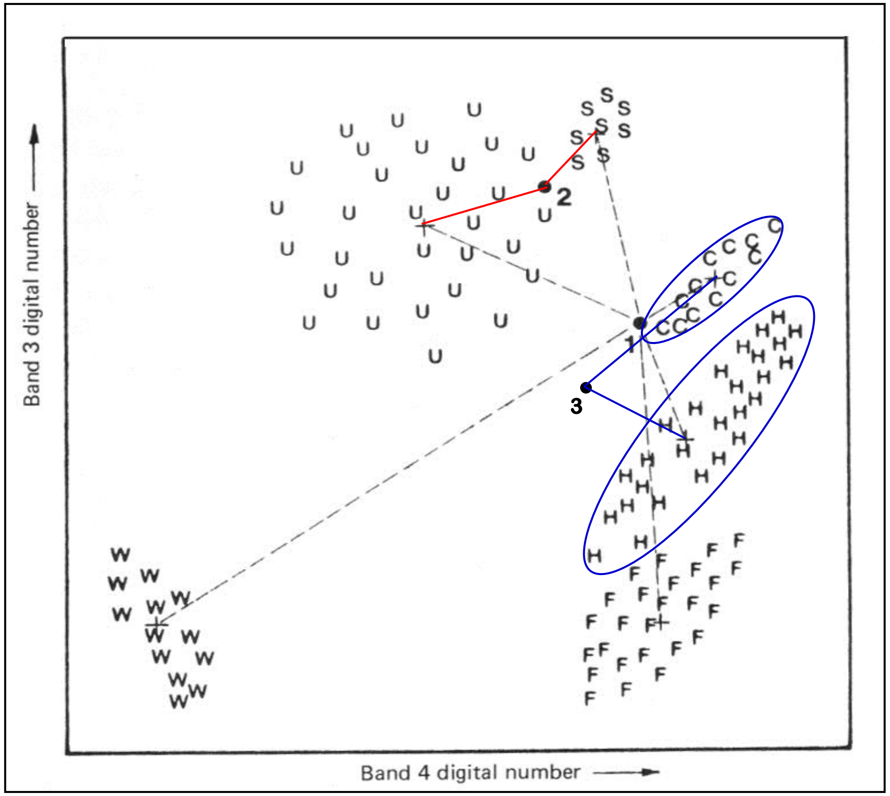
i = 1…t,
t: number of reference classes ω

- pixel x becomes a member of class ω_j if: $d_j < d_i$ für alle $i \neq j$

Lillesand & Kiefer 1999

Parametric: Minimum Distance Classifier

- the Minimum Distance Classifier is a robust, parametric classification algorithm

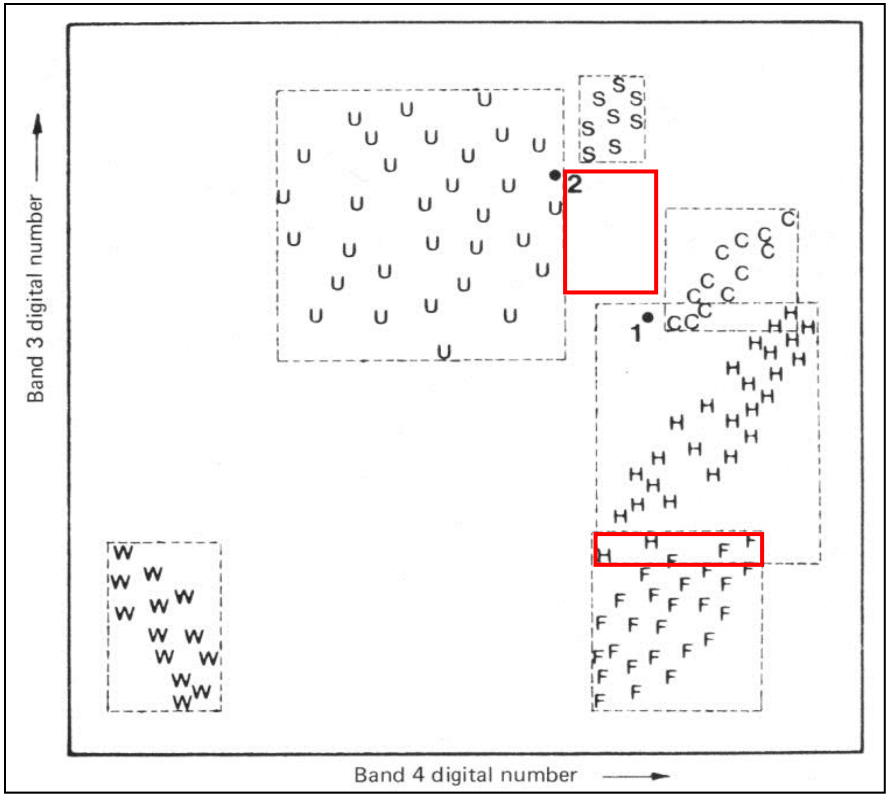


- quick calculation, even for large data sets and many classes
- limited in case of reference classes with highly varying variance or strong collinearities between spectral bands
- generally: limited if class mean is not representative for a class

Lillesand & Kiefer 1999

Another simple parametric classifier: Parallelepiped

- 2-D: rectangle; 3-D: box; > 4-D: parallelepiped

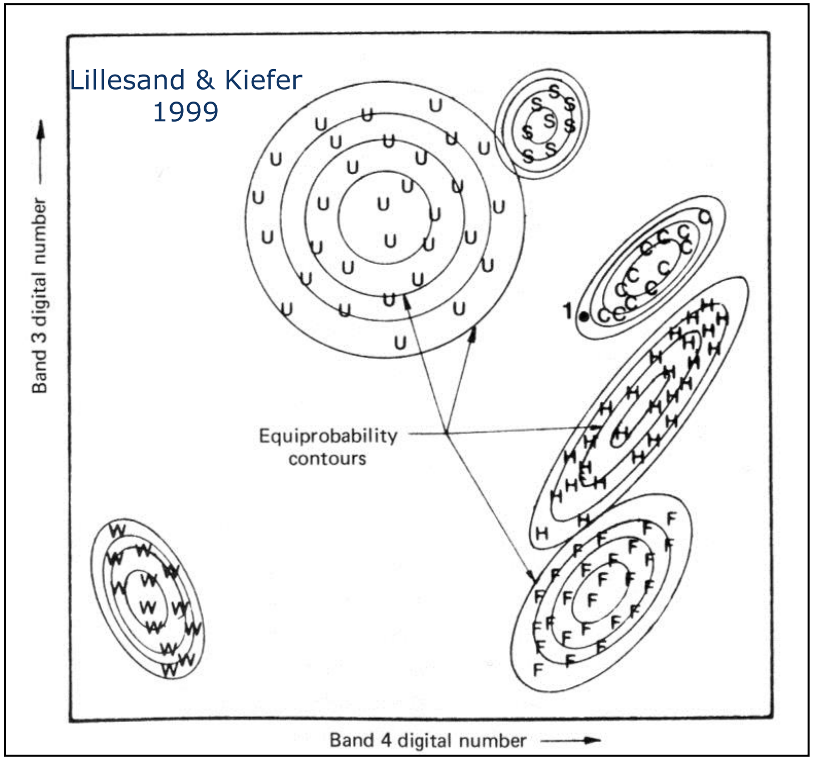


Lillesand & Kiefer 1999

- Sometimes also referred to as „Box-Classifier"
- Upper and lower boundaries of class-wise spectral sample distributions define the limits for pixel assignment to classes
- Boundaries may be defined by the spectral ranges of training pixels in each spectral band or by a given standard deviation from the band-wise spectral mean
- Potential problem: gaps and overlaps between parallelepipeds

A more advanced parametric classifier: Maximum-Likelihood

- As spectral signatures for different classes often a) overlap and b) considerably vary in their variance, it may make sense to include each classes variance (or standard deviation) and in-between class variance (co-variance)

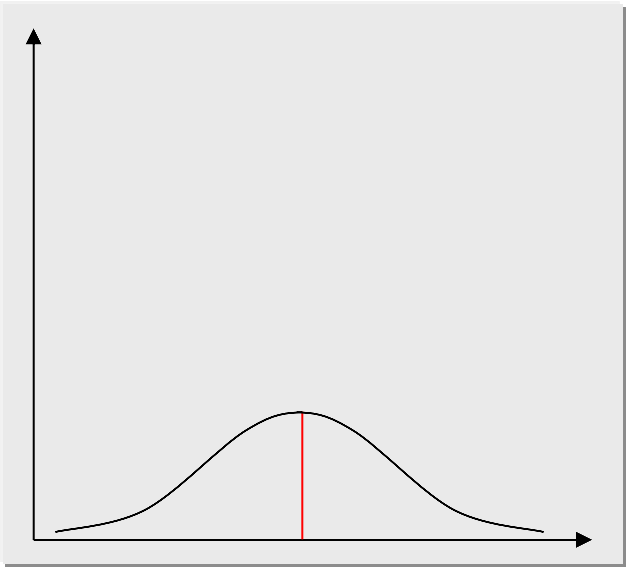


Lillesand & Kiefer 1999

- Maximum-likelihood classification handles variance and covariance based on so-called „probability densitiy functions“
- Assigning a pixel to a class is based on the multispectral class mean and the variance-covariance matrix of all spectral class definitions
- Isolines of equal assignment probabilities are calculated („equiprobability contours“)
- Assignment of each image pixel to a class is based on the maximum-likelihood of being a class member considering the equiprobability contours

Parametric: Maximum-Likelihood Classifier

- Assignment probabilities are usually calculated assuming a class-wise Gaussian normal distribution
- Assumption: symmetric distribution of spectral values of a training pixel sample around the mean (which is also the assumed to represent the median and mode
- The variance determines the form of the distribution function around the mean
- In the hypothetical case of one spectral band only, the distribution resembles a bell-shaped curve with the following parameterization:



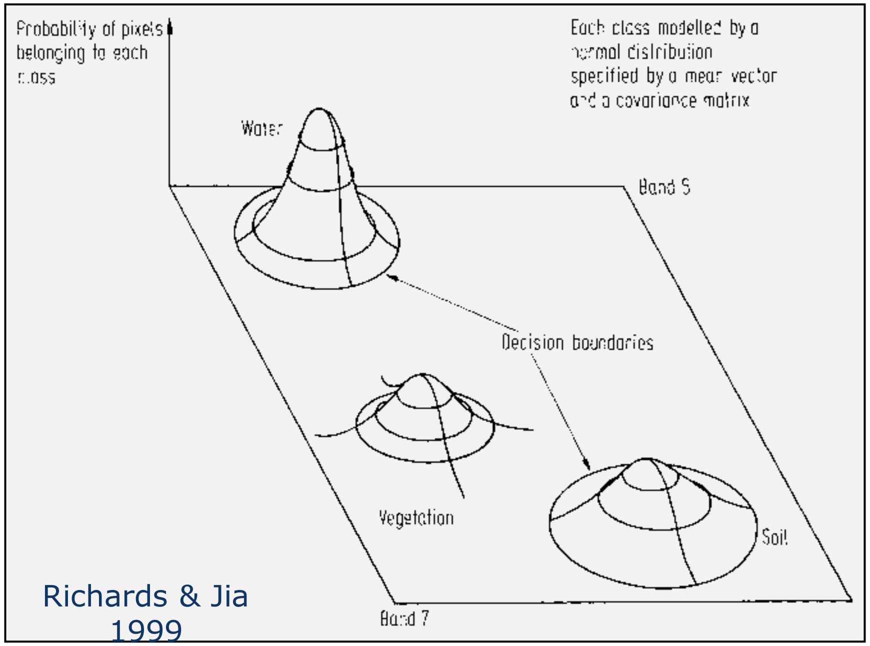
$$p(x \mid \omega_i) = \frac{1}{\sqrt{(2\pi)\sigma_i}} * e^{\sqrt{((x-\mu_i)/\sigma_i)^2}}$$

with	$p(x\|\omega_i)$:	probabiilty of $x \in \omega_i$
	x:	spectral pixel value
	ω_i:	spectral class i (i = 1,…,M; M = no of classes)
	μ_i:	mean of reference class i
	σ_i:	variance of reference class i (describing the data distribution around the mean)

Parametric: Maximum-Likelihood Classifier

- In a multidimensional case, as given for multispectral image data, the probabilities need to be calculated accordingly

- Instead of mean and variance, the mean vector across all bands and the variance-covariance-matrix will be used to define class assignments based on maximum-likelihood


Probability of pixels belonging to each class
Each class modelled by a normal distribution specified by a mean vector and a covariance matrix
Water
Band 5
Decision boundaries
Vegetation
Soil
Band 7

Richards & Jia 1999

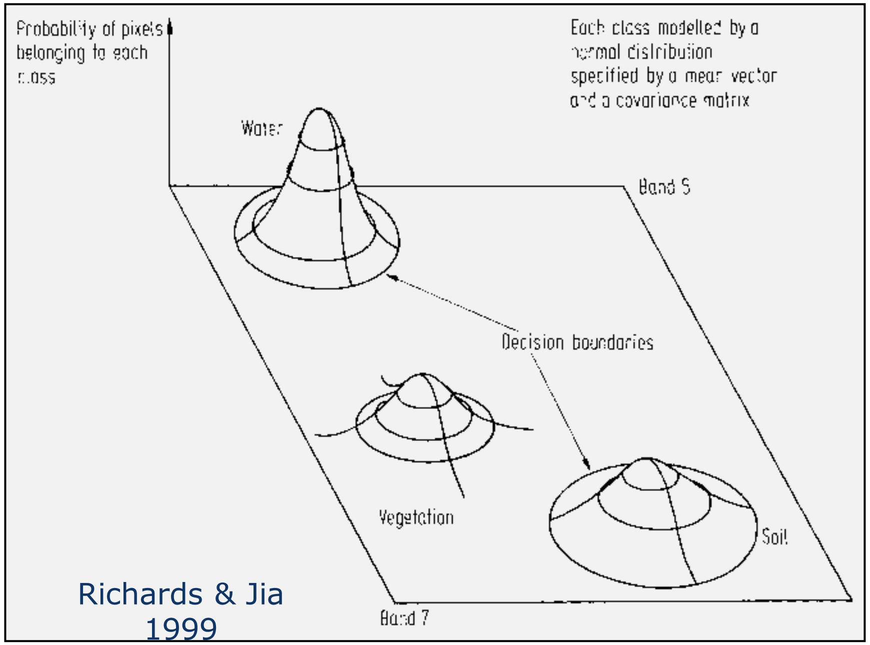
$$p(x\,|\,\omega_i) = \frac{1}{(2\pi)^{N/2}\left|\Sigma_i\right|^{1/2}} * e^{-1/2\;(x\,-\,m_i)^t\,\Sigma_i^{-1}(x\,-\,m_i)}$$

with	$p(x\|\omega_i)$:	probability of $x \in \omega_i$
	x:	multispectral vector of band-wise pixel values
	ω_i:	spectral class i (i = 1,…,M; M = no of classes)
	m_i:	mean vector for reference class i
	Σ_i:	covariance matrix for reference class i (multispectral measure of distribution)
	N:	no of bands

Parametric: Maximum-Likelihood Classifier

- Once we know the likelihood of assignment for a pixel across all spectral classes the assignment is based on the maximum likelihood

- We use the following decision rule to assigne a pixel to it‘s most likely class


Probability of pixels belonging to each class
Each class modelled by a normal distribution specified by a mean vector and a covariance matrix
Water
Band 5
Decision boundaries
Vegetation
Soil
Band 7

Richards & Jia 1999

$$x \in \omega_i \quad \text{if}$$

$$p(x\,|\,\omega_i) * p(\omega_i) \geq p(x\,|\,\omega_j) * p(\omega_j)$$

with j = 1…M,
M: no of reference classes
$p(\omega_i)$: a-priori-probability of reference class ω_i

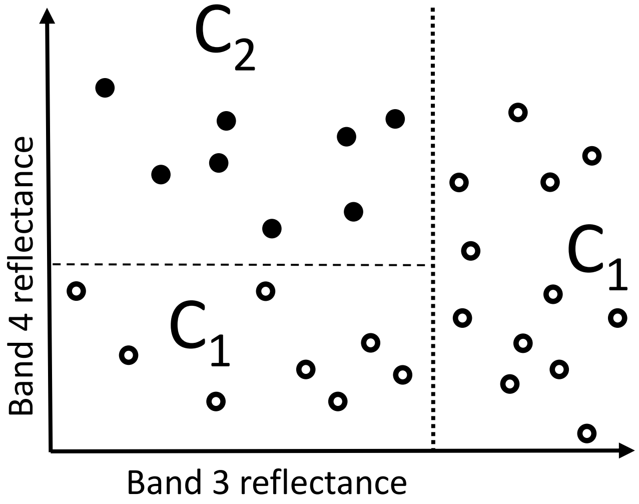
- $p(\omega_i)$, the so-called „a-priori-probability“, allows inferring knowledge on probabilities of class occurrence (e.g. from existing maps or previous classifications)

Classifiers for supervised image classification:

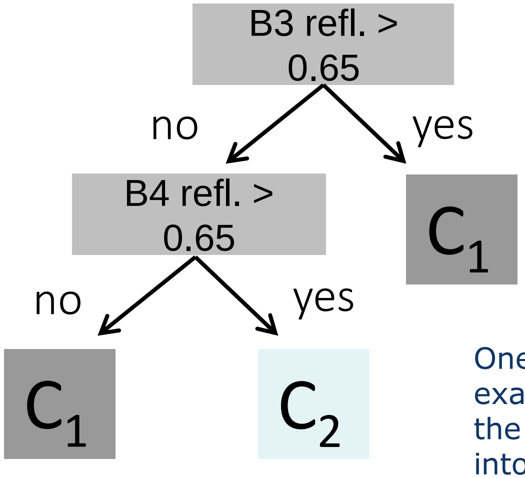
Nonparametric classifiers

Self-learning decision trees

- Decision Trees can be created interactively (you design the rules for splitting a dataset into classes) or in a self-learning fashion



2-band reflectance feature space with 2 classes to separate

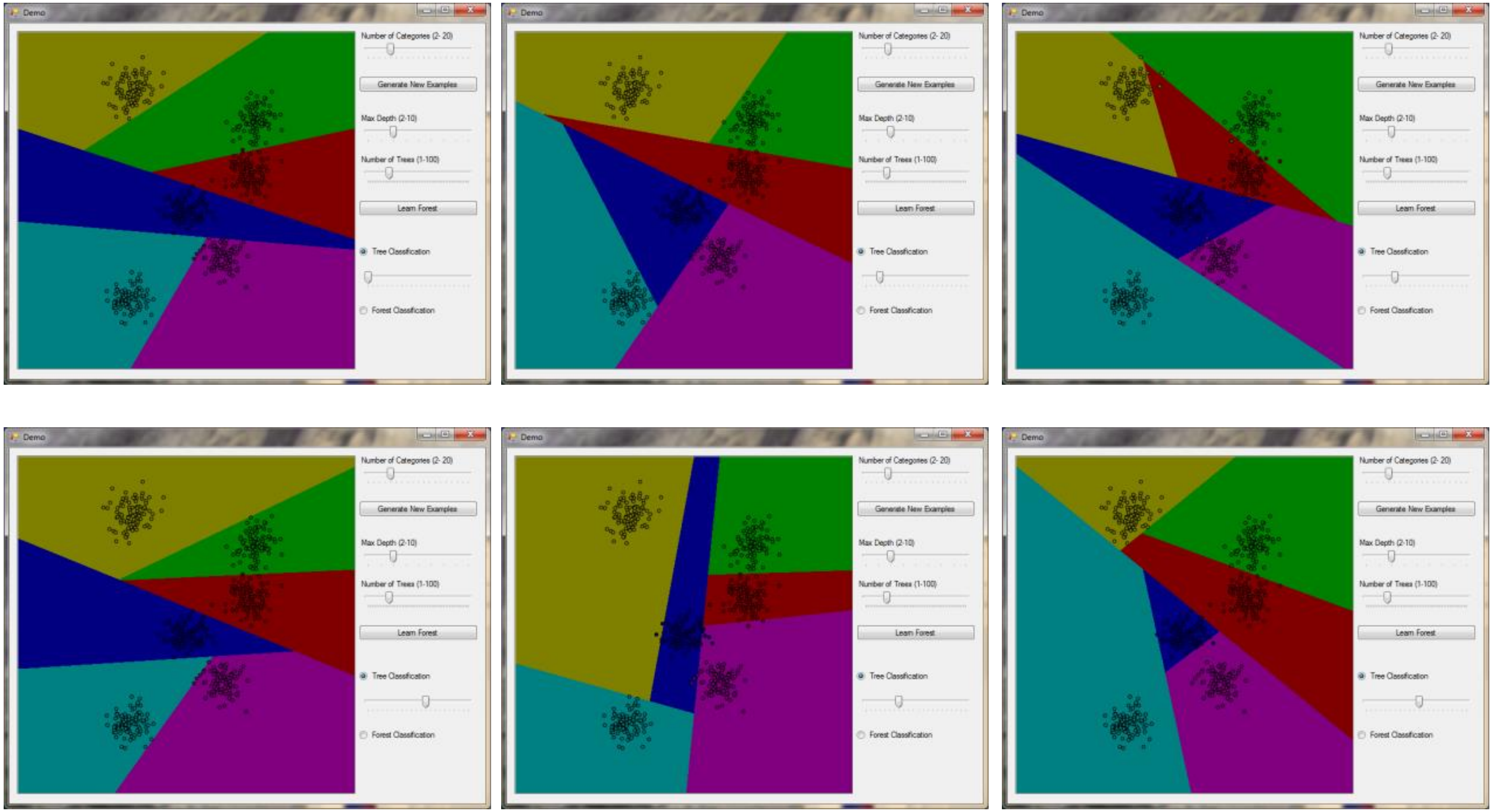


One decision tree example to split the training data into 2 classes

- Rules for subdividing a dataset into homogeneous classes are automatically derived from training data based on simple decisions (e.g. reflectance thresholds)
- Self-learning DTs split the feature space into sub-classes based on a purity measure that minimizes class heterogeneity
- The training result is then saved as a ruleset and applied to the to be classified dataset(s)

Self-learning decision trees

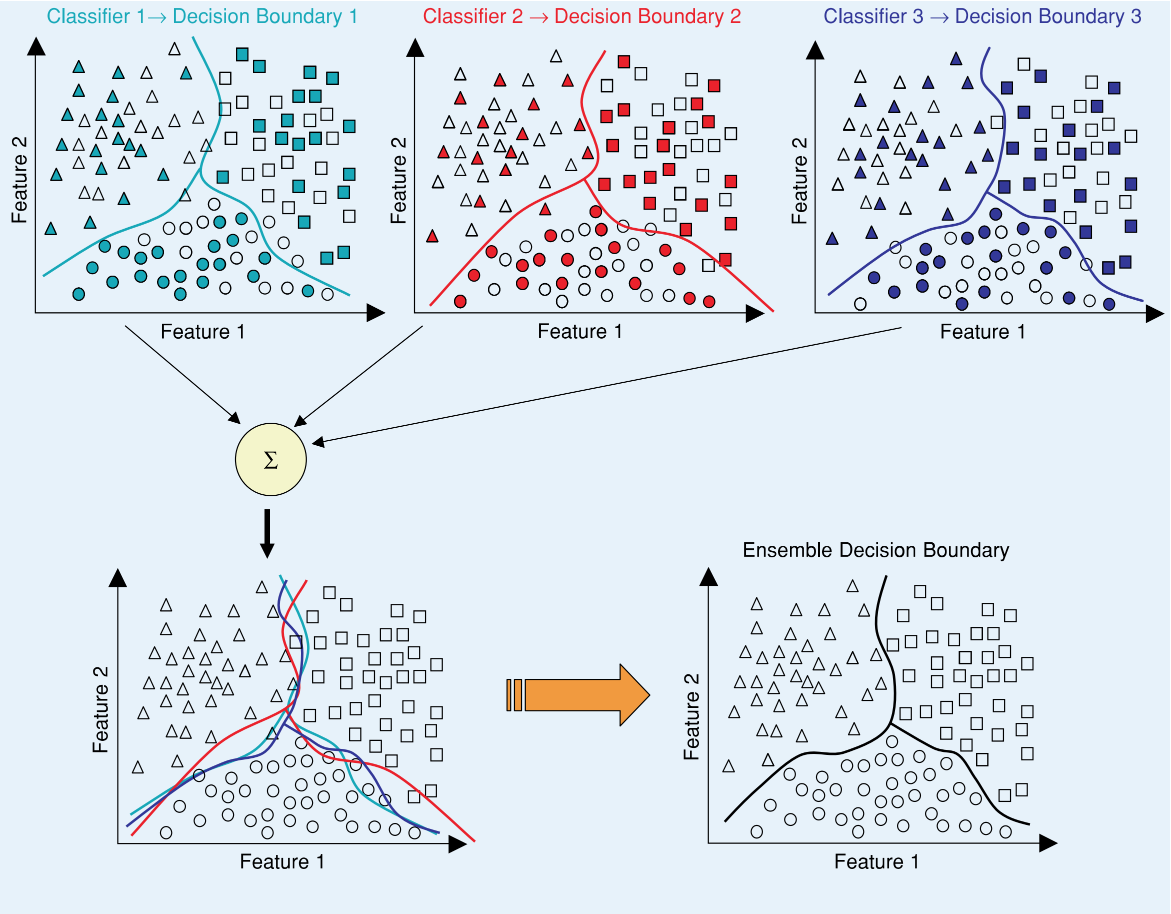
- Many different splits are possible based on the same training data set
- Using not just a 2-band feature space, hundreds or thousands of rulesets could be derived



2-band reflectance feature space with many different splits for a 6-class classification problem

Ensemble classifiers

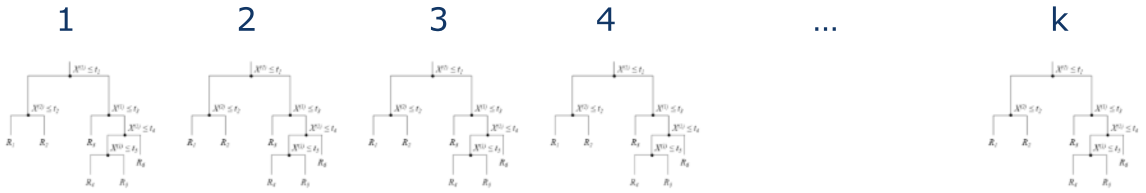
- DTs tend to poorly solve more complex classification problems (e.g. in urban environments) that need detailed rules
- Rule sets for complex classes create over-fitting: rules do not generalize well from training data to the full dataset
- To avoid over-fitting, many „weak learners“ are used instead of the perfect classification rule set: ensemble classifier
- Assumption: Ensembles from 100s or 1,000s of uncertain classifications create accurate class boundaries when combined


Classifier 1→ Decision Boundary 1
Classifier 2 → Decision Boundary 2
Classifier 3 → Decision Boundary 3
Feature 2
Feature 1
Σ
Ensemble Decision Boundary

Polikar 2006

Random Forests

- Random Forests are ensembles self-learning DTs, where each DT is randomized. The idea behind is that many weak learners can come to one strong decision (Breimann 2001)



- For each of *k* trees, a random subset of *n* training samples (e.g. from digitized training pixel spectra) is selected to create a DT
- all samples (incl. the already used ones) are then used again during further DT generation: bootstrap aggregating (short: bagging)
- Then a random feature subset (e.g. bands of a satellite image) is selected at each node of each tree to create decision rules
- The unused training samples (complement of n) is used to estimate each tree's classification error, leading to an aggregated *Out-Of-Bag* (OOB) error
- The class assignment of each image pixel is based on the majority vote across the k classifiers created at the training stage

Random Forests

- Parameters that need to be set are:

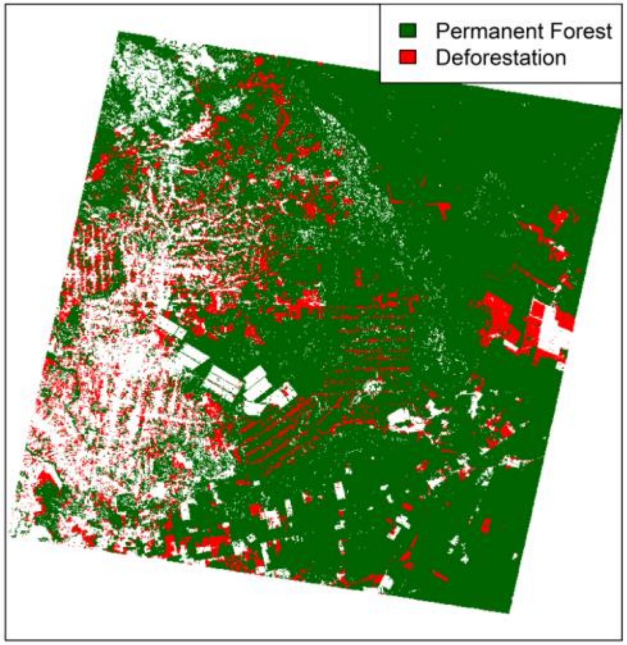
  - Number of trees: Should be large enough that each case is included at least once

  - Size of bagged sample (size of n): Generally has not much influence on the outcome, but depends on the number of trees. Can be important for imbalanced training data.

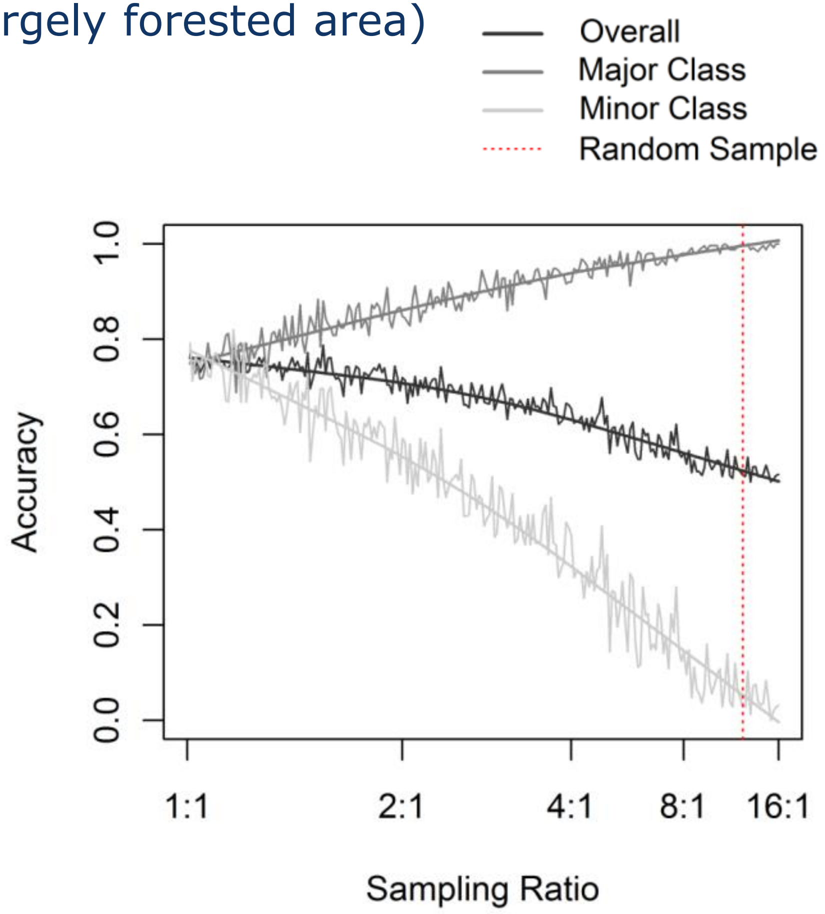
  - Number of features for creating a split in a tree: a rule-of-thumb value is sqrt(no. of all available features) for RF classification

Random Forests

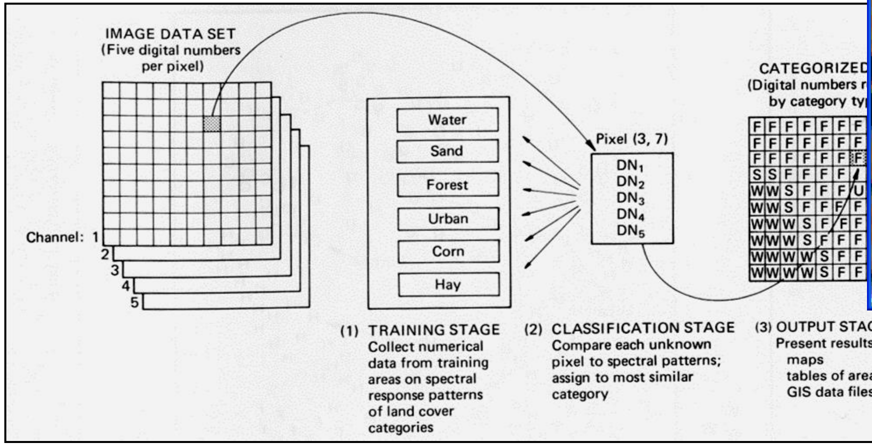
- Imbalanced training data means that one class is more frequent than the other/others
- Imbalanced training data can bias RF classifiers
- The minor class (e.g. deforestation in a largely forested area) will be underestimated



The class ratio is close to 1:16



Summary

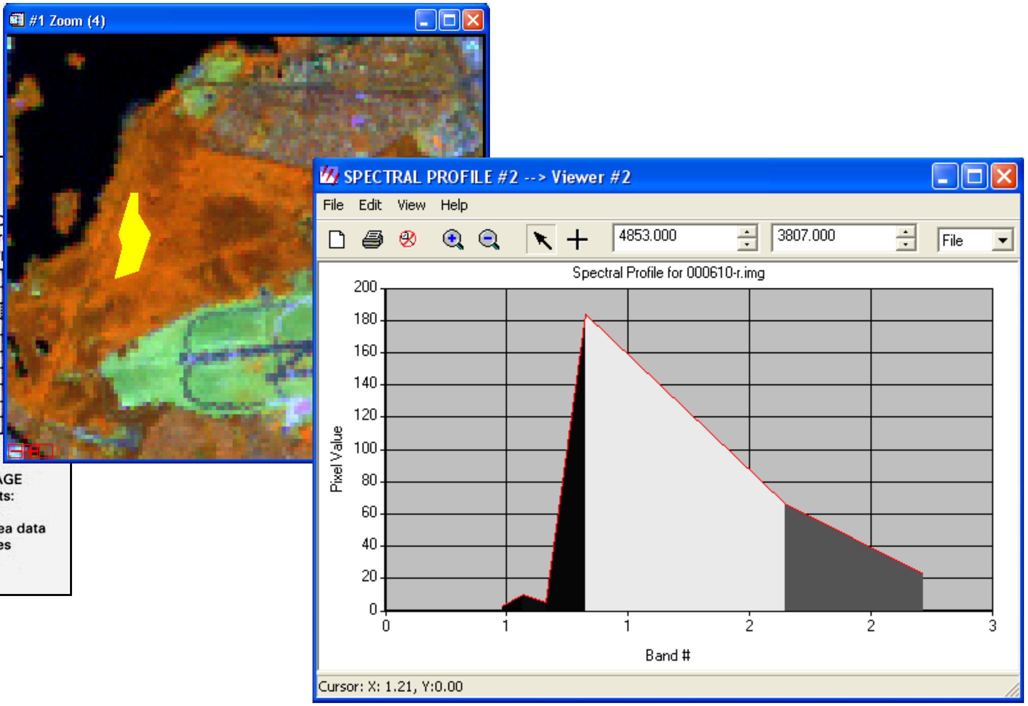


F F F F F F F
F F F F F F F
F F F F F F F
S S F F F F
W W S F F F U
W W S F F F F
W W W S F F F
W W W S F F F
W W W W S F F
W W W W S F F

(1) TRAINING STAGE
Collect numerical data from training areas on spectral response patterns of land cover categories

(2) CLASSIFICATION STAGE
Compare each unknown pixel to spectral patterns; assign to most similar category

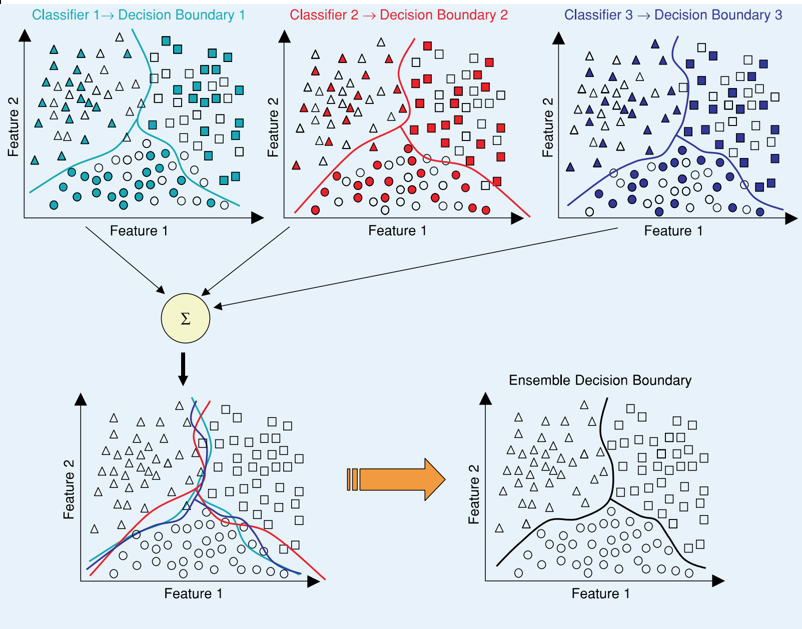
(3) OUTPUT STAGE
Present results:
maps
tables of area data
GIS data files

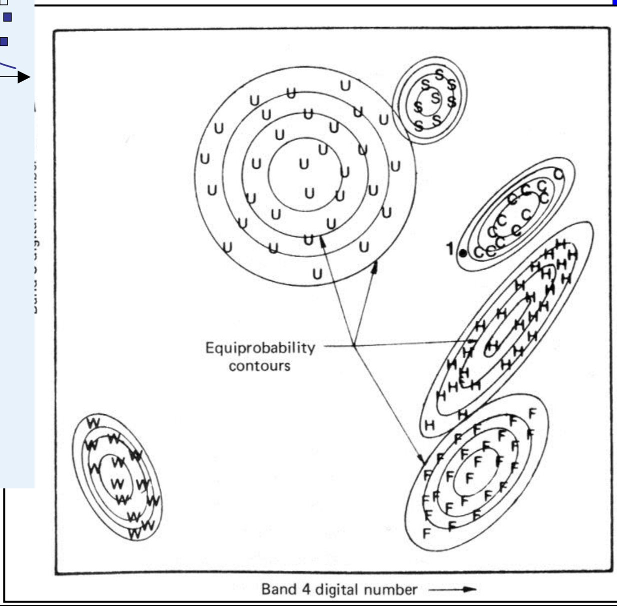


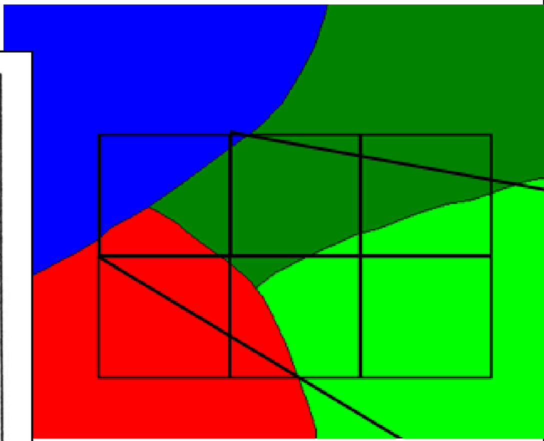
4853.000

3807.000

Cursor: X: 1.21, Y:0.00


Classifier 1 → Decision Boundary 1
Classifier 2 → Decision Boundary 2
Classifier 3 → Decision Boundary 3
Feature 2
Feature 1
Feature 2
Feature 1
Feature 2
Feature 1
Σ
Feature 2
Feature 1
Ensemble Decision Boundary
Feature 2
Feature 1





Recapitulation for next session

(1) Read

Horning, N. (2010). Random Forests: An algorithm for image classification and generation of continuous fields data sets. In, *Proceeding of International Conference on Geoinformatics for Spatial Infrastructure Development in Earth and Allied Sciences (pp. 9-11).* Provided in Moodle

(2) What is strange in fig. 1 of the paper? What‵s the reason?

Outlook

Next week we will focus on:

Change detection

Thanks for your attention!