Parallel & Scalable Machine Learning

Introduction to Machine Learning Algorithms

Dr. –Ing. Gabriele Cavallaro
Postdoctoral Researcher
High Productivity Data Processing Group
Juelich Supercomputing Centre

Lecture 7 - 18/02/2020

PATTERN RECOGNITION SYSTEMS

COURSE OUTLINE

- Parallel and Scalable Machine Learning Driven by HPC
- Introduction to Machine Learning Fundamentals
- Supervised Learning with a Simple Learning Model
- Artificial Neural Networks (ANNs)
- Introduction to Statistical Learning Theory
- Validation and Regularization
- Pattern Recognition Systems
- Parallel and Distributed Training of ANN
- Supervised Learning with Deep Learning
- Unsupervised Learning – Clustering
- Clustering with HPC
- Introduction to Deep Reinforcement Learning

OUTLINE

- Introduction to Remote Sensing Application

- Pattern Recognition System

  - Feature Extraction

  - Feature Selection

- Good Practice with data

  - Sampling methods

REMOTE SENSING

Remote (without physical contact) Sensing (measurement of information)

- Obtain information about the *atmosphere* and *surface* of Earth **without needing to be in contact with it**

- Achieved by **sensing** and recording **emitted or reflected energy**
  - Toward processing and interpreting the retrieved information for decision-making

- Measurement of **radiation of different wavelengths** reflected or emitted from distant objects or materials

- They may be categorized by **class/type**, substance, and spatial distribution
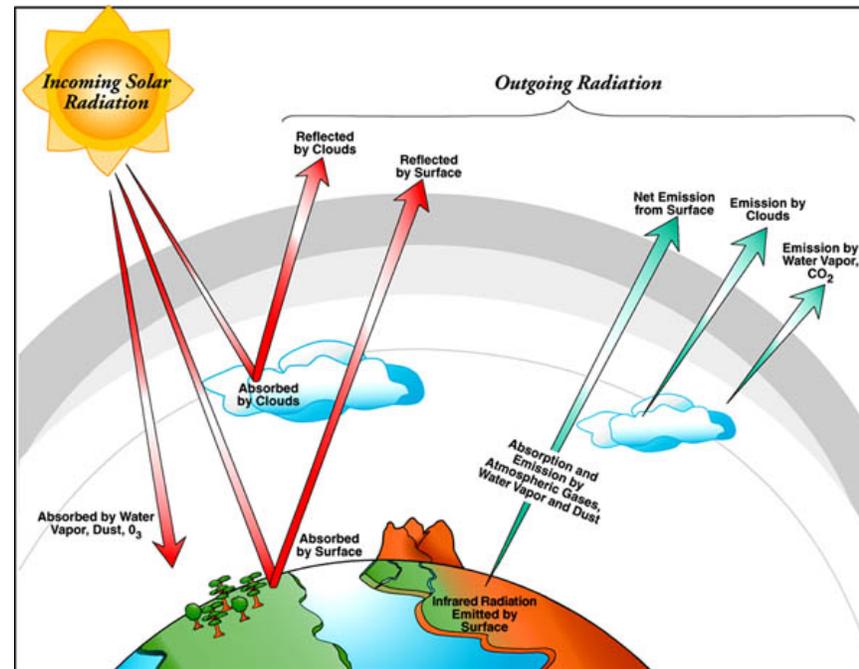
*[2] The Earth-Atmosphere Energy Balance*

The term remote sensing was first used in the United States in the 1950s by Ms. Evelyn Pruitt of the U.S. Office of Naval Research

[1] Satellite (1960)

REMOTE SENSING

The Process (Passive)

(1) Illumination by the sun

(2) Travel through the atmosphere

(3) Interactions with the target

(4) Recording of energy by the sensor

(5) Transmission, absorption, reflection, and emission

(6) Retrieval, interpretation, and analysis
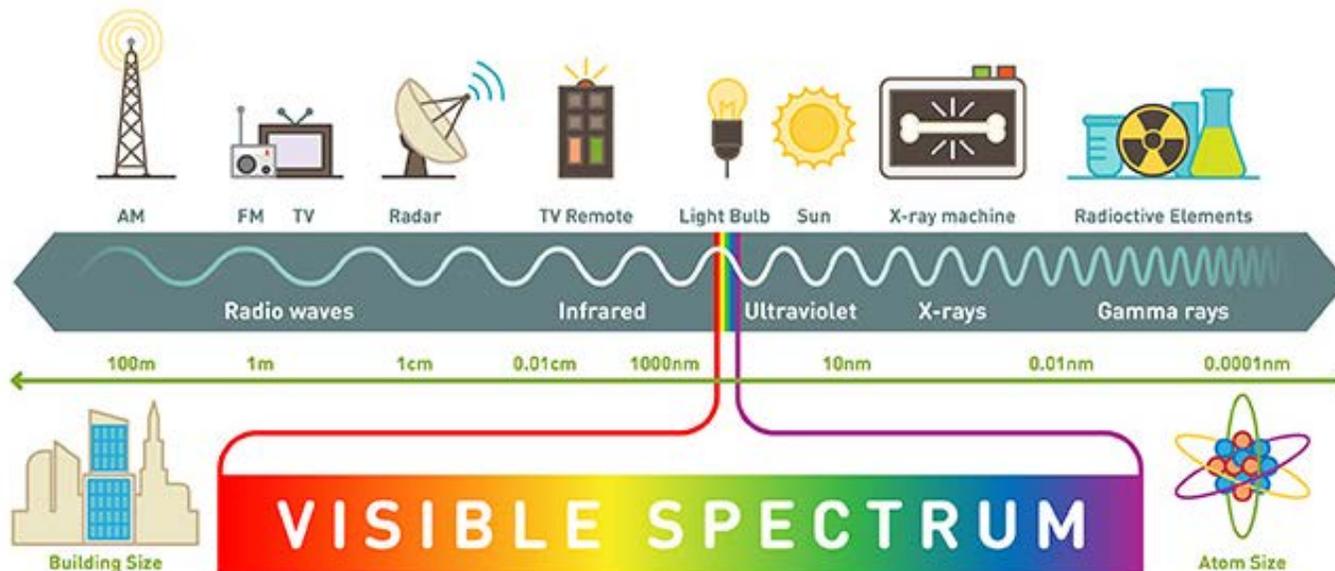
(7) Decision making for applications
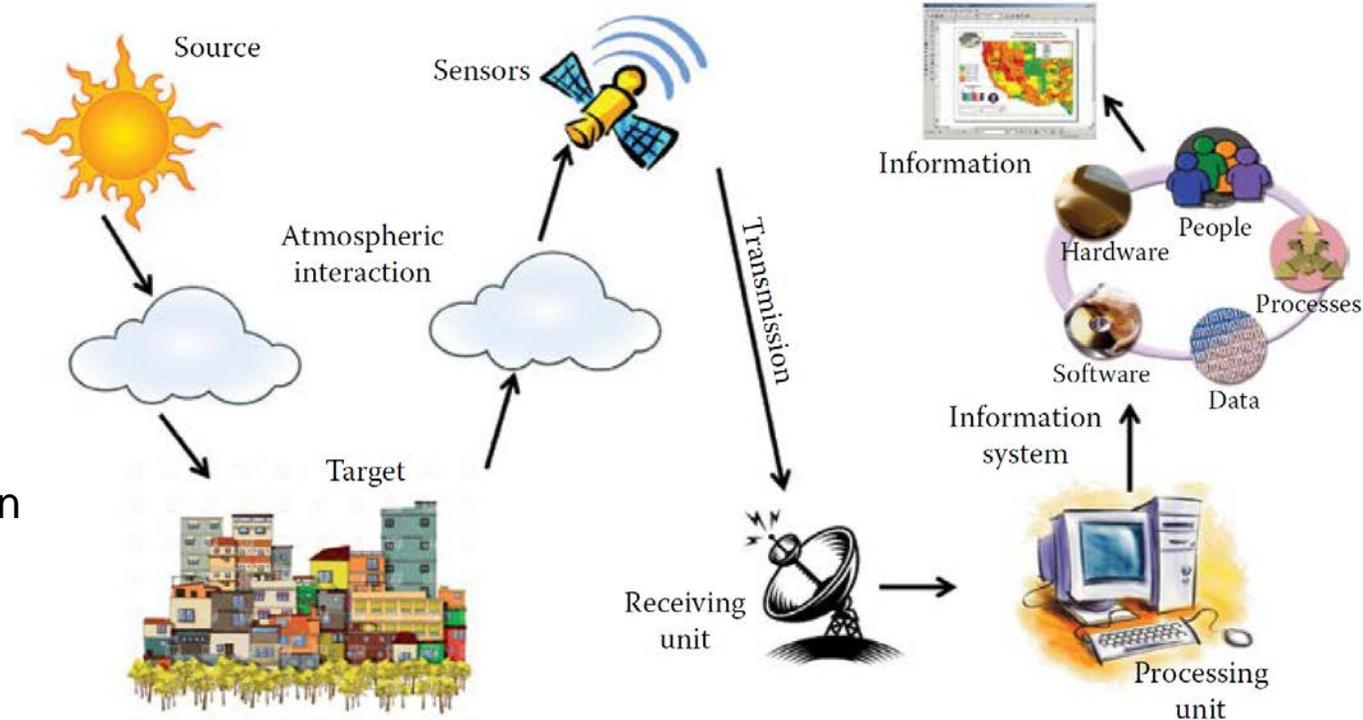
[3] N. B. Chang and K. Bai

ELECTROMAGNETIC (EM) SPECTRUM

What is Sensed?

- Continuous **set of radiation** sorted according to **wavelength** (or **frequency**)

  - Subdivisions are established for convenience and by traditions within different disciplines
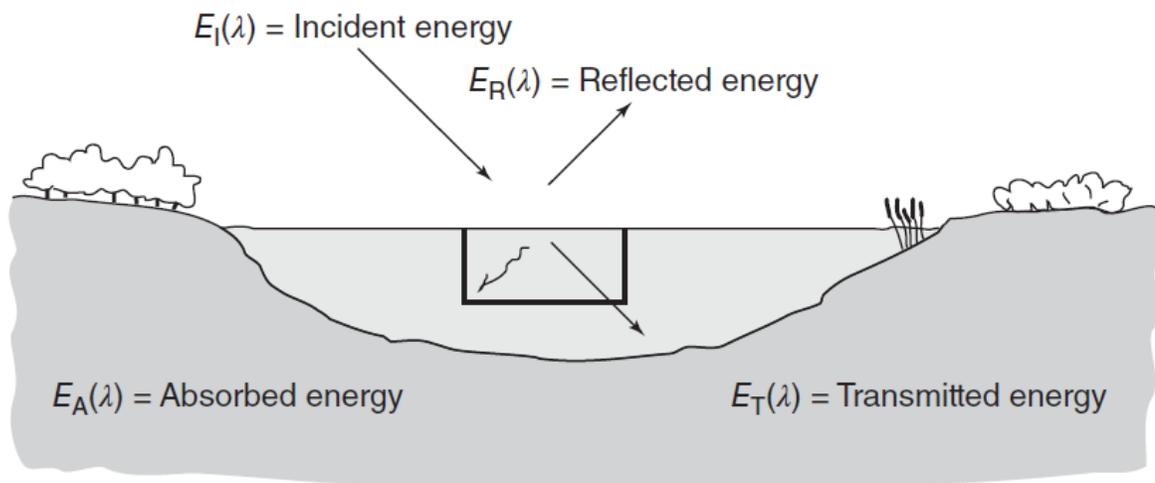
[4] Electromagnetic spectrum

*All this energy is inherently similar and propagates in accordance with the **wave model***

- There is neither a source nor a remote sensing system that "works" over the whole **EM spectrum**

ENERGY INTERACTIONS WITH THE EARTH SURFACE FEATURES

- Fractions of the **energy incident** on the element are **reflected, absorbed, and/or transmitted**

- Due to the **conservation of energy** theorem, their interrelationship is

$$E_I(\lambda) = E_R(\lambda) + E_A(\lambda) + E_T(\lambda)$$

$E_I(\lambda)$ = Incident energy

$E_R(\lambda)$ = Reflected energy

$E_A(\lambda)$ = Absorbed energy

$E_T(\lambda)$ = Transmitted energy

[5] K. Tempfli et al.

- $\rho(\lambda) = \dfrac{E_R(\lambda)}{E_I(\lambda)}$ Reflection coefficient

- $\tau(\lambda) = \dfrac{E_T(\lambda)}{E_I(\lambda)}$ Transmission coefficient

- $\alpha(\lambda) = \dfrac{E_A(\lambda)}{E_I(\lambda)}$ Absorption coefficient

$$\rho(\lambda) + \tau(\lambda) + \alpha(\lambda) = 1$$

- with all energy components being a function of wavelength

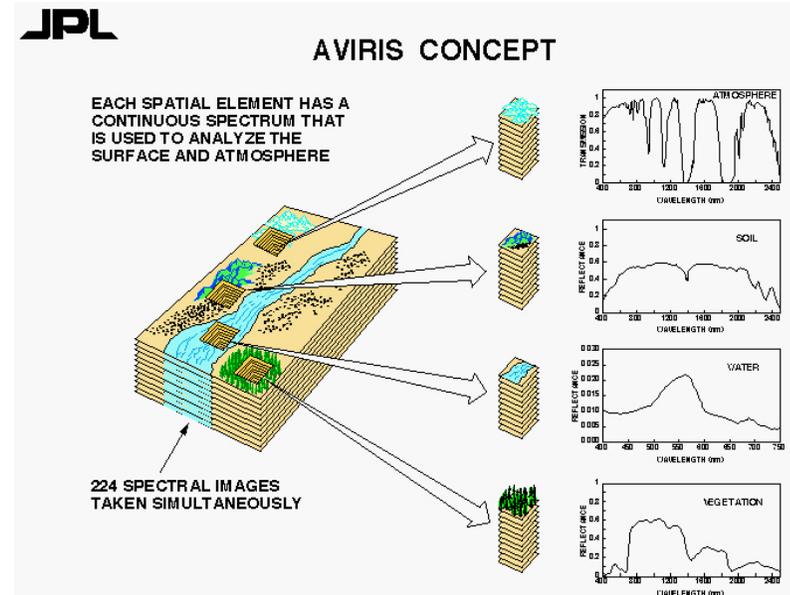# REFLECTANCE PROPERTIES OF SURFACE FEATURES

Spectral reflectance

- The **energy balance relationship** can be expressed in the form

$$E_R(\lambda) = E_I(\lambda) - [E_A(\lambda) + E_T(\lambda)]$$

- The reflectance properties of features can be quantified by measuring the portion of incident energy that is reflected

  – This is measured as a function of wavelength

  – It is called **spectral reflectance**

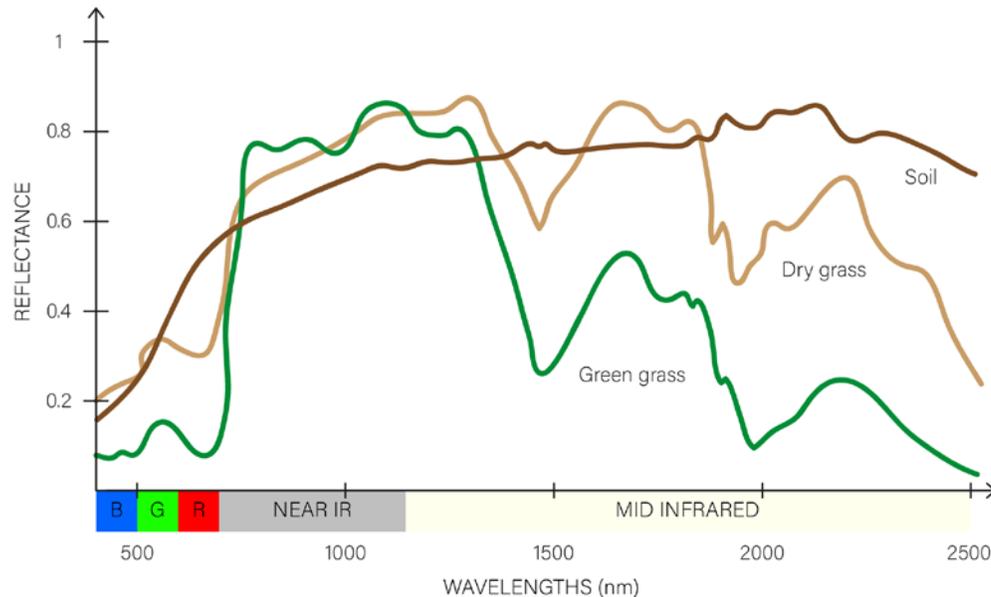$$\rho(\lambda) = \frac{E_r(\lambda)}{E_i(\lambda)} * 100$$

By analyzing the **spectral reflectance** it is possible to discriminate between **different land covers**

[6] AVIRIS Concept

SPECTRAL REFLECTANCE

Spectral Signature

- The **Spectral reflectance** of Land Cover (LC) classes are often **spectrally separable**

- However, the **degree of separation varies** among and within **spectral regions**


- E.g., Soil and Grass might reflect nearly equally in visible spectrum
- Yet they can be separated in the infrared spectrum

- Since **spectral reflectance's** often permit an assessment of the type and/or condition of the classes

  – They are often referred as **spectral signatures**

INFORMATIVE AND SPECTRAL CLASSES

Satellite images cannot map land covers in all its complexity and diversity

- Two types of classes are defined:
  - **Informative classes**: categories of interest to the users of the data (**land cover** classes)
  - **Spectral classes**: groups of uniform pixels with respect to the **spectral reflectance** in their spectral channels

- **One-to-one matches** between **informative and spectral classes** is rare
  - **One informative class** can correspond **to several spectral classes**
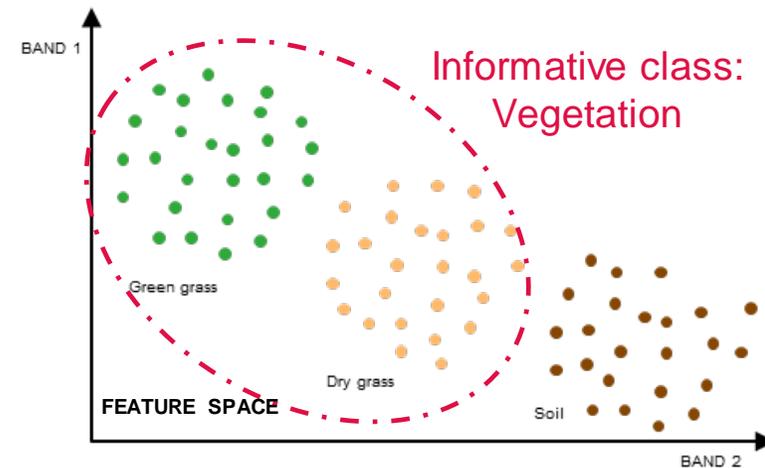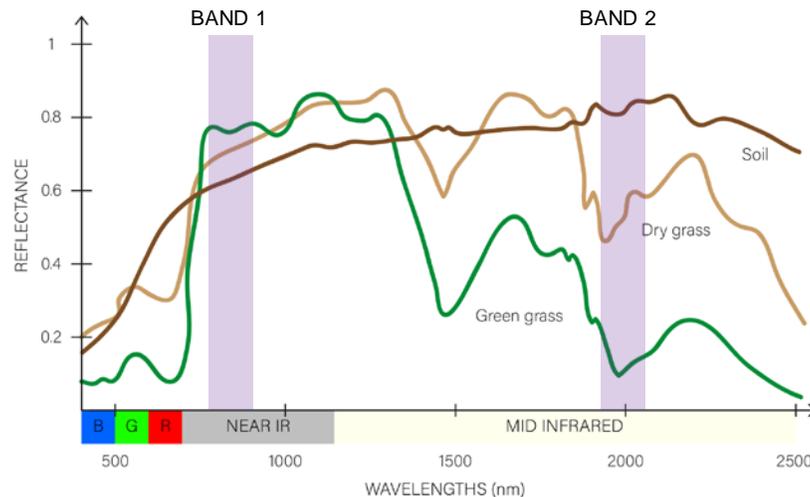
SPECTRAL RESPONSE PATTERNS

The term signature tends to imply a pattern that is absolute and unique

- This is not the case with the **spectral patterns** observed in the **natural world**

  – Grassland have variations in time, species composition, density, and vigor

  – Other factors produce additional variations even **within spectrally uniform classes**
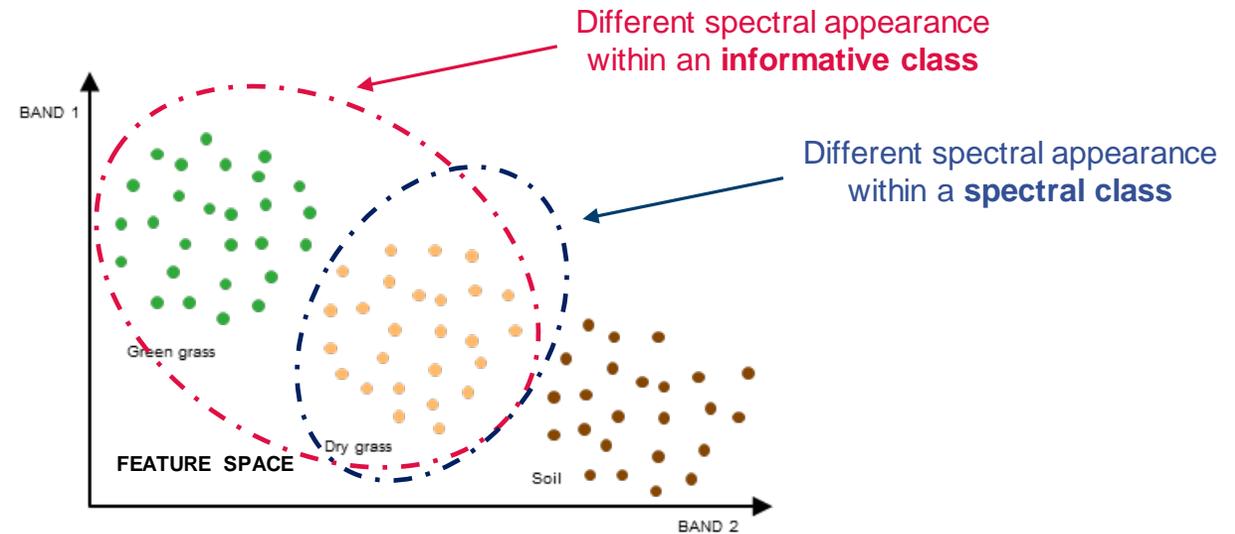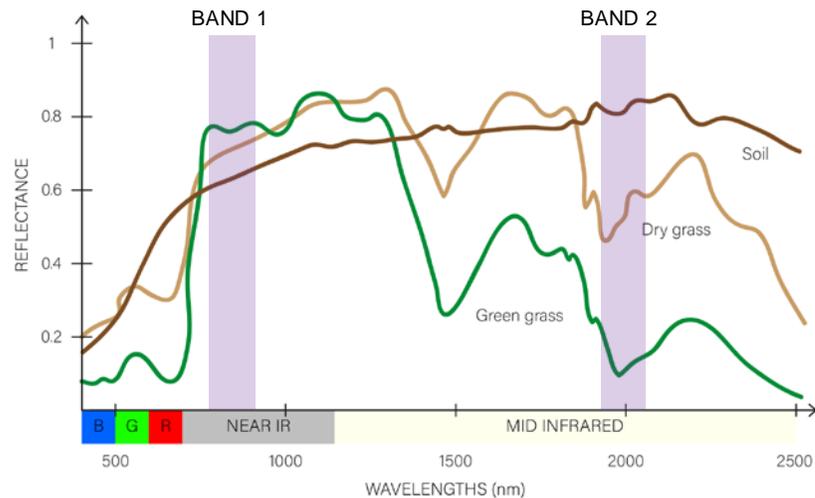
- They result in **spectral "response patterns"** rather than in **spectral "signatures"**

SPECTRAL RESPONSE PATTERNS

Temporal and spatial effects

- **Temporal effects:** factors that **change** the spectral characteristics of a **class over time**

[7] Tree all four seasons illustrations

- **Spatial effects:** same classes at a given time have different characteristics at different geographic locations

  – Entirely different soils, climates, and cultivation practices exist
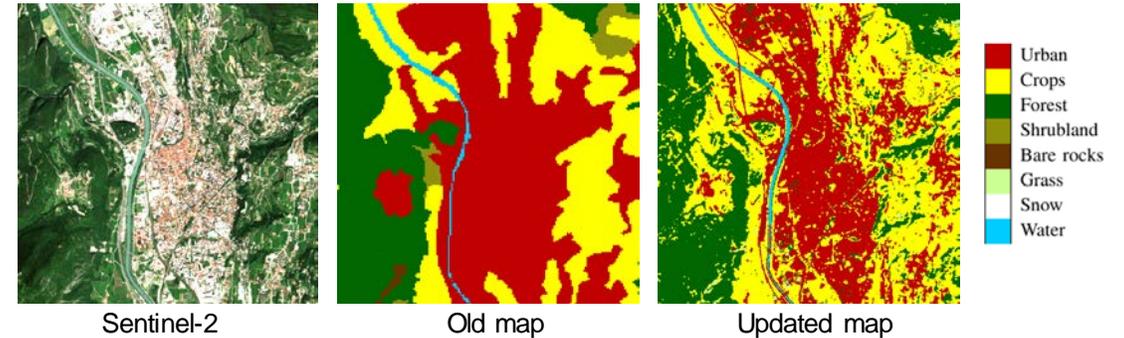

[8] Farm in China

[9] Farm in California

RECOGNITION PHASE

Extract information from the image and made available to the end user

- The result depends on the specification application considered, e.g.,

  - **Thematic maps** of the territory

  - **Change maps** between 2 different dates

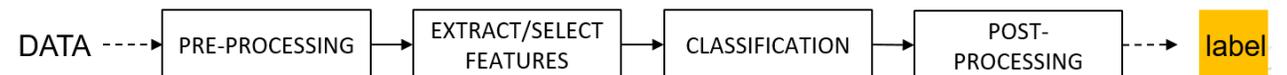 - Environmental risk maps

  - 3D Topographic maps

Sentinel-2   Old map   Updated map

Urban
Crops
Forest
Shrubland
Bare rocks
Grass
Snow
Water

[10] C. Paris et al.

- Two main strategies:

  - **Not automatic** (widespread in the past and still used today)

    o With the help of photo interpreters (**photointerpretation**)

  - **Automatic recognition techniques**

    o E.g., **Pattern recognition systems**

[11] Satellite Images

DATA ---→ PRE-PROCESSING → EXTRACT/SELECT FEATURES → CLASSIFICATION → POST-PROCESSING ---→ label

ELEMENTS OF IMAGE INTERPRETATION

The visual interpretation of aerial and space images is a complex process

- **Representation** of features from an overhead, often unfamiliar, **vertical perspective**

- **Wavelengths outside of the visible portion** of the spectrum

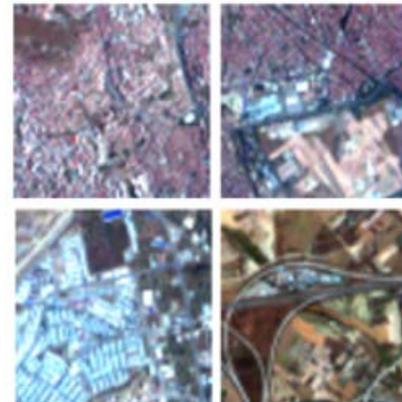- **Unfamiliar scales and resolutions**

Computer Vision

Remote Sensing

Nose, Eyes, Mouth

Wheels, License Plate, Headlights

Door, Windows, Steps

© MIT 6.S191: Introduction to Deep Learning
introtodeeplearning.com

Hyperspectral images: hundreds of channels

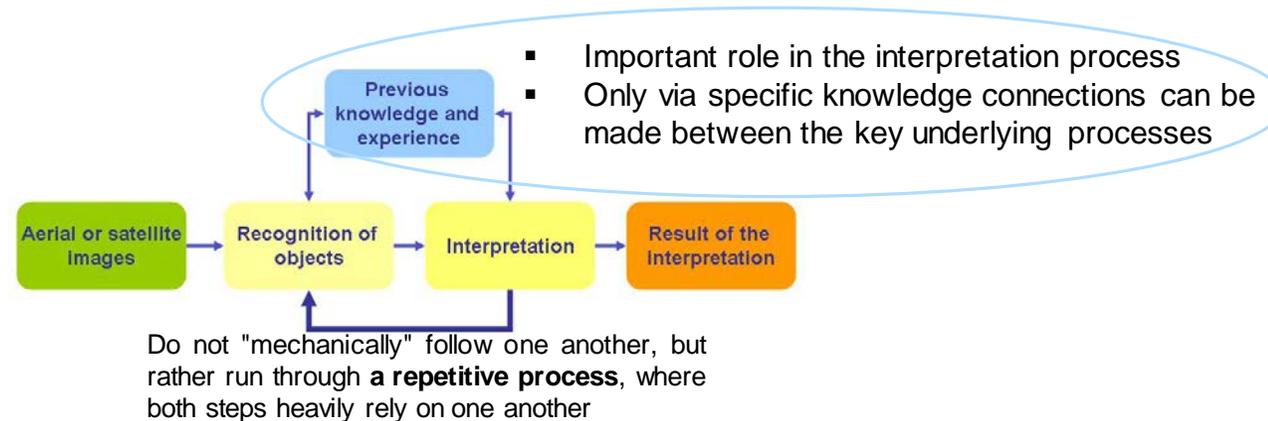*SAR images: noisy data*

[12] F. Ömrüuzun et al.

PHOTOINTERPRETATION

Two levels process

1. The **recognition of objects** such as streets, fields, rivers, etc.

   o It relates to the following say: "*I can recognize only what I already know* "

2. A **true interpretation** can be deduced through conclusions (from previously recognized objects) of situations

- Important role in the interpretation process
- Only via specific knowledge connections can be made between the key underlying processes

[13] Interpretation Process

Do not "mechanically" follow one another, but rather run through **a repetitive process**, where both steps heavily rely on one another

- A human operator is **not able to quickly process** the enormous amount of data that the sensors provide

- Only **qualitative** analyzes are possible

- The analyzes are **subjective**

HOW DETAILED IS THE INTERPRETATION?

Data resolutions constrain the level of information (i.e., level of interpretation)

- E.g., Hierarchical thematic classification

Level 1	Level 2	Level 3
Agricultural areas	2.1. Arable land	2.1.1. Non-irrigated arable land 2.1.2. Permanently irrigated land 2.1.3. Rice fields
	2.2. Permanent crops	2.2.1. Vineyards 2.2.2. Fruit trees and berry plantations 2.2.3. Olive groves
	2.3. Pastures	2.3.1. Pastures
	2.4. Heterogeneous agricultural areas	2.4.1. Annual crops associated with permanent crops 2.4.2. Complex cultivation patterns 2.4.3. Land principally occupied by agriculture, with significant areas of natural vegetation 2.4.4. Agro-forestry areas

Level 1 (~80m)   Level 2 (~20-30m)   Level 3 (<5m)

AUTOMATIC RECOGNITION

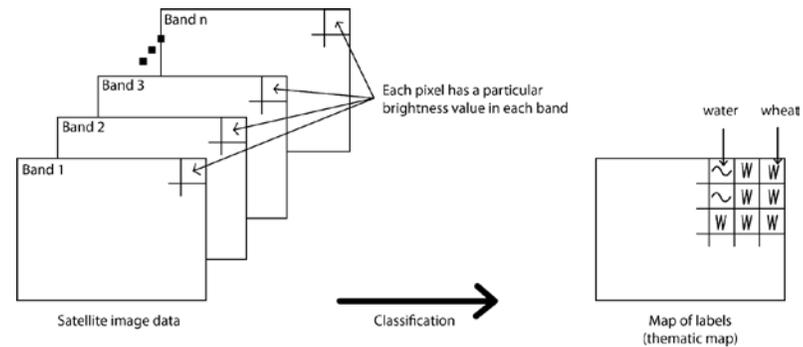## Automatic Classification

- Main advantages

  – Can **process large amounts** of data (multispectral and multitemporal) in **a short time**

  – Perform **quantitative assessments**

  – The analyzes are **objective**

- Main disadvantage

  – Difficult to include all the information that could be identify by the human brain

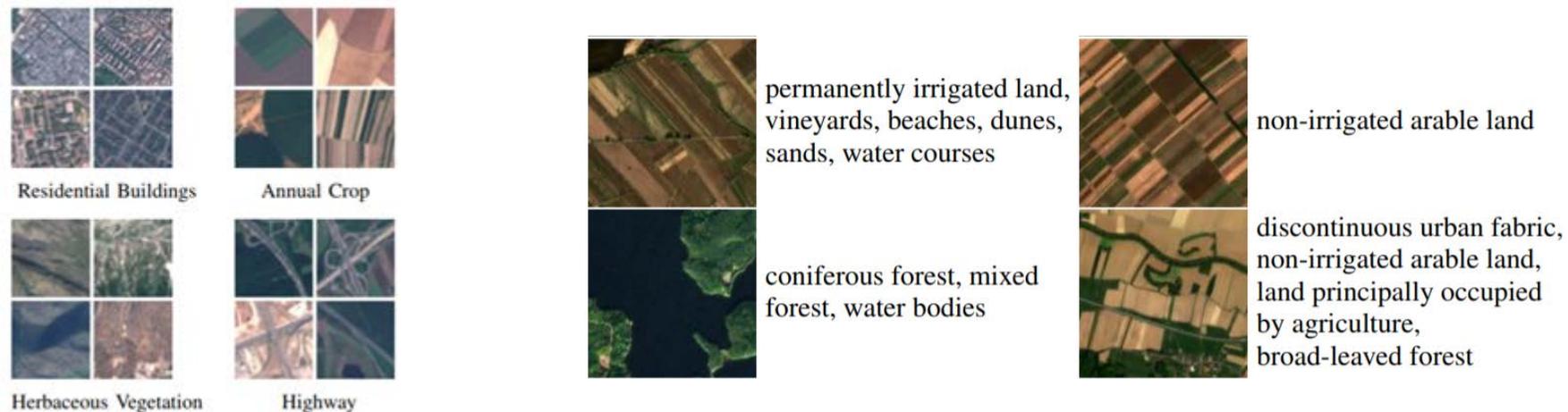- **Classification** is one of the most typical applications of the **recognition phase**

Thematic classes: ■ trees, ▢ asphalt, ■ bitumen, ■ gravel, ■ metal sheets, ▢ shadow, ■ meadows, ■ self-blocking bricks, ■ bare soil.

CLASSIFICATION TASKS

Output variable takes a class label

- **Pixel-wise** classification

- **Patch-based** classification (with single or multiple land-cover class labels)

[14] P. Helber et al.

[15] G. Sumbul et al.

PATTERN RECOGNITION

Classification is a task that falls into the general category of pattern recognition

- *Pattern*: a form, template, **composite of features**, or model (or, more abstractly, a set of rules)
 - Which can be used to make or to generate objects or parts of an object

Spectral bands, spatial arrangement patterns

Completely
Deterministic

Completely
Random

E.g., Crystal structure

E.g., White noise

[16] Remote Sensing
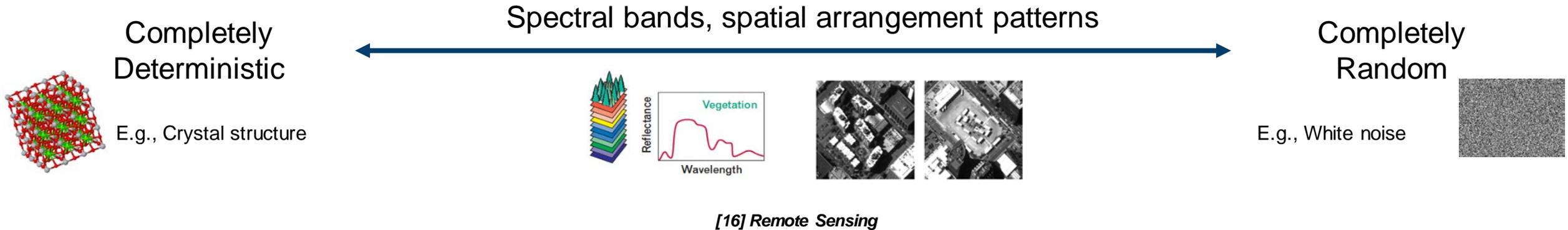
- **Pattern recognition:** automatic discovery of patterns in data through the use of tools from
  - Statistics, probability, computational geometry, **machine (deep) learning**, signal processing, and algorithm design
- Use of these patterns to take actions such as classifying the data into different **classes**

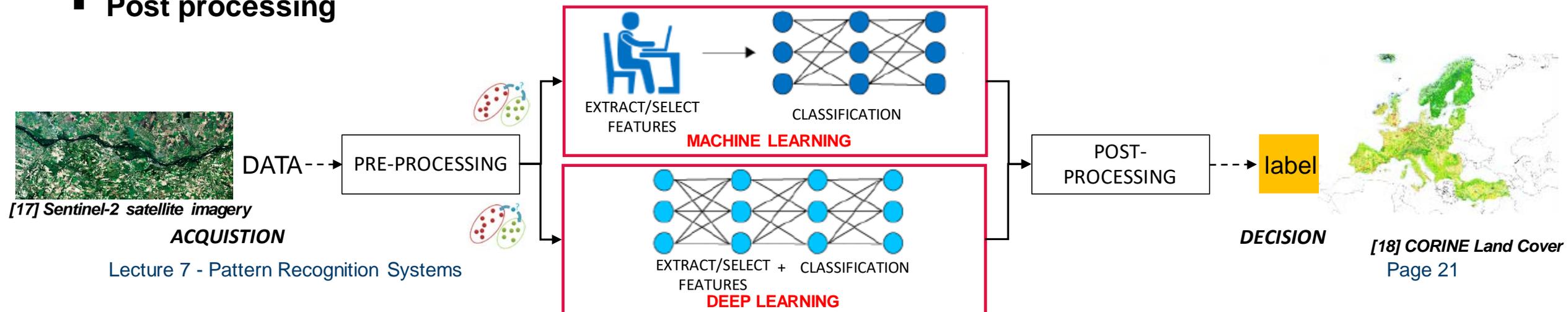# MACHINE LEARNING

Form of Learning

- **Supervised learning:** correct responses for input data are given

  - "teacher" signal, correct "outcomes", "labels" for the data

  - Classic frameworks: $\boxed{\textbf{classification}, \textbf{regression}}$
    THIS LECTURE

- **Unsupervised learning:** only data are given

  - Find "hidden" structure, patterns

  - Classical frameworks: **clustering**, **dimensionality reduction**

- **Reinforcement learning:** data including (sparse) **reward** $r(X)$

  - Discover actions a that minimize total future reward R

  - **Active** learning: experience depends on choice of a

$V(s)$ internal model
$\pi(s, a)$
r reward
environment
action a
s
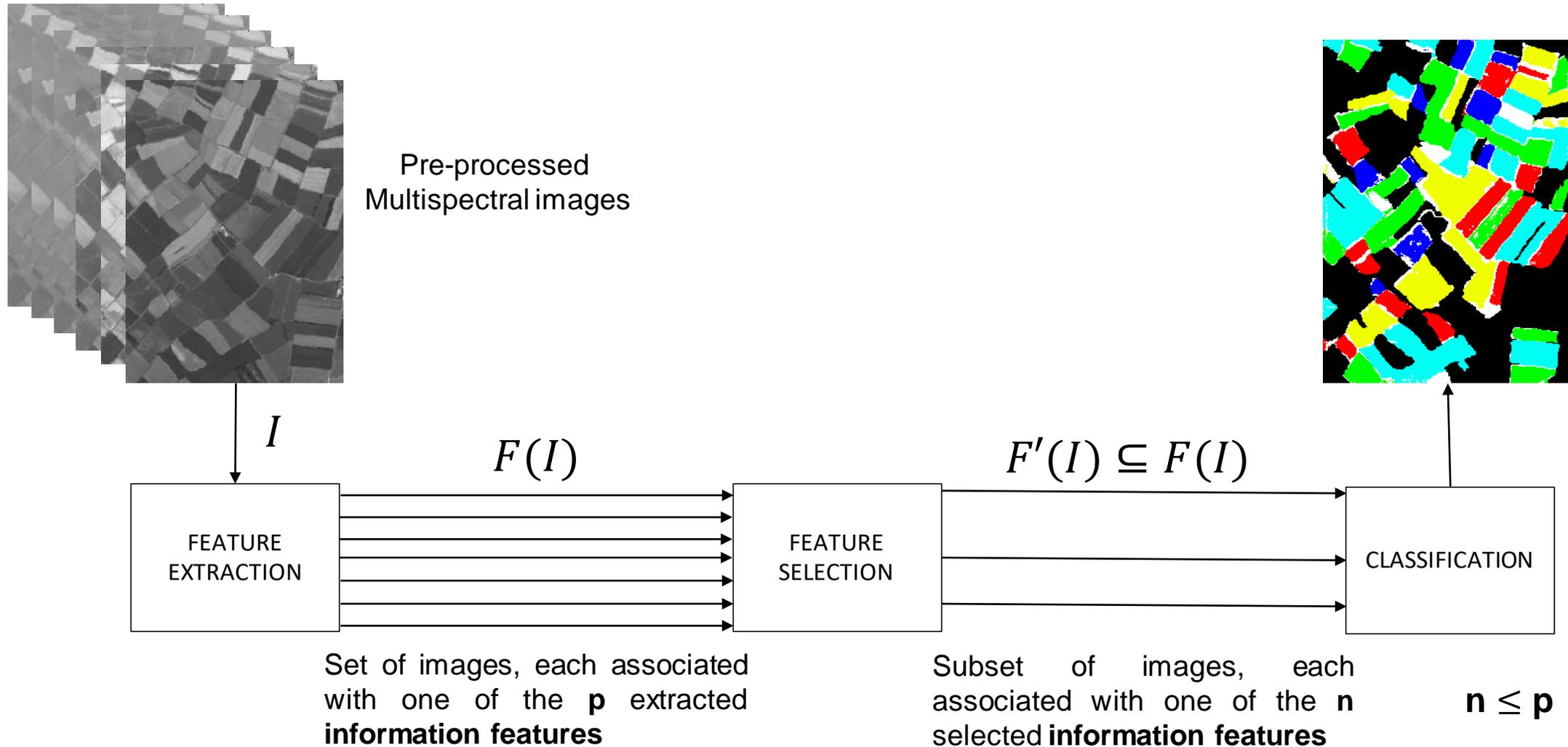sensory input (states)
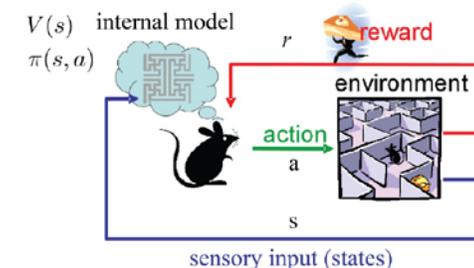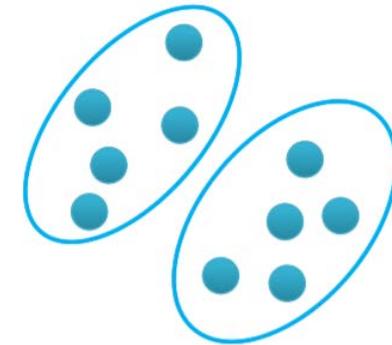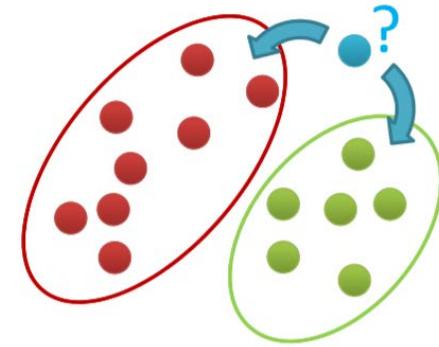
PATTERN RECOGNITION SYSTEM

Block Scheme

- **Pre-processing**
    - Atmospheric and geometrics corrections, filtering, etc.,

- **Feature extraction/selection**
    - Extraction of information parameters
    - Selection of information parameters

- **Classification**
    - Based on the information parameters previously extracted and selected

- **Post processing**

[17] Sentinel-2 satellite imagery

ACQUISTION

DATA

PRE-PROCESSING

EXTRACT/SELECT FEATURES

CLASSIFICATION

MACHINE LEARNING

EXTRACT/SELECT + FEATURES CLASSIFICATION

DEEP LEARNING

POST-PROCESSING

label

DECISION

[18] CORINE Land Cover

FEATURE EXTRACTION AND SELECTION

Pre-processed Multispectral images

$$I$$

$$F(I)$$

$$F'(I) \subseteq F(I)$$

| FEATURE EXTRACTION | | FEATURE SELECTION | | CLASSIFICATION |

Set of images, each associated with one of the **p** extracted **information features**

Subset of images, each associated with one of the **n** selected **information features**

$$n \leq p$$

FEATURE EXTRACTION

Extract the features of the images to be classified

- The choice of the features depends on the classification problem, however they should be:

  - Simple to extract

  - Insensitive to noise

  - Discriminative of patterns within different classes

$$F(I)$$

$$I \longrightarrow \boxed{\text{FEATURE EXTRACTION}}$$
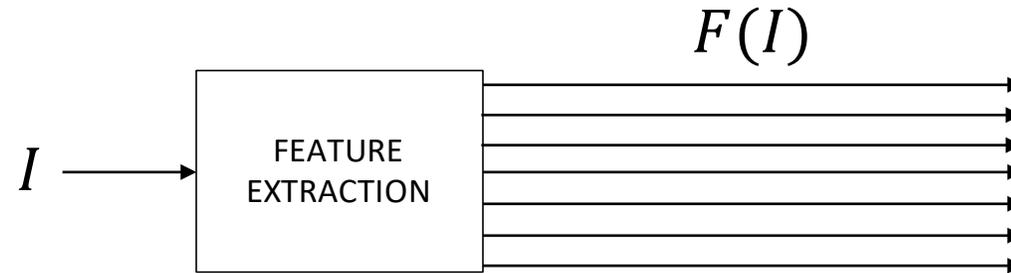
- The output of the **feature extraction module** are features such as:

  - Original bands (unchanged)

  - Relationships between bands

  - Image texture

BANDS RATIOS

Example: Vegetation index

- **Ratio of DNs** in spectral band m to those in band n, pixel-by-pixel

$$R_{mn}(x,y) = \frac{DN_m(x,y)}{DN_n(x,y)}$$

- E.g., Ratios of NIR to visible bands can enhance radiance differences between soils and vegetation

    - **Ratio Vegetation Index (RVI)**

$$RVI = \frac{NIR}{RED}$$

    - **Normalized Difference Vegetation Index (NDVI)**

$$NDVI = \frac{NIR - RED}{NIR + RED} = \frac{RVI - 1}{RVI + 1}$$

- Soil will exhibit ratio values near one, while vegetation will show a relatively larger ratio
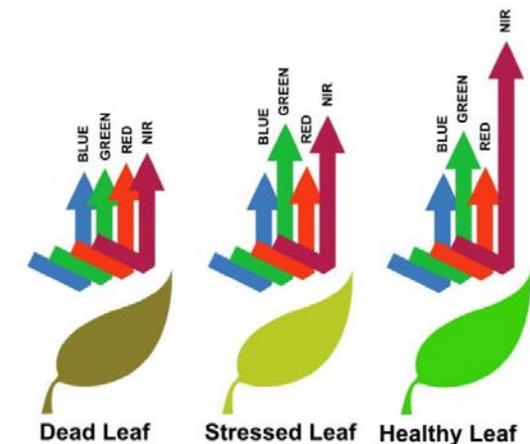
Dead Leaf   Stressed Leaf   Healthy Leaf   *[19] NDVI*

NORMALIZED DIFFERENCE VEGETATION INDEX (NDVI)

Image example
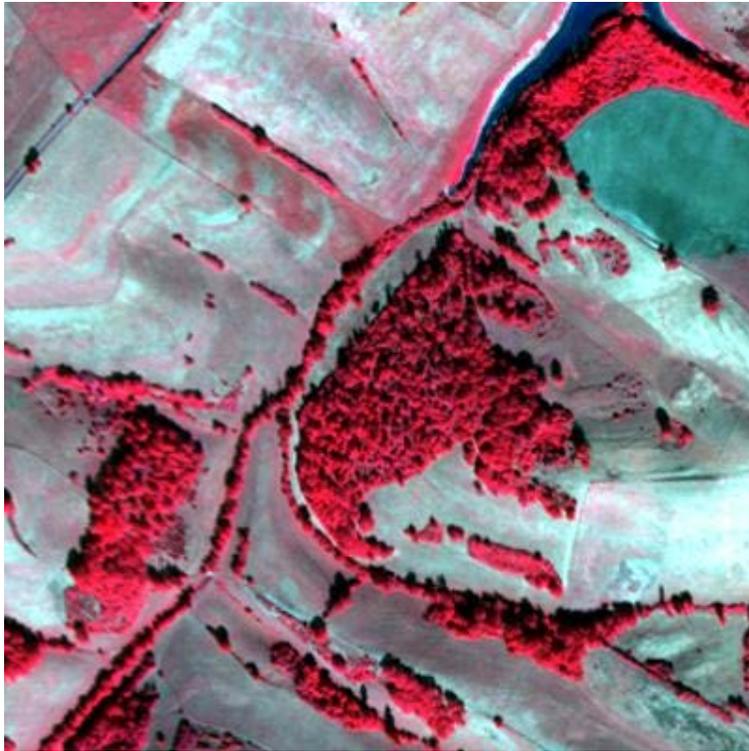
False colour composite

NDVI

FEATURE EXTRACTION

Approaches

- **Pixel-based**: approach based on pixel (spectral patterns)
- **Region-based**: approach based on regions (spatial patterns)
- **Hybrid** (spectral and spatial patterns)

- The choice depends on the **application**, type of **images** and **computational resources** available
  - This decision conditions the entire classification phase
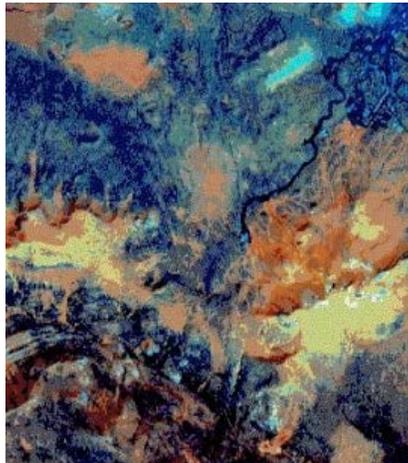
Images with very uneven areas
pixel-based
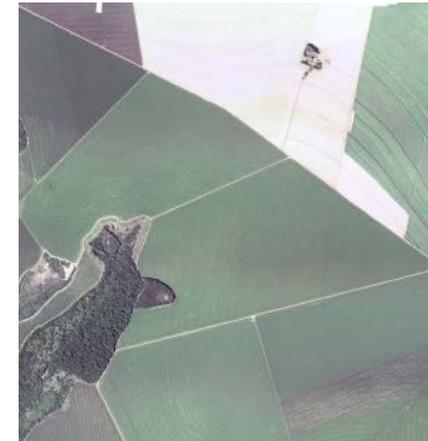
Images with homogeneous areas
region-based or hybrid

SPECTRAL CLASSIFIERS

Pixel-based: analyse spectral patterns

- Each pixel is individually considered (i.e., as values isolated from their neighbors)
  - Using the values from **several features** (e.g., **spectral bands**, **temporal bands**,…)

- Pixels with similar spectral reflectance are **grouped** together in one class
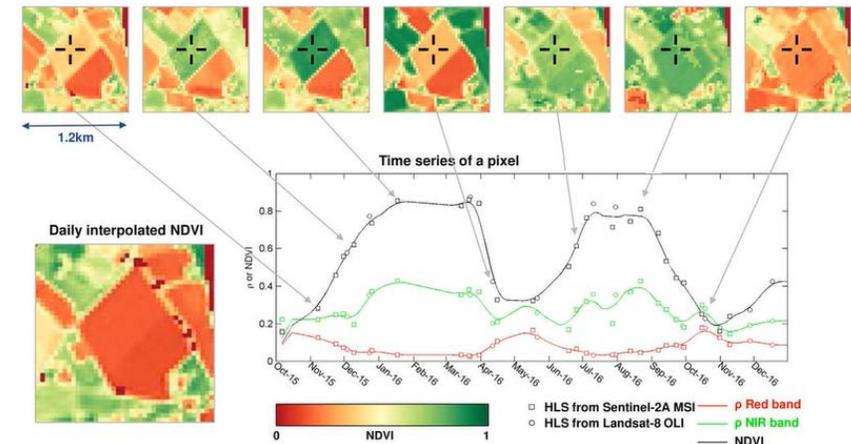
[20] Modified from Future US Land Imaging

- **No attention** is paid to the **neighbors** of the pixel being classified

FEATURE EXTRACTION

Pixel-based

- The features are extracted for **each pixel** of the image to be classified

- The **output** includes **many images as many features extracted**
  - Each image contains the information relating to one of the features extracted

Pre-processed
Multispectral images

I

FEATURE
EXTRACTION

$F(I)$

KURTOSIS

VARIANCE

ENTROPY

ORIGINAL BANDS
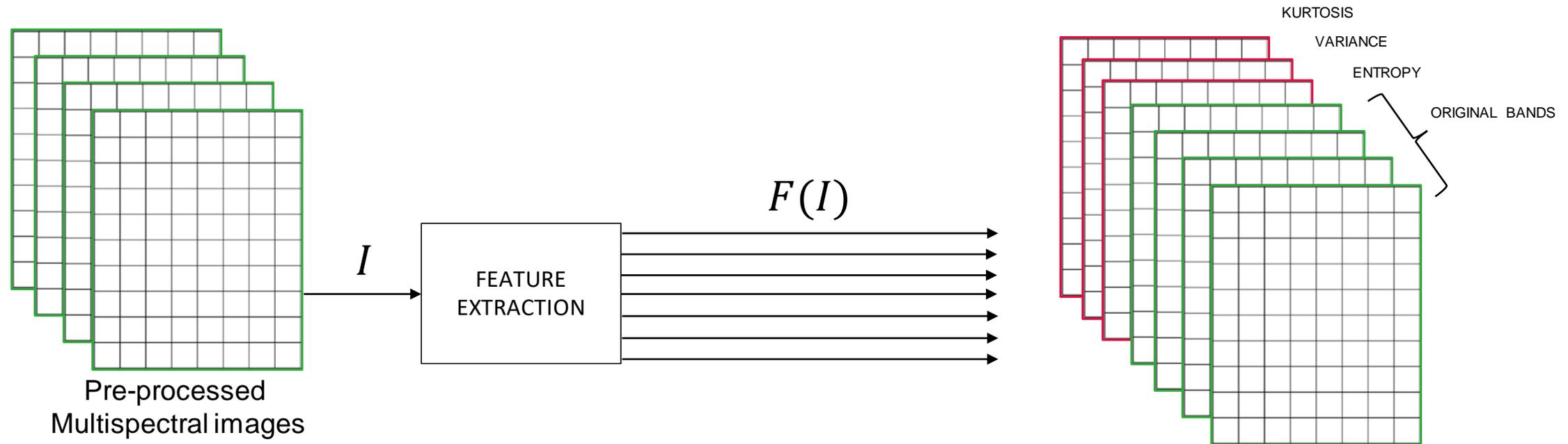
Images associated with the
extracted features

FEATURE EXTRACTION

Pixel-based Example: Principal Component Analysis (PCA)

- **Nearby bands** tend to be **correlated** (correlation means redundancy - images "look alike")

- Theoretically n bands = n dimensional data

- The "actual" dimension required to represent data with negligible information loss is lower

- **PCA: finds the linear subspace that shows the largest variances**
  - I.e., Eigenvalue decomposition of the covariance matrix

n = 103 bands

Affine transformation to decorrelate the axis

103 components

Pick K dimensions with the highest amount of variance

K = 4 Principal Components

Original   *PCA*

OA   71.82%   70.55%
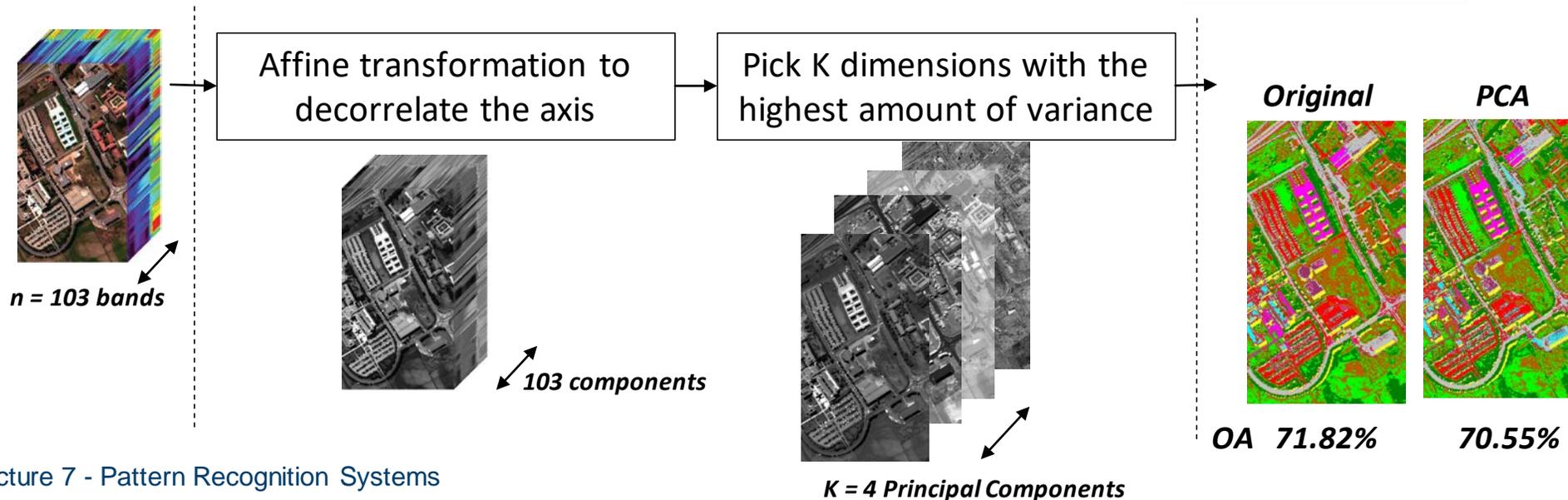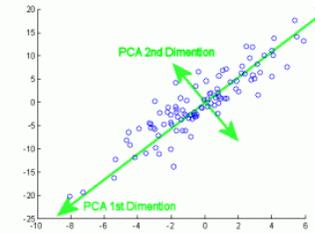
SPATIAL CLASSIFIERS

Region-based: analyse spatial patterns

- Categorization of image **pixels based on their spatial relationship** with pixels surrounding them
  - Consider image texture, pixel proximity, feature size, shape, directionality, repetition, and context
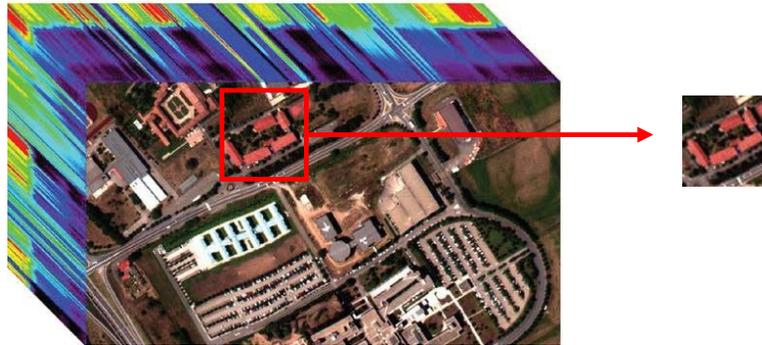
WorldView-2 Panchromatic - Resolution 0.46 [m]

- Attempt to **replicate** the **spatial synthesis** done by the human analyst during the **visual interpretation process**

- They are **more complex and computationally intensive** than spectral classifiers

FEATURE EXTRACTION

Region-based

- The original image is **segmented** first

- Therefore, for each region, a series of **descriptive parameters** are extracted
  - E.g., area, perimeter, average DN, etc.

$$R = [R_1, R_2, \dots, R_Q]$$

SEGMENTATION

FEATURE EXTRACTION

$F(I)$

I

Pre-processed Multispectral images

area, perimeter, average DN, etc.

area, perimeter, average DN, etc.

area, perimeter, average DN, etc.

area, perimeter, average DN, etc.

area, perimeter, average DN, etc.
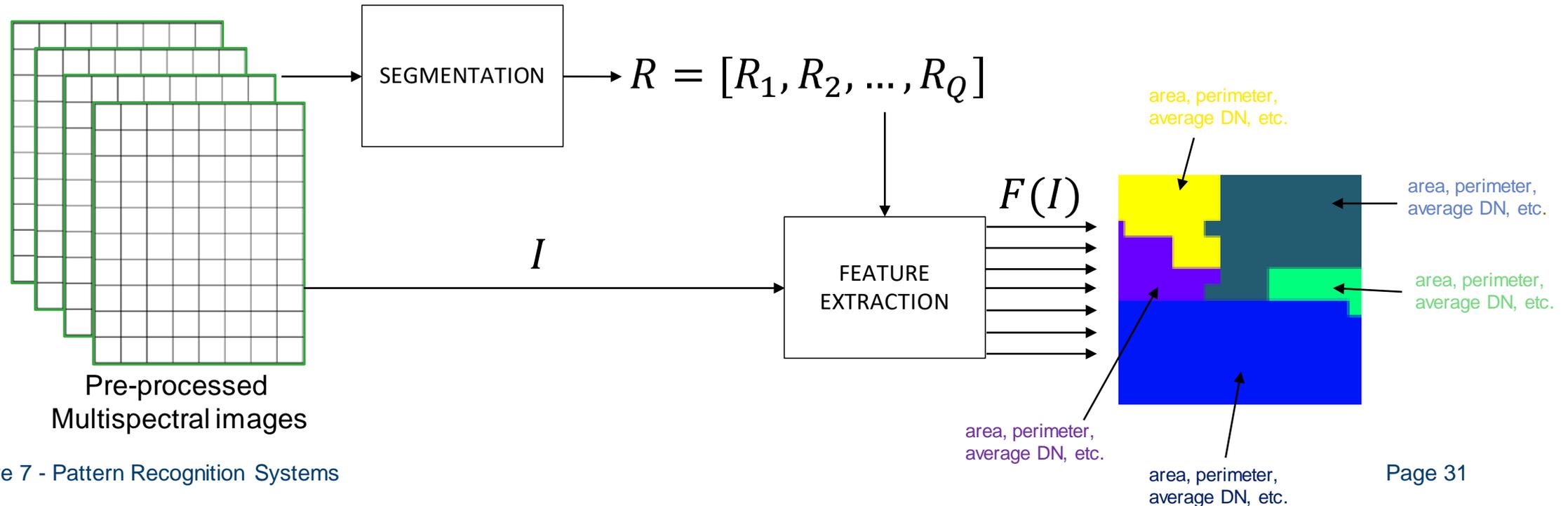
SPECTRAL-SPATIAL CLASSIFIERS
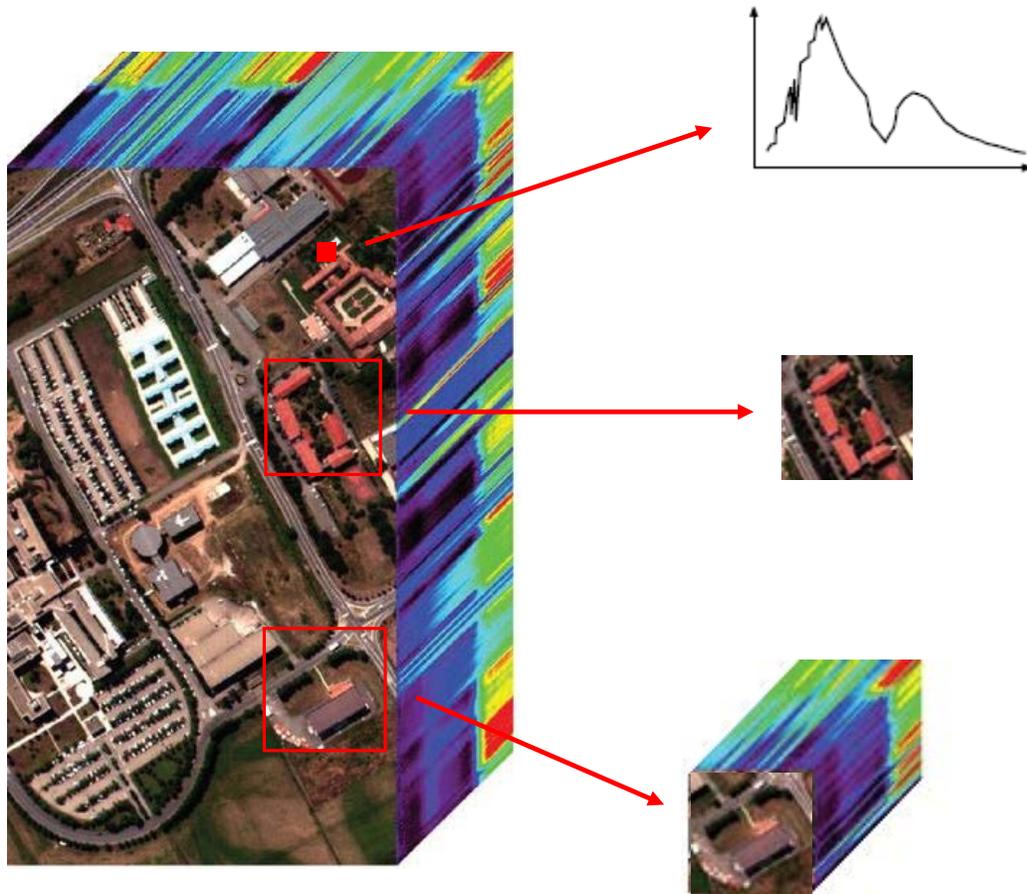
1. **Spectral:**
   - Physically-based spectral features
   - Statistical multivariate methods: linear (**PCA**) and nonlinear

2. **Spatial/contextual**
   - Standard image processing descriptors
   - Advanced computer vision descriptors

3. **Spatial-spectral**
   - Extract features from spectral patches or regions

FEATURE SPACE

Geometrical visualization of the information associated with the feature vectors

■ Feature vectors can be represented in a space with p dimensions

– With p the number of the features extracted

■ This space is called **feature space**

– E.g., p=3 $\qquad \boldsymbol{f}' = [0.3, 0.1, \dots, 0.8\ ]^t$
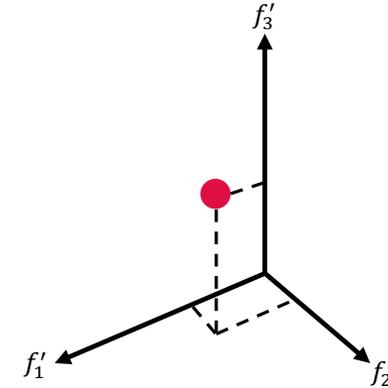
■ In the feature space the vectors tend to form **patterns/clusters** that correspond to the different **land cover classes**

■ The **shape and size of these clusters** depend on several factors such as

– The **features used** to represent the scene considered

– The specific **type of coverage**
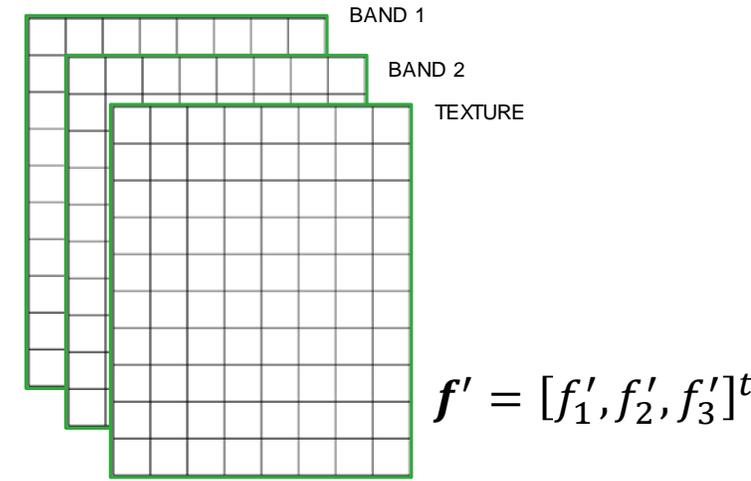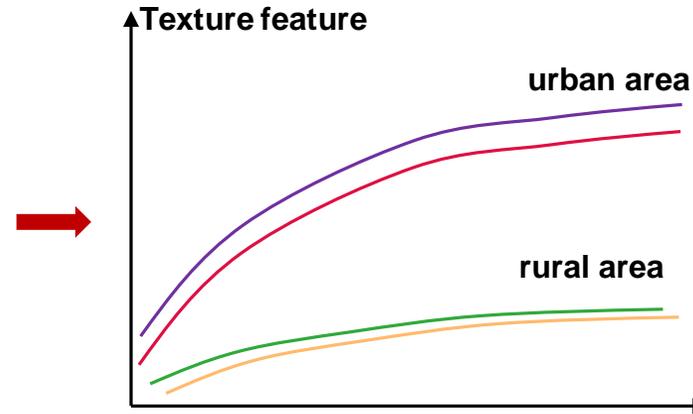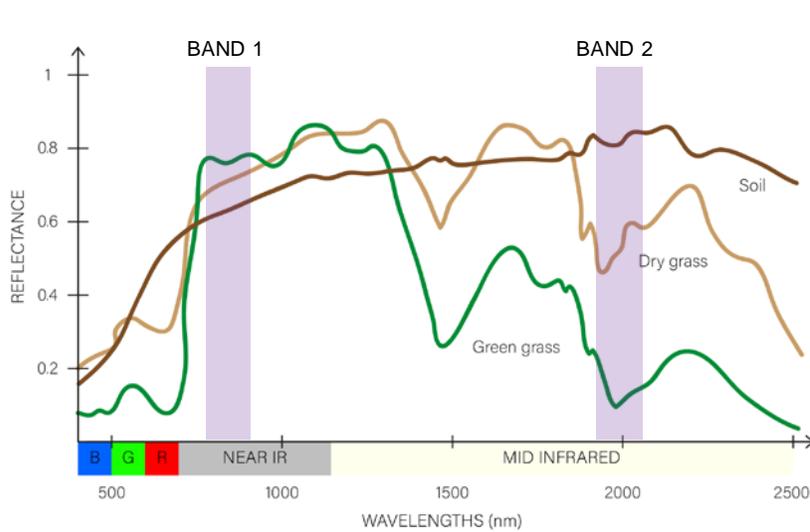
– The **noise**

– **Topographic effects**

EXAMPLE OF FEATURE SPACE
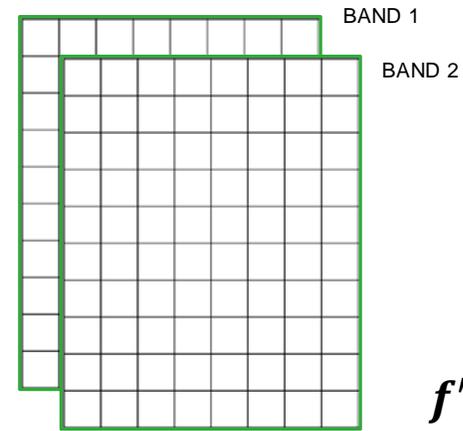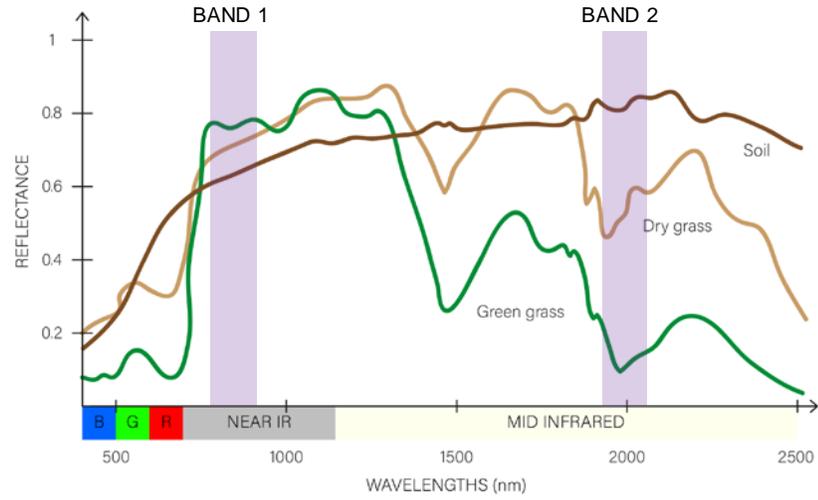
$$\boldsymbol{f}' = [f_1', f_2']^t$$

EXAMPLE OF FEATURE SPACE

$$\boldsymbol{f}' = [f_1', f_2', f_3']^t$$

The texture feature allows the identification of two subclasses of the class grass
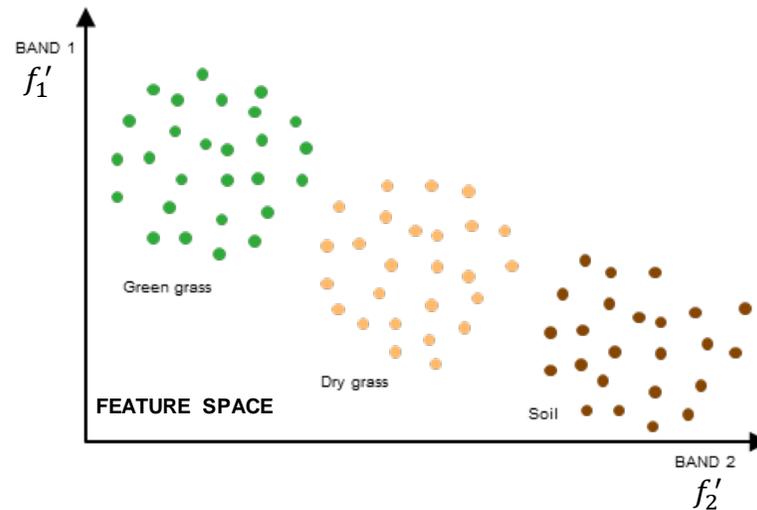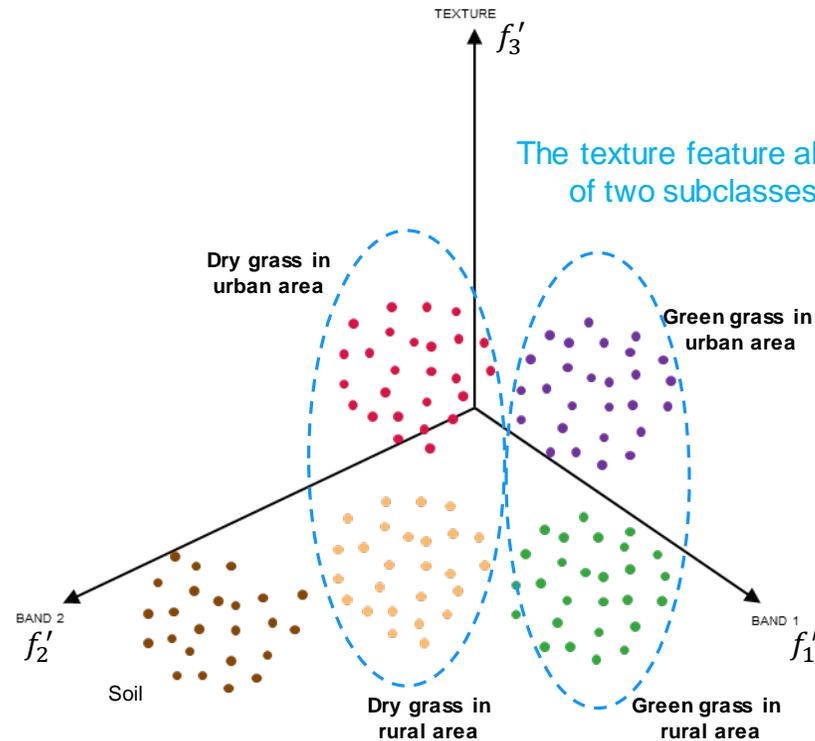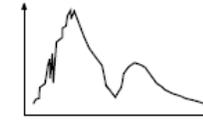
STANDARDIZATION OF THE FEATURES

Common requirement for many machine learning algorithms

- At the output of the feature extractor, each pixel of region are represented by a **feature vector** f:

$$f = [f_1, f_2, \ldots, f_p]^t$$

- The range of values of the features can vary widely, e.g., different scales, outliers, etc.

  - Classification algorithms cannot work properly if the range of all features is not rescaled

- Example:

  - Red looks more similar to blue than yellow

  - However, the features on the x- and y-axes have different scales

  - Therefore, the observed similarity might be an artifact of unscaled data

  - After normalization (with z-score), all the features have the same scale

  - Now, red is actually more similar to yellow
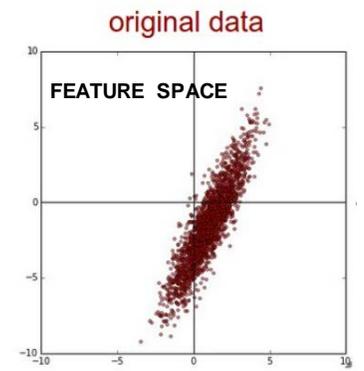
[21] Data Preparation
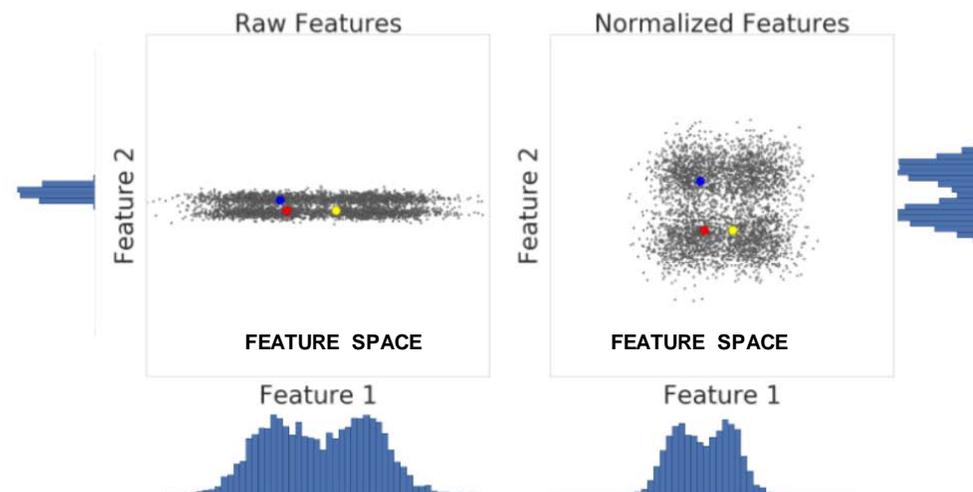
FEATURE SCALING

Standard approaches

- **Mean Subtraction:** subtract the mean of each parameter $f_i (i = 1, ..., p)$

  - Geometric interpretation: center the data cloud around the origin of every dimension
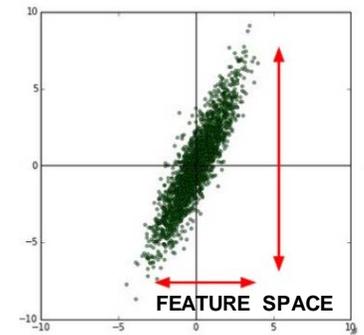
$$f_i' = f_i - f_i^{mean}$$

- **Normalization:** normalize the data dimensions so that they are of approximately the same scale

  - Calculates the minimum value f_i^{min} and maximum f_i^{max} and apply the **transformation**:

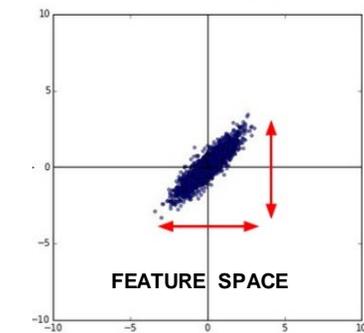$$f_i' = \frac{f_i - f_i^{min}}{f_i^{max} - f_i^{min}}$$

original data

FEATURE SPACE

$f = [f_1, f_2, ..., f_p]^t$

zero-centered data

FEATURE SPACE

$f' = [f_1', f_2', ..., f_p']^t$

normalized data

FEATURE SPACE

$f' = [f_1', f_2', ..., f_p']^t$

[22] Common data reprocessing

FEATURE SELECTION

- Select the **subset of n parameters, among all the p extracts**
- That allow **the best possible distinction** between the different classes that has be identified $(\text{n} \leq p)$

KURTOSIS
VARIANCE
ENTROPY
ORIGINAL BANDS

$F(I)$

FEATURE SELECTION

$F'(I) \subseteq F(I)$
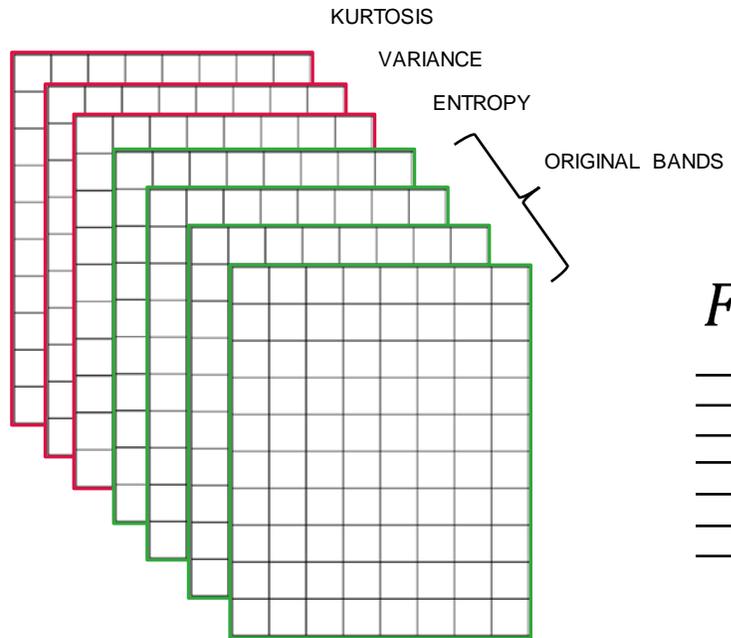
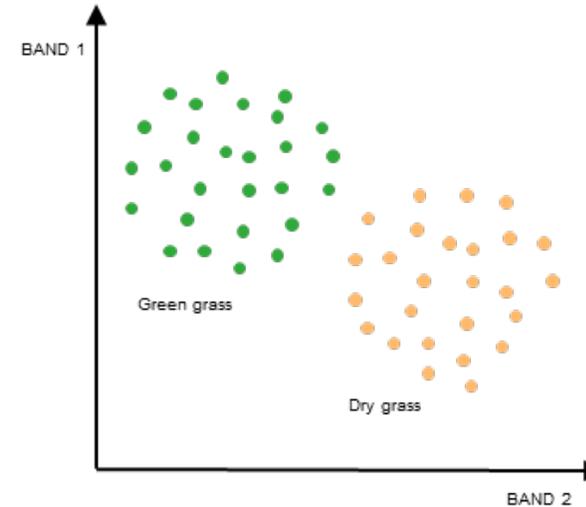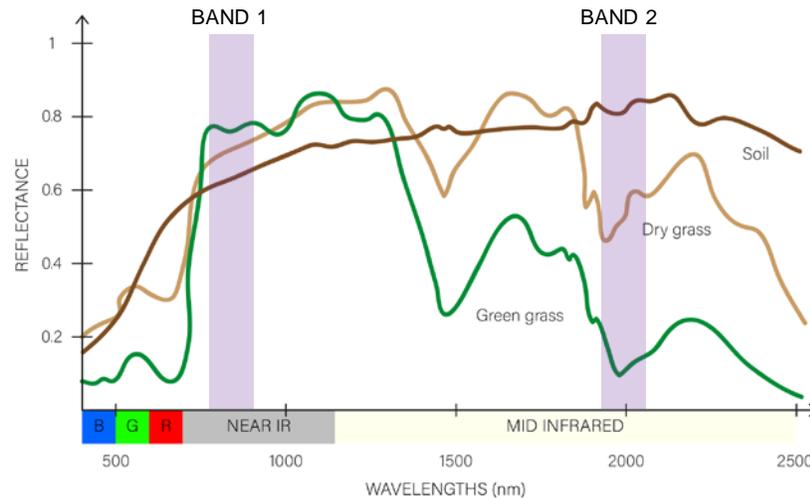Images associated with the p extracted parameters
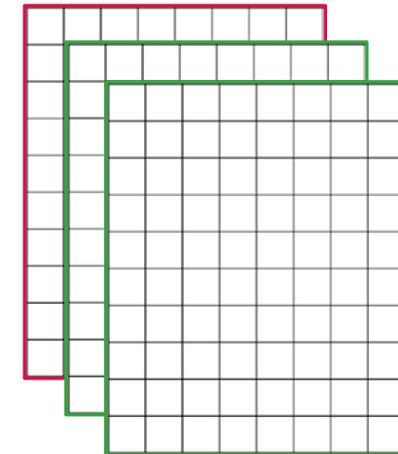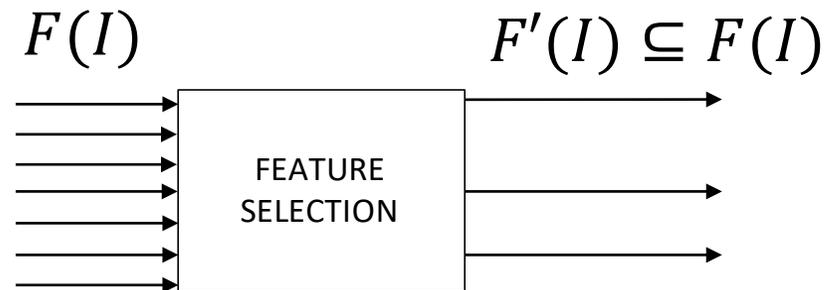
Images associated with the p extracted parameters

FEATURE SELECTION EXAMPLE

Classification of two land cover classes

- Problem: distinguish only between the 'green grass" and ''dry grass" classes


- The task of the feature selection step is to identify that

  - Only **band 2** is useful for discriminating the classes

- Reduce the number of features (spectral bands) and pass them to the classifier

  - Which makes a final decision

FEATURE SELECTION

Properties

- The selected features must meet 2 basic conditions

  - Vectors belonging to the **same class must be as close as possible in the space of the feature**
    - I.e., They must originate compact clusters

  - Clusters consisting of vectors belonging **to different classes must be as far apart as possible**
   - I.e., They must be well separated
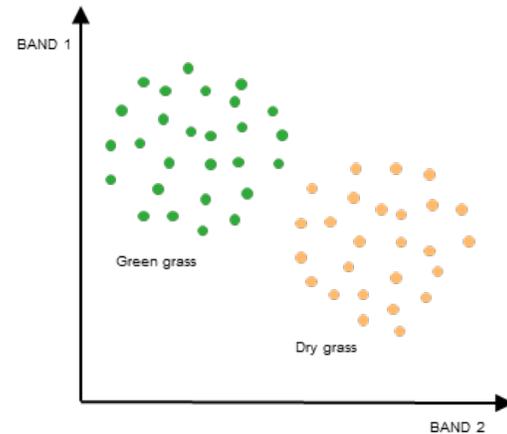
- Conditions of **compact and well-separated cluster** are necessary for effective **automatic classification phase**

SUPERVISED CLASSIFICATION

Access classifier performance

- If annotated samples are available, the classifier parameters are learned in a supervised way

- How to estimate the generalization error: split the **groundtruth** into three disjoint sets
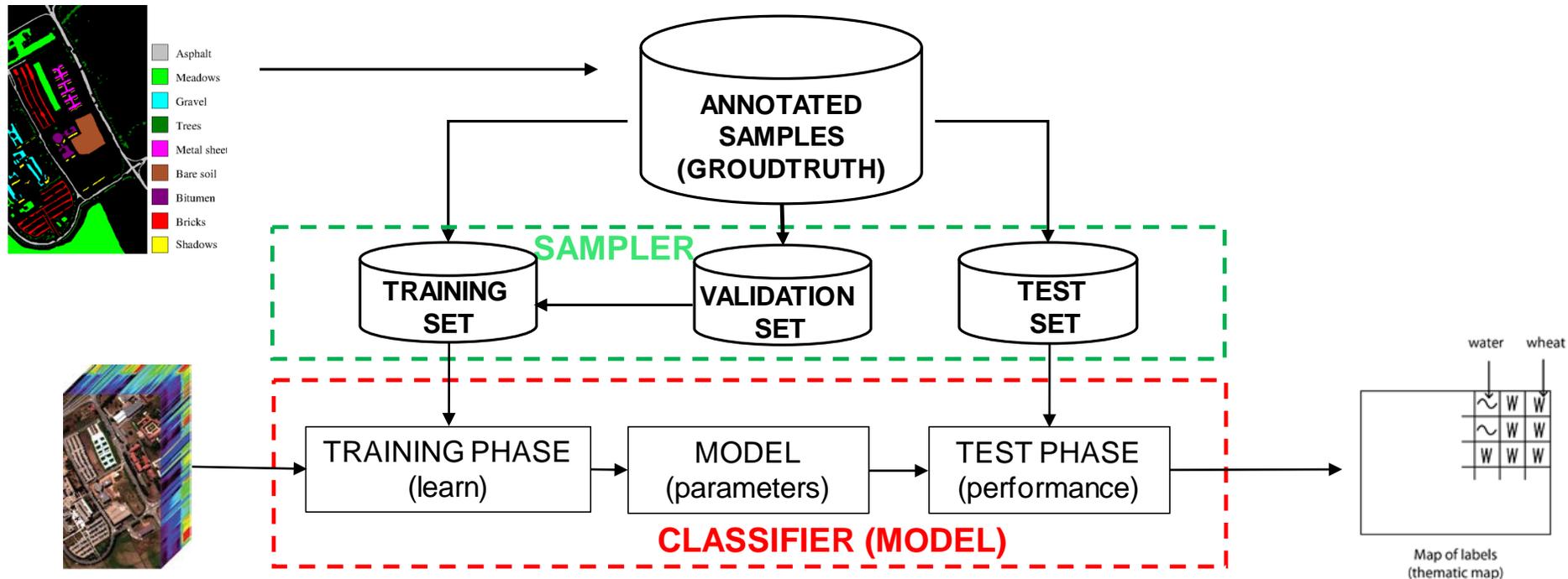
- **Performance**: usually more influenced by the amount and quality of the training samples (i.e., sampling design) rather than the classifier/model complexity

TRAINING, VALIDATION AND TEST SETS

- **Training set**: used to train the model

  – How do we ensure that the model is not overfitting to the data in the training set?

- **Validation set**: used to validate the model during training

  – Its classification is based only on the model that is learnt from the training set

  – Help to adjust the hyperparameters (e.g., number of hidden layers, learning rate, etc..)

- **Test set**: used to test the model after it has been trained

TRAINING SET

VALIDATION SET
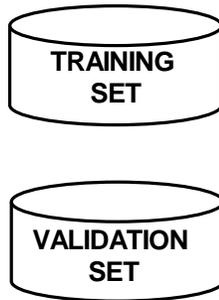
```
Epoch 7/10
2152/2152 [==============================] – 34s 16ms/step – loss: 0.2539 – acc: 0.9021 – v
al_loss: 0.4802 – val_acc: 0.8374
Epoch 8/10
2152/2152 [==============================] – 34s 16ms/step – loss: 0.2087 – acc: 0.9198 – v
al_loss: 0.4097 – val_acc: 0.8589
Epoch 9/10
2152/2152 [==============================] – 34s 16ms/step – loss: 0.2465 – acc: 0.9080 – v
al_loss: 0.8998 – val_acc: 0.7405
Epoch 10/10
2152/2152 [==================    ======] – 34s 16ms/step – loss: 0.3857 – acc: 0.8568 – v
al_loss: 0.4116 – val_acc: 0.8549
```

EPOCHS AND ITERATIONS

- **1 Epoch**: **entire training** set passed forward and backward through the network in once
  - The training set is divided in batches since the data can be too large
- **1 iteration**: **entire batch** passed forward and backward through the network in once
  - If 1000 training samples and batch size set to 500, it means 2 iterations to complete 1 Epoch
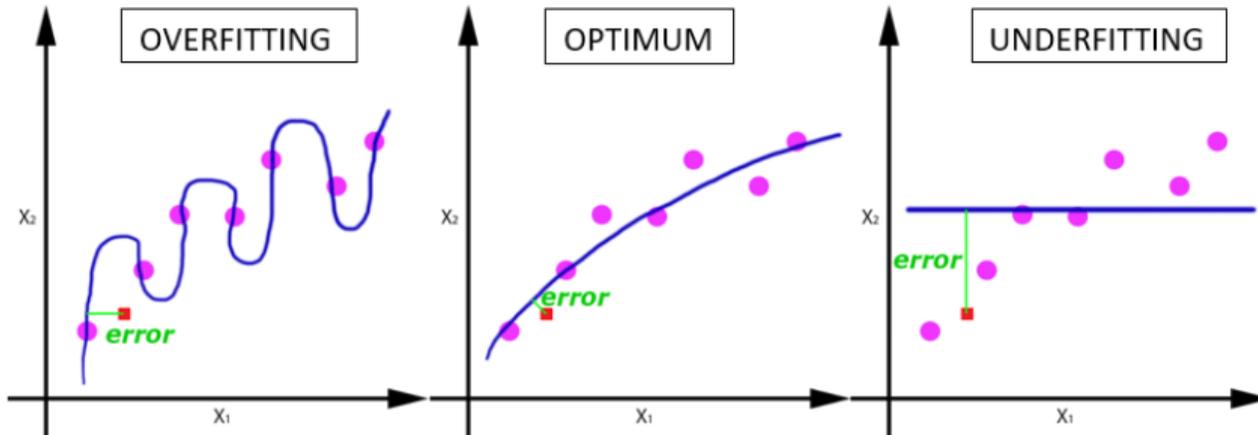

[23] Epoch vs Batch Size vs Iterations
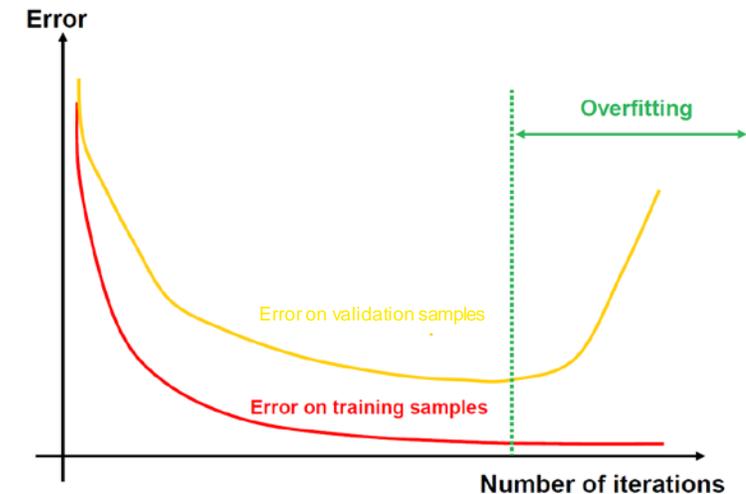
*What is the right numbers of epochs?*

GOOD PRACTICE WITH DATA

- Multiple images: train images and test images

- Large image: split image (left-right / top bottom)

- Small image: do not use it

- **Small image: use it correctly**

- It is important to prevent that testing samples were not seen during the training

The 2D receptive field of a CNN can involuntarily include samples from the test set, making the network overfit and biasing the evaluation
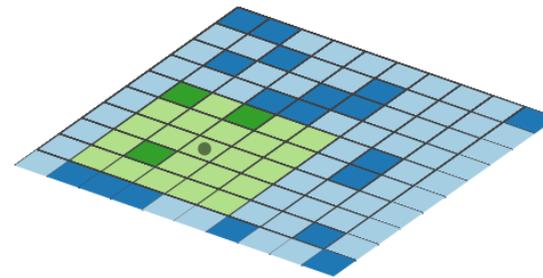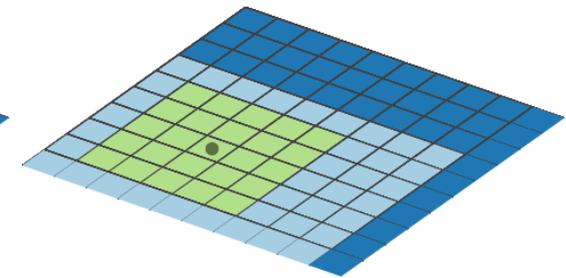
Random train/test   Disjoint train/test

Train   Test   Receptive field   Test pixel in receptive field   •: center pixel

[24] N. Audebert et al.

SAMPLING STRATEGIES

- **Random sampling:**
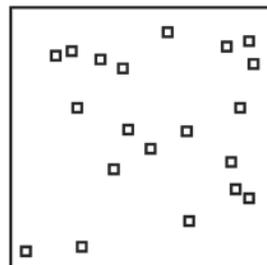 - Randomly select training samples within the area of each class
  - Often used, but bad idea if generalization required

- **Patch sampling**:
  - Image is divided into blocks, test samples are from blocks that haven't been used for training

- **Cluster sampling:**
  - Train on one area, test on another area
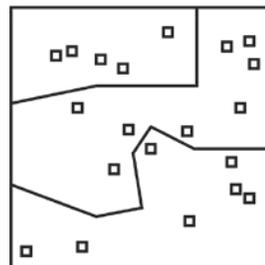
Random sampling   Stratified random sampling   Systematic sampling   Cluster sampling

GENERALIZATION

- Randomly sampling over the whole image is not a realistic use case

- It is a poor indication of generalization power

- Neighboring pixels will be highly correlated, which means that the test set will be very close to the train set
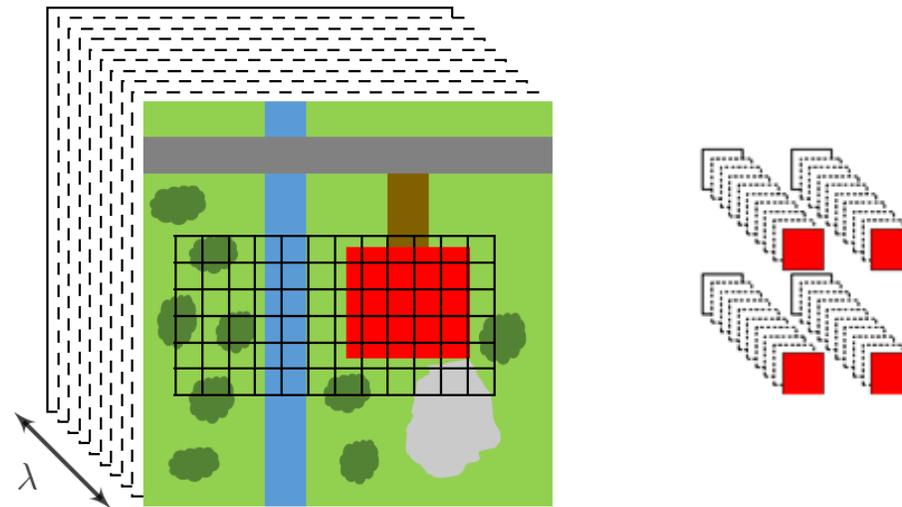
BENCHMARK SITE FOR REMOTE SENSING

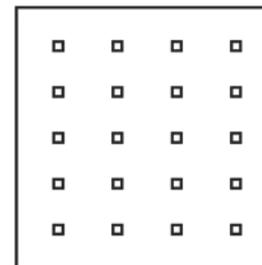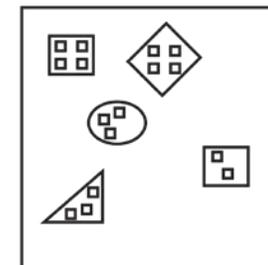- Well-defined train/test splits where the samples are extracted from significantly disjoint parts of the image

- Defined by the IEEE GRSS on the Data and Algorithm Standard Evaluation (DASE) website

[25] GRSS

- The ground truth is divided based on the connected components instead of the pixels, which allows the evaluation to actually measure how the model generalizes to new geo-entities

Train (random)   Test (random)   Train (disjoint)   Test (disjoint)

[24] N. Audebert et al.
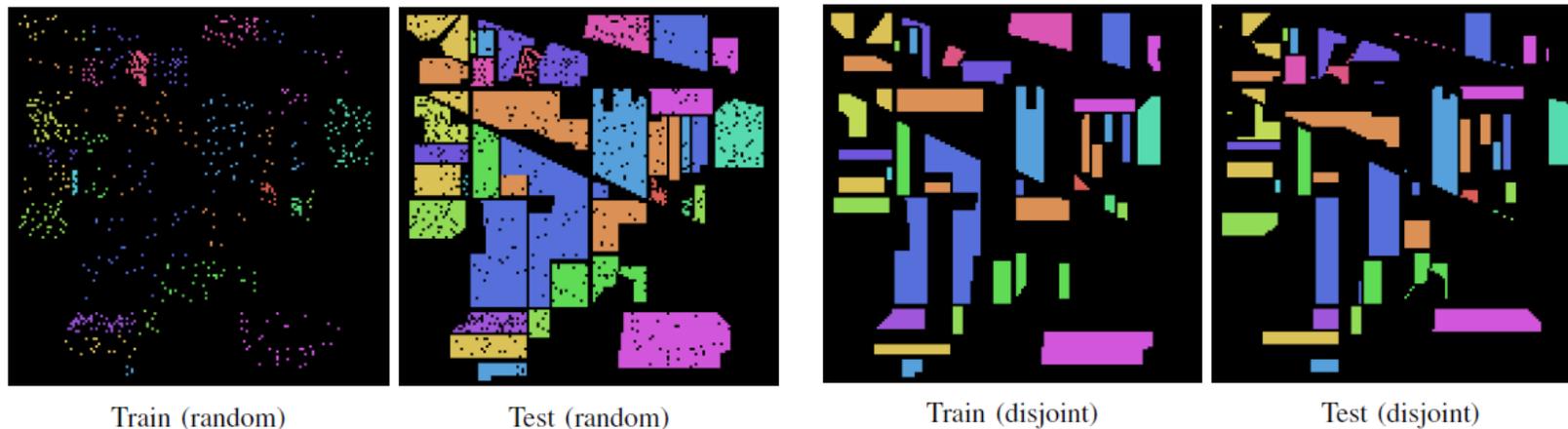
HOW TO COMPARE THE OBTAINED RESULTS WITH THE TEST DATA?

- **Photo interpretation**

  – E.g., visual comparison of classification maps

  – It can identify weaknesses of the classifier

Estimated class posterior
Blue = low probability; Red = high probability

[56] R. Hänsch

- **Metrics**

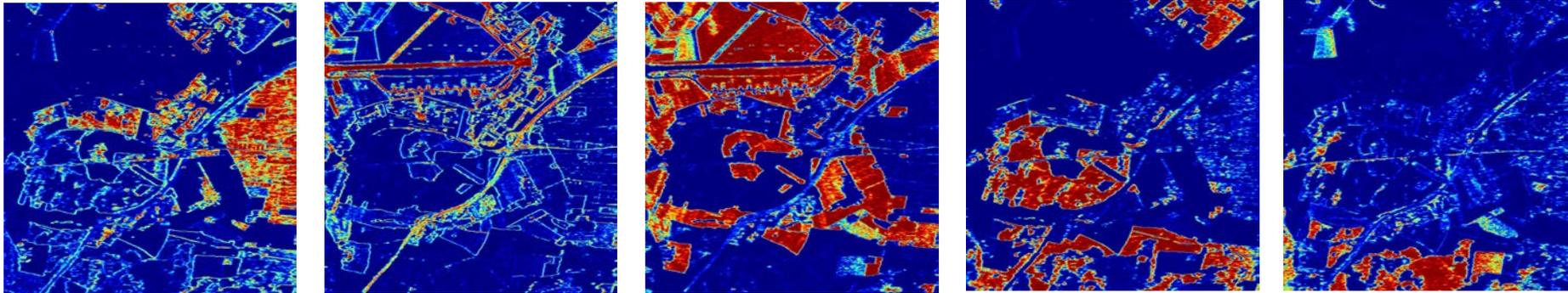 – Accuracy measures: overall, class-specific, average, kappa coefficient, F1 score, etc.

REFERENCES

- [1] CORONA: American's First Satellite Program: first photograph

  Online: https://www.oneonta.edu/faculty/baumanpr/geosat2/RS%20History%20II/RS-History-Part-2.html

- [2] The Earth-Atmosphere Energy Balance

  Online: http://theatmosphere.pbworks.com/w/page/27058542/The%20Earth-Atmosphere%20Energy%20Balance

- [3] N. B. Chang and K. Bai, "Multisensor data fusion and machine learning for environmental remote sensing", in Multisensor Data Fusion and Machine Learning for Environmental Remote Sensing, 2018

  Online: https://doi.org/10.1201/b20703

- [4] Electromagnetic spectrum

  Online: https://www.computerhope.com/jargon/e/elecspec.htm

- [5] K. Tempfli, N. Kerle, G. C. Huurneman, L. L. F. Janssen, Principles of Remote Sensing: An Introductory TextBook

  Online: https://webapps.itc.utwente.nl/librarywww/papers_2009/general/principlesremotesensing.pdf

- [6] AVIRIS Concept

  Online: https://aviris.jpl.nasa.gov/aviris/concept.html

- [7] Tree all four seasons illustrations

  Online: https://www.shutterstock.com/search/tree+all+four+seasons?image_type=illustration

- [8] Farm in China

  Online: http://factsanddetails.com/china/cat9/sub63/item348.html

- [9] Farm in California

  Online: https://www.nytimes.com/2012/10/14/magazine/californias-central-valley-land-of-a-billion-vegetables.html

- [10] Paris, C., Bruzzone, L., & Fernandez-Prieto, D. (2019). A Novel Approach to the Unsupervised Update of Land-Cover Maps by Classification of Time Series of Multispectral Images. IEEE Transactions on Geoscience and Remote Sensing. https://doi.org/10.1109/TGRS.2018.2890404

REFERENCES

- [11] Satellite Images Visual Interpretation Online Training Course

  Online: https://www.youtube.com/watch?v=eDND1sGMqF8

- [12] F. Ömrüuzun, B. Demir, L. Bruzzone and Y. Y. Çetin, "Content based hyperspectral image retrieval using bag of endmembers image descriptors," 8th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Los Angeles, CA, 2016, pp. 1-4.

- [13] Schematic Presentation of the Interpretation Process. Source: Albertz 2007 with modifications

  Online: https://seos-project.eu/remotesensing/remotesensing-c04-p01.html

- [14] P. Helber, B. Bischke, A. Dengel, and D. Borth, "EuroSAT: A Novel Dataset and Deep Learning Benchmark for Land Use and Land Cover Classification," Computing Research Repository - arXiv, vol. abs/1709.0, 2017

- [15] G. Sumbul, M. Charfuelan, B. Demir, V. Markl, BigEarthNet: A Large-Scale Benchmark Archive for Remote Sensing Image Understanding, IEEE International Conference on Geoscience and Remote Sensing Symposium, Yokohama, Japan, 2019.

- [16] Remote Sensing for All Sectors

  Online: https://www.zu.ac.ae/main/en/colleges/colleges/__college_of_natural_and_health_sciences/research/_hyperspectral/_pages/Initiatives.aspx

- [17] Sentinel-2 Satellite Imagery

  Online: http://geocento.com/satellite-imagery-gallery/sentinel-2/

- [18] CORINE Land Cover

  Online: https://land.copernicus.eu/pan-european/corine-land-cover

- [19] Normalized Difference Vegetation Index (NDVI)

  Online: http://www.agasyst.com/portals/NDVI.html

- [20] Future US Land Imaging Jeffrey Masek, NASA GSFC

  Online: https://slideplayer.com/slide/12177269/

REFERENCES

- [21] Data Preparation and Feature Engineering for Machine Learning course

  Online: https://developers.google.com/machine-learning/clustering/prepare-data

- [22] Common Data Processing Pipeline

  Online: http://cs231n.github.io/neural-networks-2/

- [23] Epoch vs Batch Size vs Iterations

  Online: https://towardsdatascience.com/epoch-vs-iterations-vs-batch-size-4dfb9c7ce9c9

- [24] Audebert, N., Saux, B., and Lefèvre, S. (2019). Deep Learning for Classification of Hyperspectral Data: A Comparative Review. https://doi.org/10.1109/mgrs.2019.2912563

- [25] GRSS Data and Algorithm Standard Evaluation website

  Online: http://dase.grss-ieee.org/

- [26] R. Hänsch, A. Ley and O. Hellwich, "Correct and still wrong: The relationship between sampling strategies and the estimation of the generalization error," 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, TX, 2017, pp. 3672-3675.