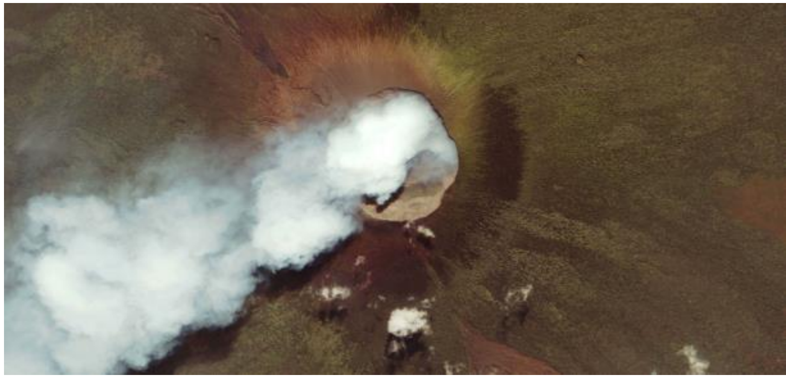
Satellite Image Indexing and Retrieval

Henri Maître
Télécom ParisTech

Why do we need Remote Sensing

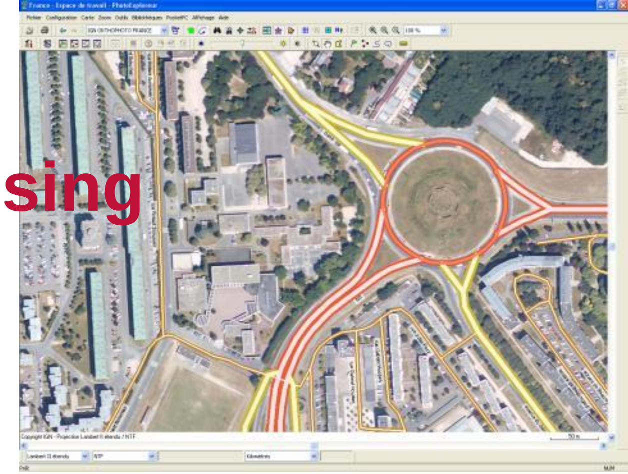
- **Environnement:**
  - Meteorology: short-term weather prediction
  - Climate: long-term monitoring
  - GMES = Global Monitoring for Environment and Security: survey of natural and man-made catastrophies
    - volcanos
    - earthquake, tsunamis, floods
    - Industrial hazards
    - Marine pollution



Why do we need Remote Sensing

- **Agriculture :**
  - Survey and evaluation of crop & farming production
  - Fish & Aquaculture resource management
  - Forestry resource planning
  - Water management, dams, watering
  - Desertification & urban pressure

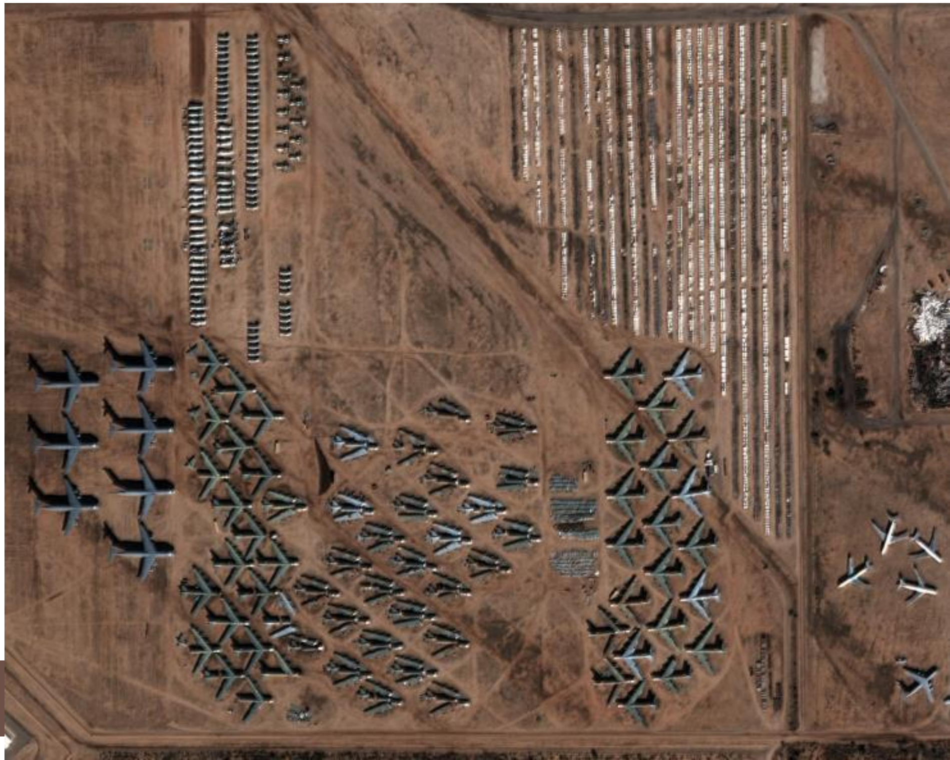
Why do we need Remote Sensing



- **Town & country planning:**
  - Mapping and inventories
  - Constructions & public work: trains, airports, harbours, dams
  - Cities and Mega-cities management
  - Management of moving populations, displacements, installation
  - Climatic impact management
  - Crisis Management: fires, floods, …

Why do we need Remote Sensing

- **Defence & Security applications:**
  - Military deployment preparation
  - Intelligence and survey of national territory



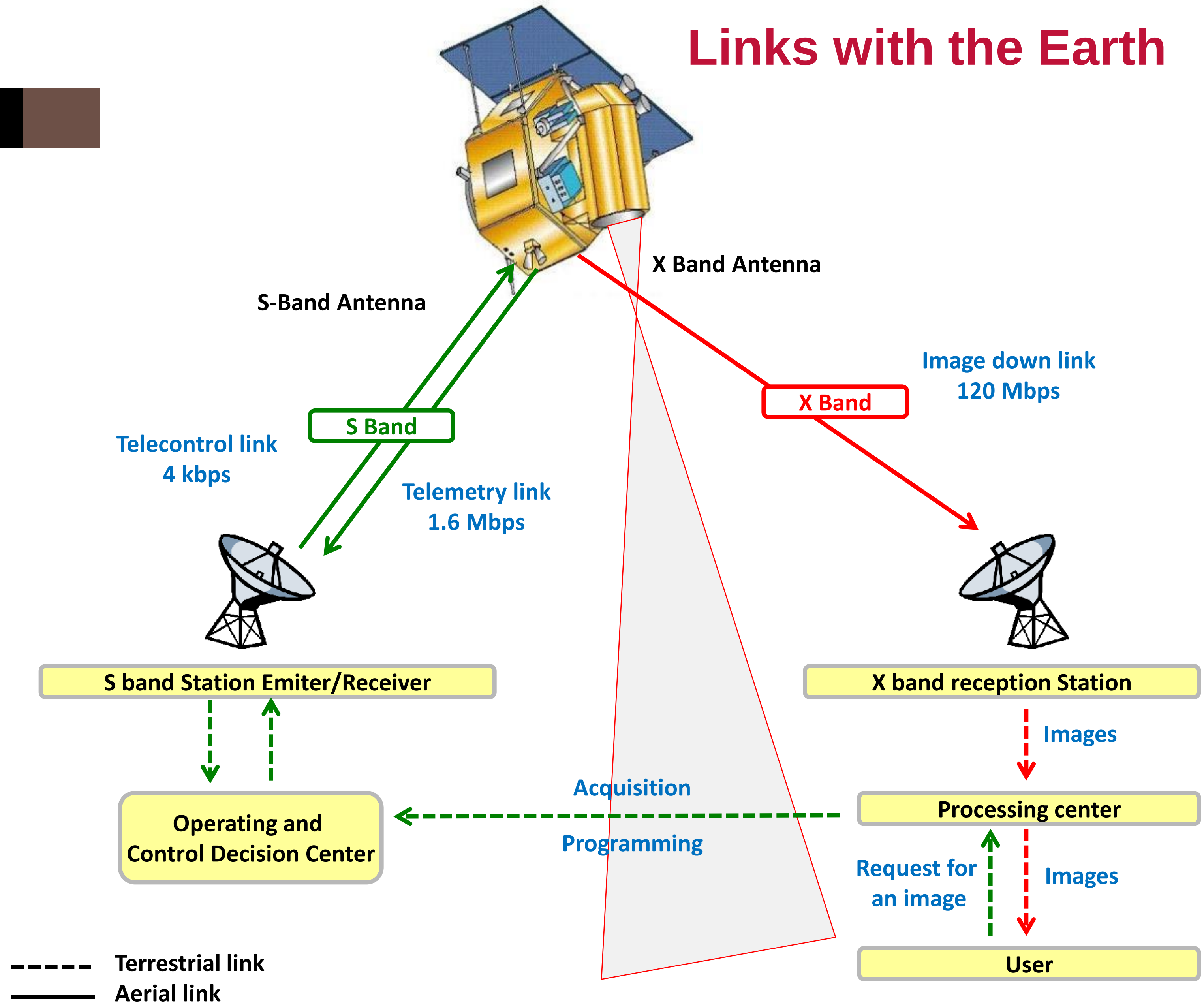
How is prepared a remote sensing program

How is prepared a remote sensing program

- **Conceive the sensor:** application, customers
- **Determine which satellite / which booster**
- **Conceive the ground-station and the data management process :** economical, social and business issues

➔ 15 to 20 years

Links with the Earth

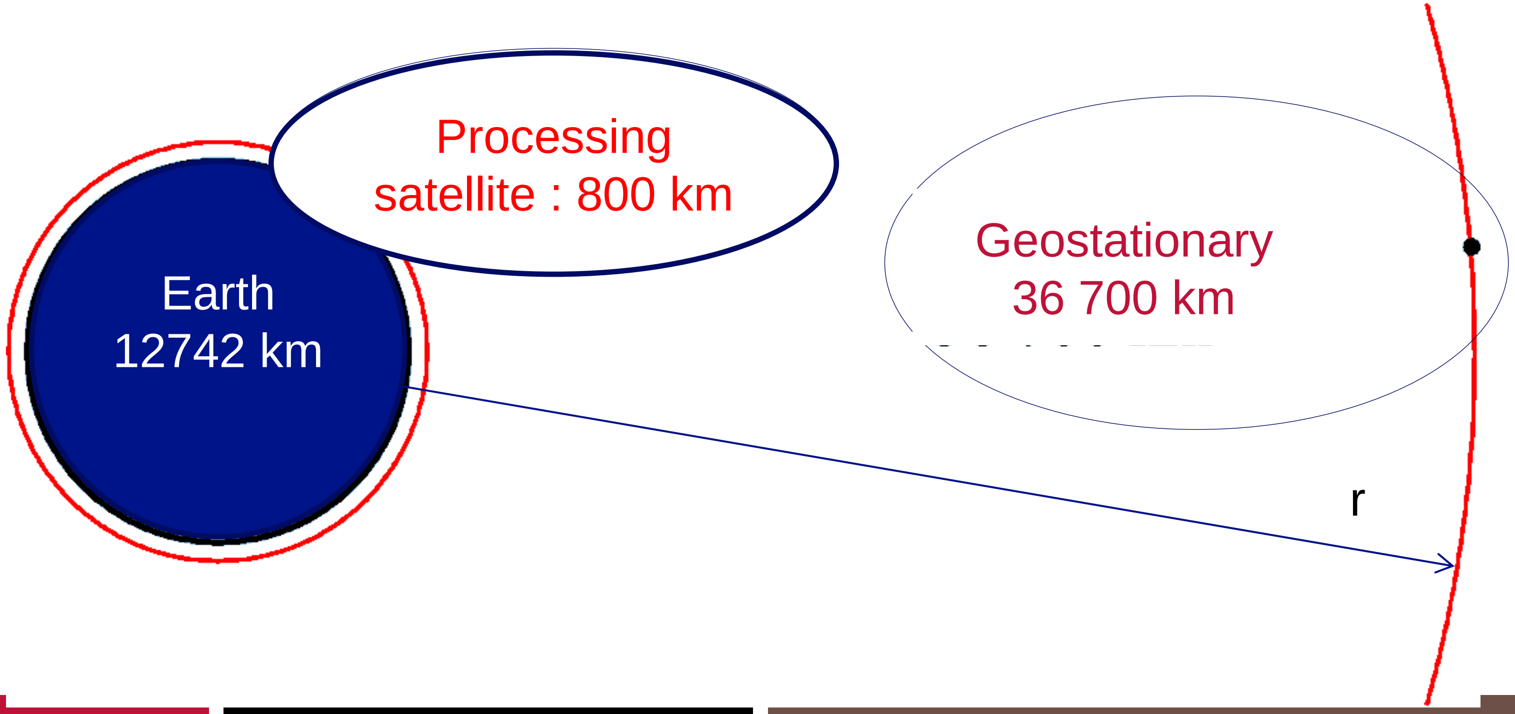


Satellite : orbit choice

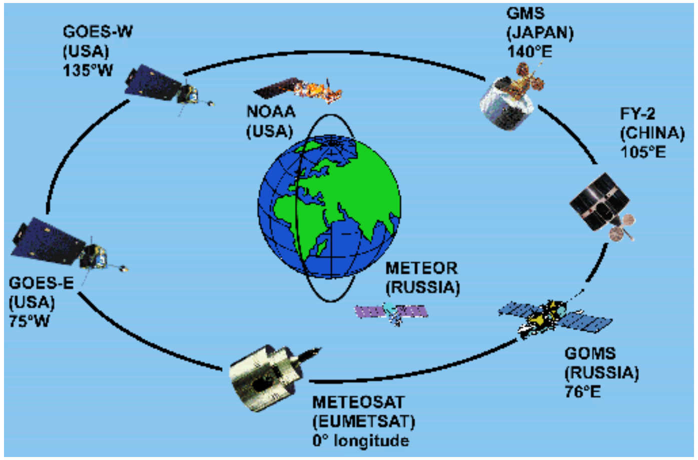
$$\vec{F} = -\frac{\mu\, m}{r^2}\frac{\vec{r}}{r}$$

- **Mecanics laws:**
  - Newton = centripedal force
  - Satellite speed = driving force

    ➔ elliptical or circular trajectory (Kepler)

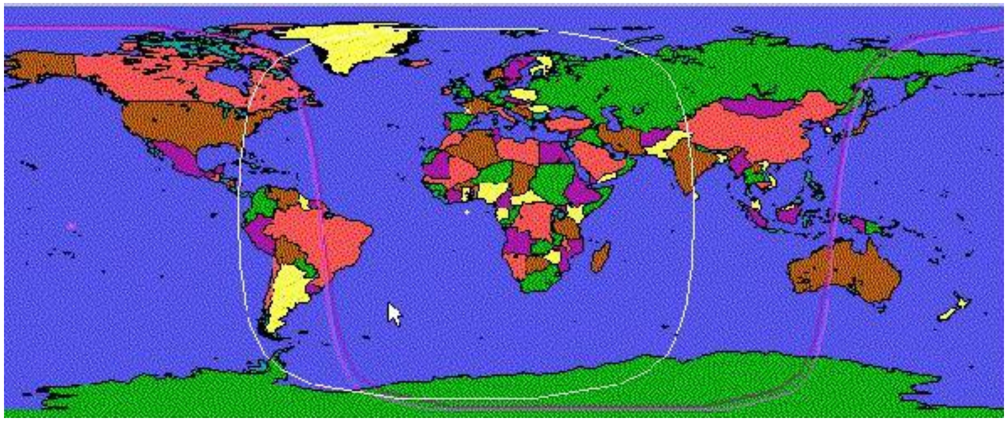


Orbit choice



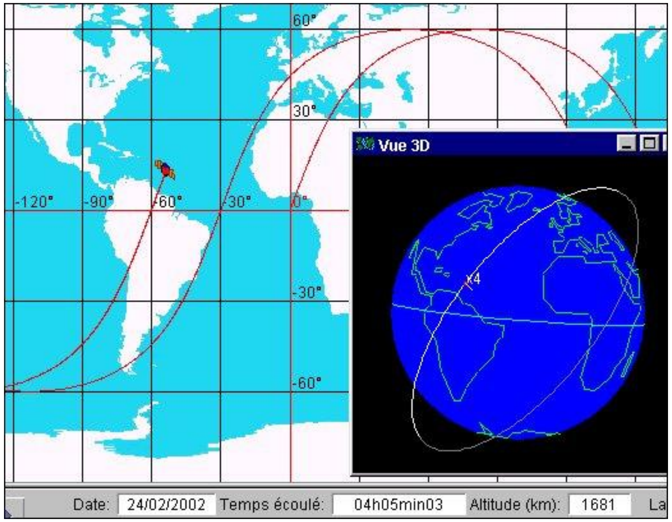
- **1) Geostationary**
  - In the Equator Plane
  - Always at vertical of the same point on the Equator
  - Altitude ~ 36 700 km
  - Field of view: ~1/3 Earth: always the same

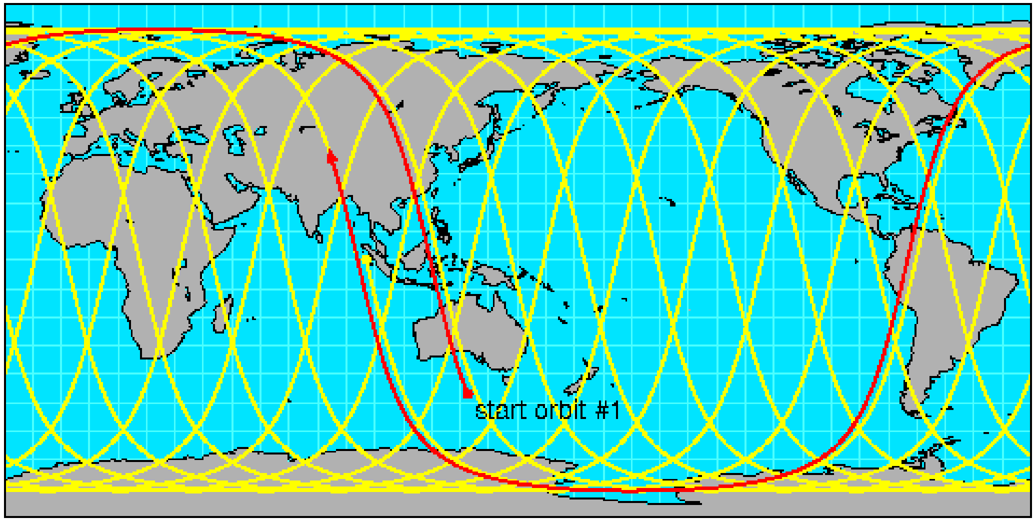
  - Applications : **meteo, survey of catastrophies**, telecoms, TV



Orbit choice

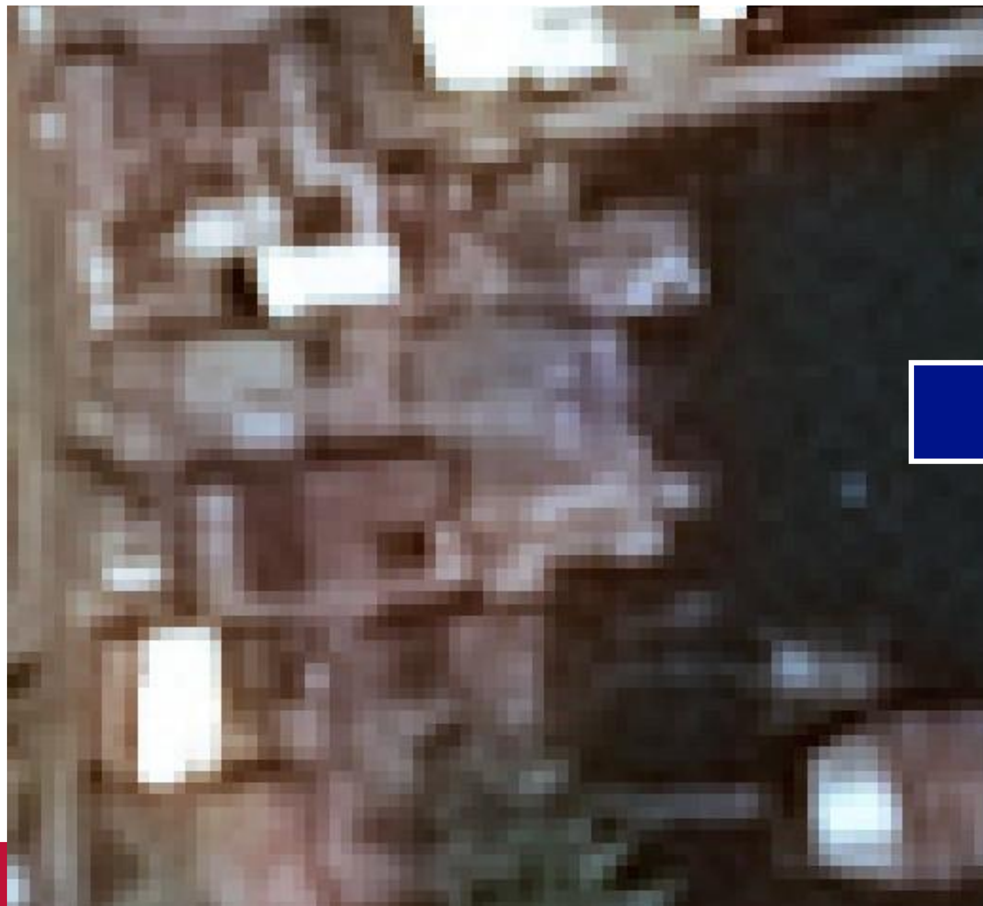
- **2) Processing satellite (low orbit)**
  - Altitude ~ 800 km (down to 250 km)
  - Circular ~ N/S
  - Trajectory : ± polar
  - ~ 15 revolutions / day
  - Heliosynchrons





Choice of resolution

- **Pixel size = smallest measured piece of ground**
  - from 30 cm to 10 km



SPOT 5

Δx = 2,5m

On Ground resolution

- **Depends on:**
  - Sensor :

    Photosites size: δx

    $G = \frac{f}{D}$ = enlargement

    $\Delta x = \frac{\delta x}{G}$ = smallest detail

  - The camera lens

    $\delta' x = \frac{\lambda f}{d}$ = diffraction limited resolution

$$\Delta x = \frac{\lambda f}{Gd}$$

➔ Smallest detail

f = focal lens
~ 1 m

D = satellite-Earth distance
~ 1 000 km = 10^6 m

λ = wave length
= 0,5 . 10^{-6} m

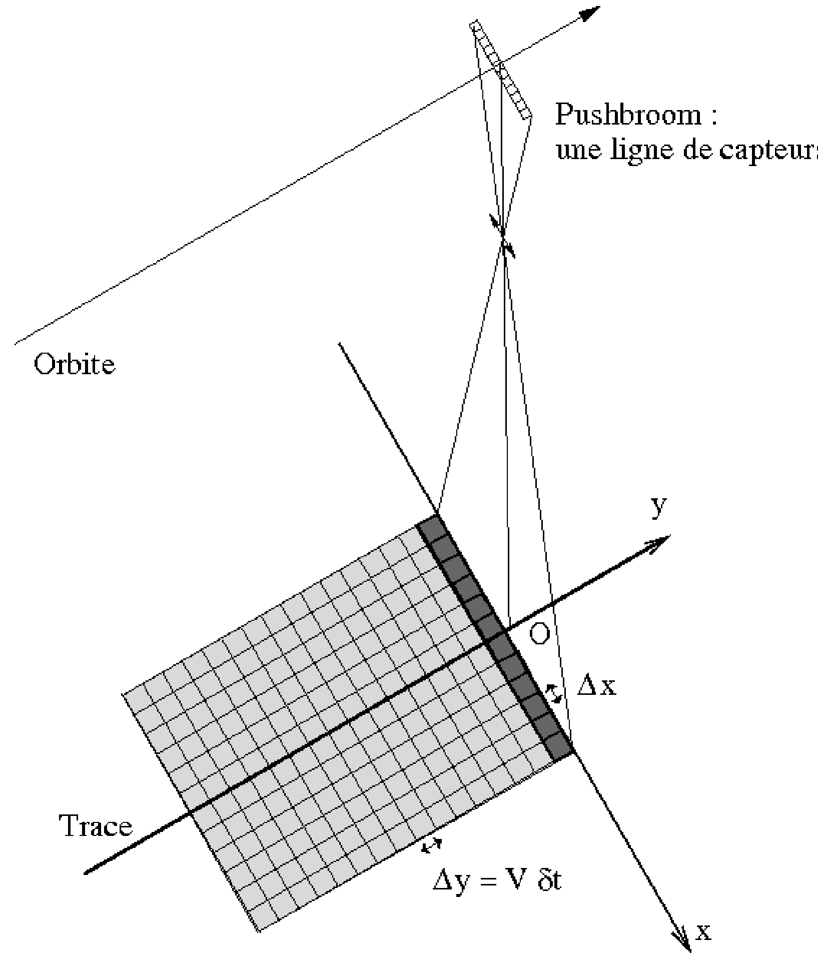
d = lens diameter
~ 0,5 m

$$\boldsymbol{\Delta x = 1\ m}$$

If the photosite measures 10^{-6} m

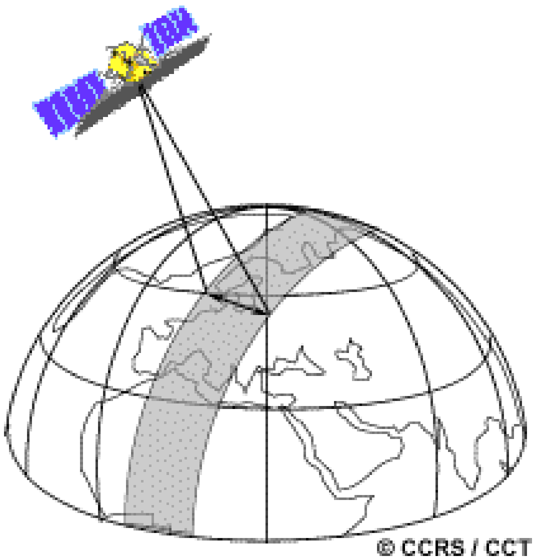
Often *push-broom* *sensor*

- **Sensor size along track:**
  - On line sensor
  - = speed x aperture time
- **In the other direction**
  - Number of sensors on a linene
  - from 6 000 to 40 000
- **Resolution :**
  - Depends on the lens

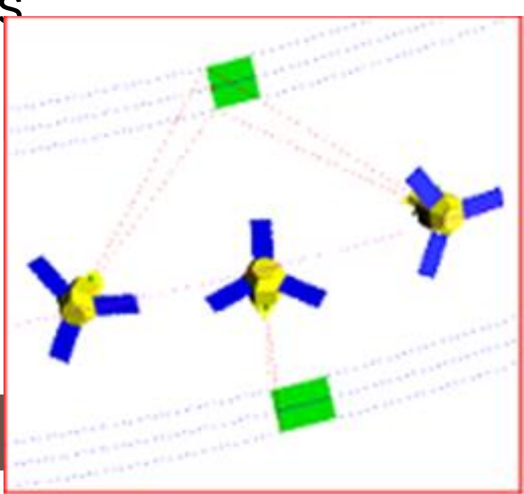


Swath choice

- **Swath = image width**
  - from 10 km to 10 000 km
  - = from 3 000 to 40 000 pixels / line
  - Given by the sensor size
  - Limited by the communication link with Ear

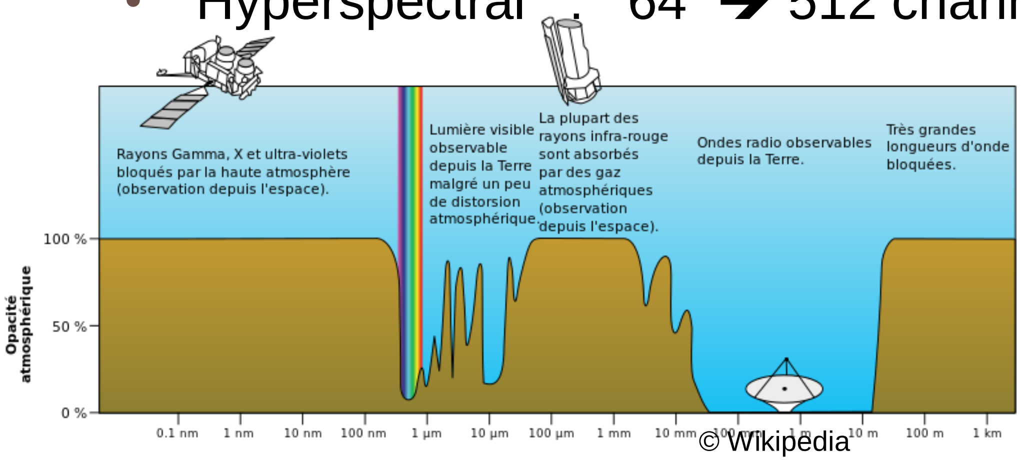


- **Revisit delay**
- **15 min for geostationnary sat. (to dump the memory)**
  - from 1h30 (min) to 1 mois for processing satellites
  - But … sensor agility!

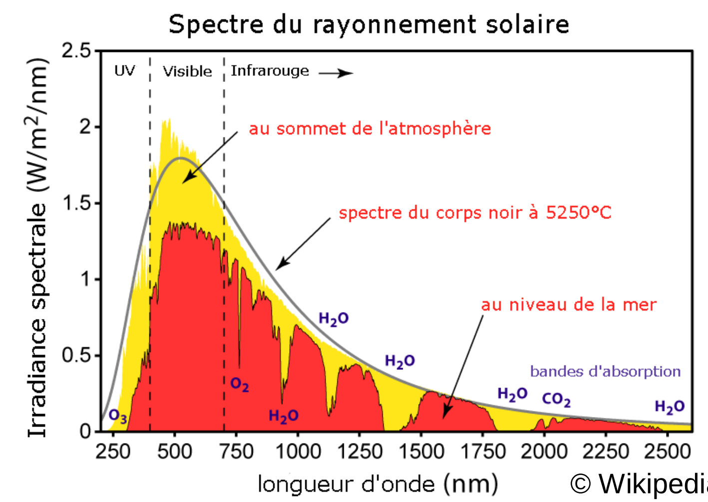


Which wave length?

- **1 – Passive sensors: measure the energy sent back from Sun by Earth or the energy radiated by Earth**
  - Emitted from the Sun (Wien's law) x Atmosphere transparency x Ground Reflexion
  - Black and White (Panchromatic)
  - Visible = Blue - Green - Red
  - Visible and Near Infra-Red : G - R - IR = false colors
  - Multispectral : 7 ➔ 20 channels
  - Hyperspectral : 64 ➔ 512 channels



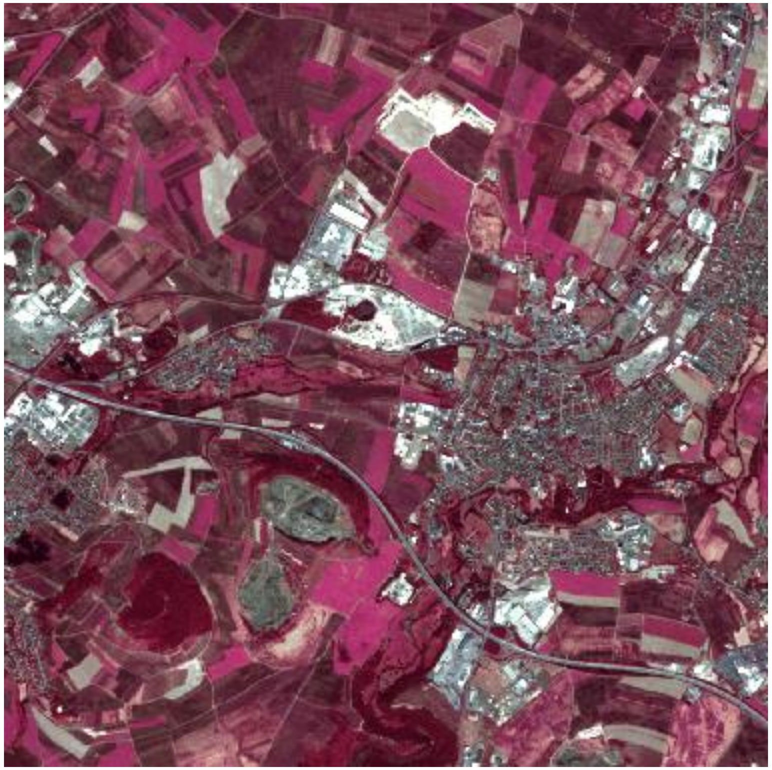
© Wikipedia



© Wikipedia

False colors : NIR-R-G → R-G-B

vegetation = red



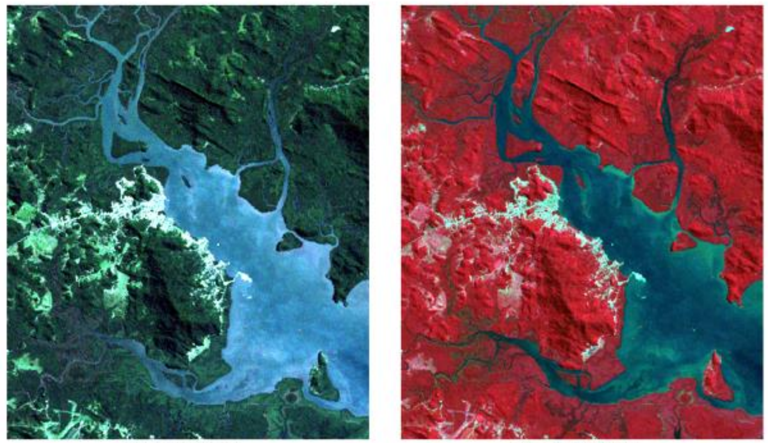
Multispectral image visualisation

Landsat = 7 channels

1
2 R
3
4 G
5
6 + B
7

41(7+5)

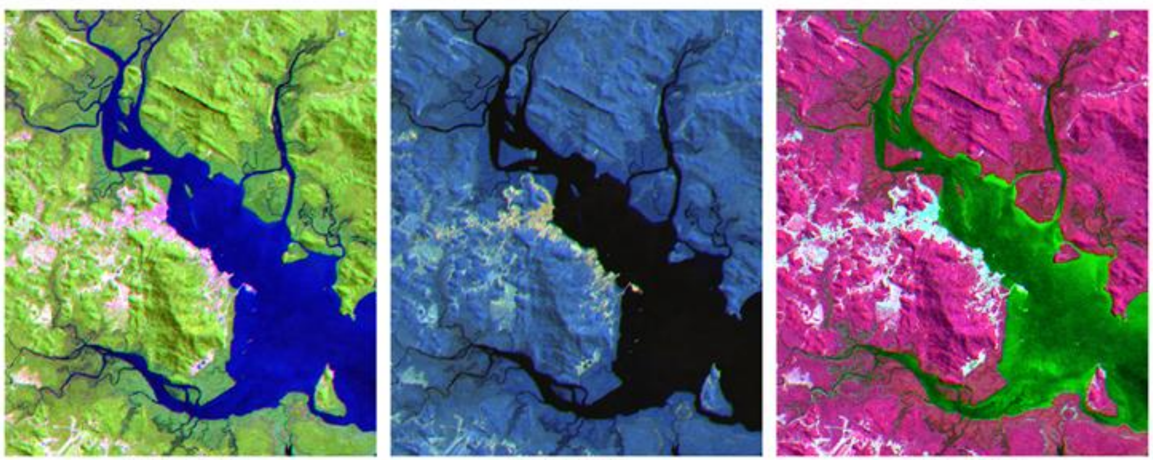
321



(a) combinaison 321 (b) combinaison 432

432

542



(c) combinaison 542 (d) combinaison 754 (e) combinaison 435

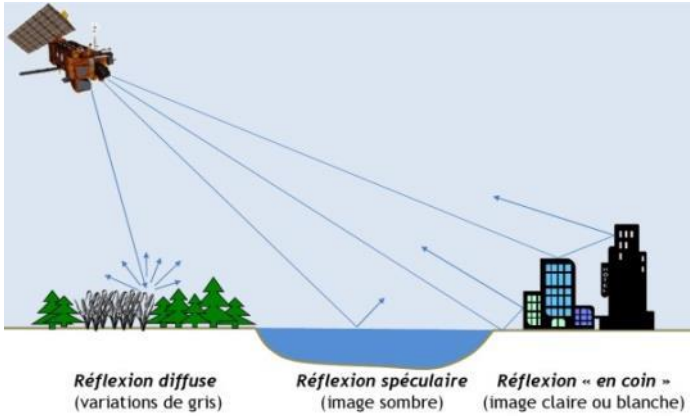
435

754

Which wave length?

- **2 – Active sensors: EM emiter + receiver**

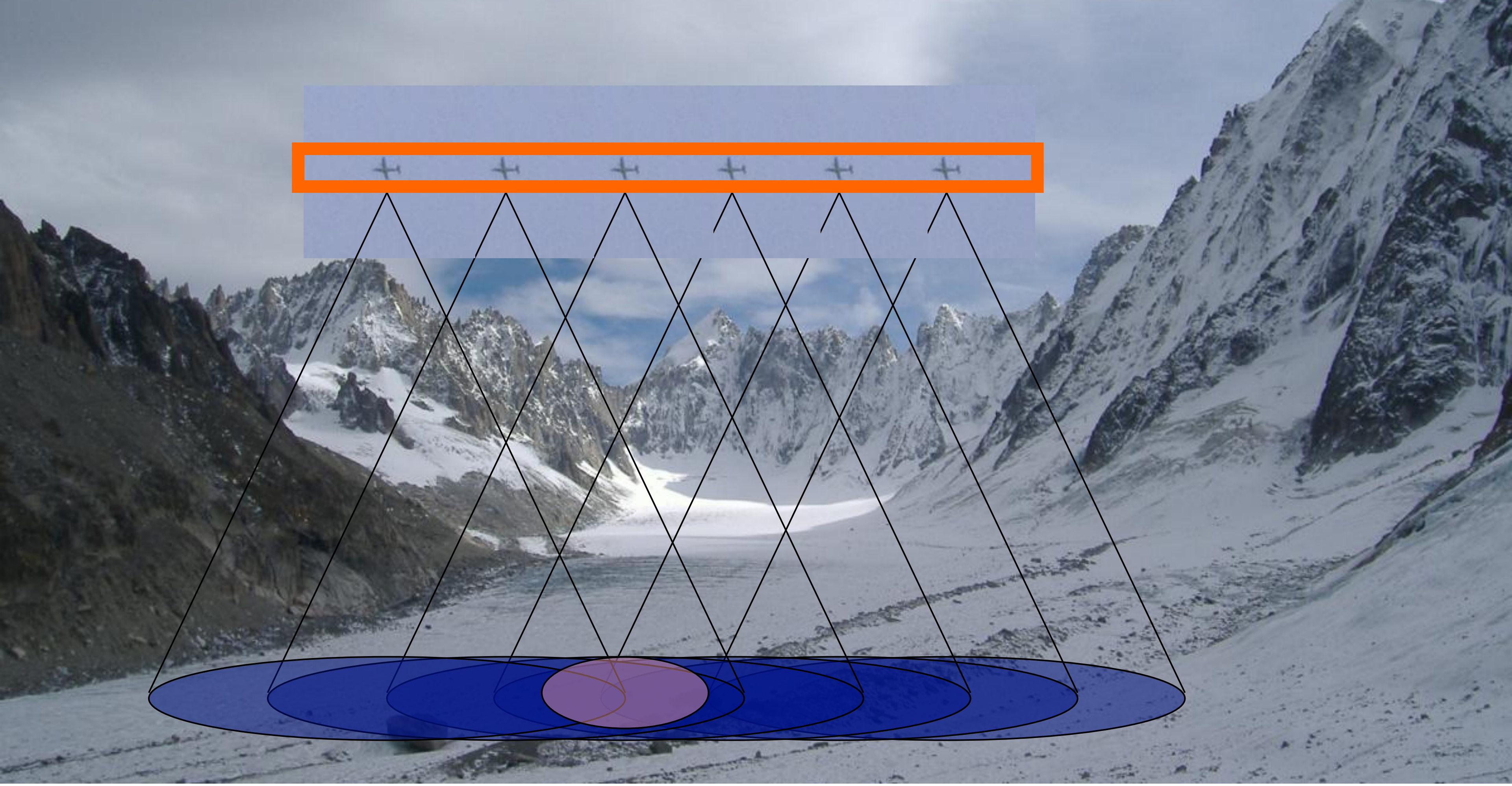
  radar = Micro waves: λ= 1 cm to 10 m



- But low resolution : $\Delta x = \dfrac{\lambda f}{Gd}$
- With complex processing: Synthetic Aperture Radar ➔ hi resolution

Real antenna is too small, it covers a very large field

When moving, superposition of the seen areas



One point is seen from several antenna positions

From computation we obtain an accurate information = synthetic antenna

Satellite images = big data !

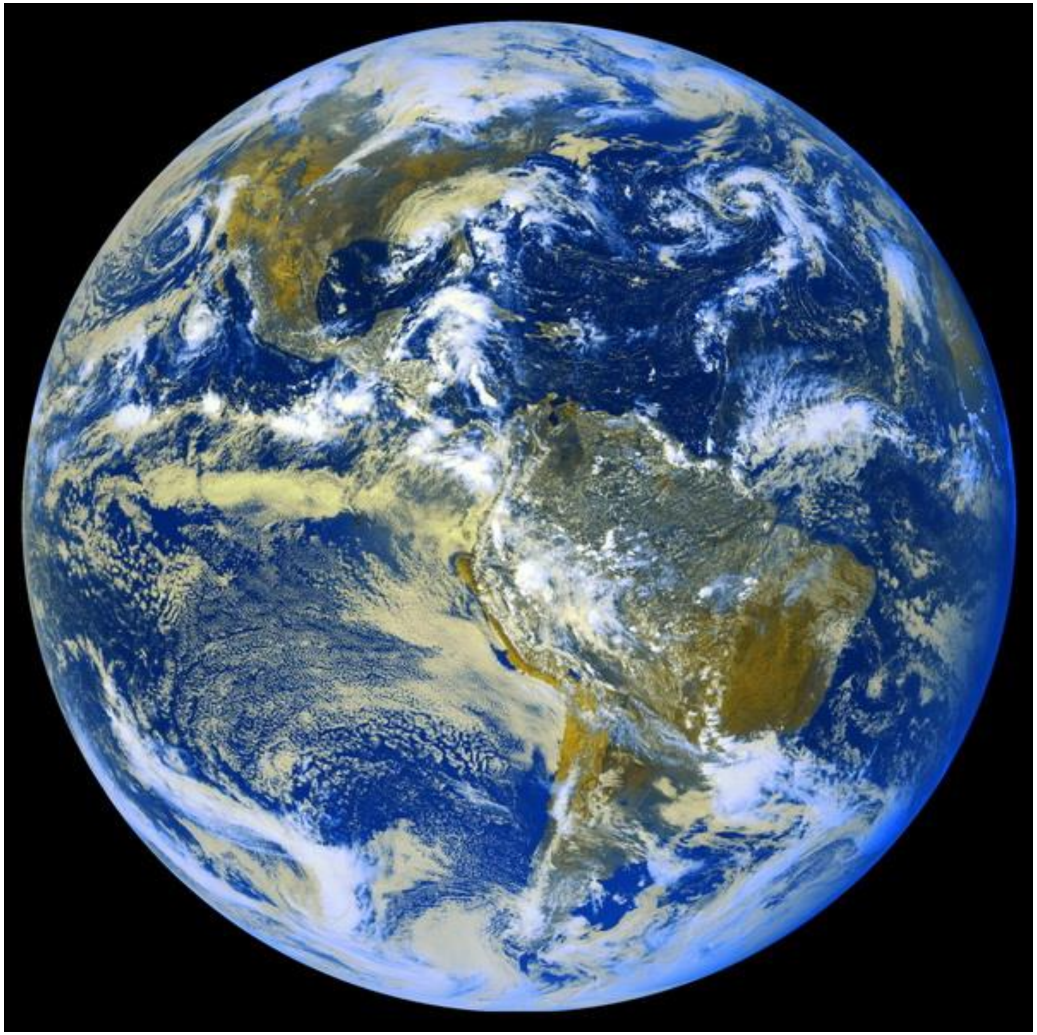
Television HD	1 280 x 720 pixels
Television 4k	4 000 x 2 000 pixels
PC display screen	1 600 x 1 200 pixels
Photo camera	5 000 x 4 000 pixels
Spot 1 … 4	**6 000 x 6 000 pixels**
SPOT 5	**24 000 x 24 000 pixels**
Quickbird	**40 000 x 40 000 pixels**

1 600 000 000 pixels = 1,6 Gpixels

= 800 PC display screens

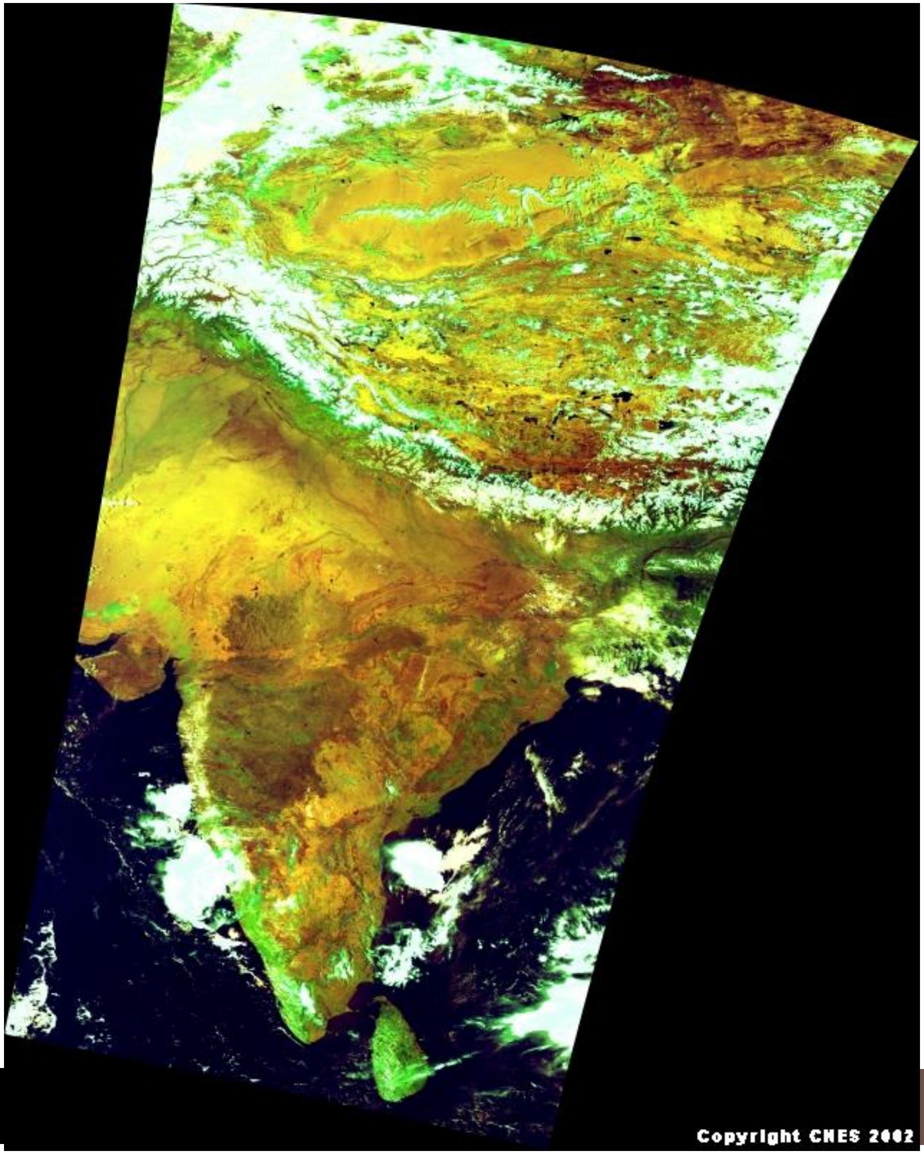
1 SPOT 5 image = 10 s of satellite run

Meteo satellites: very low resolution



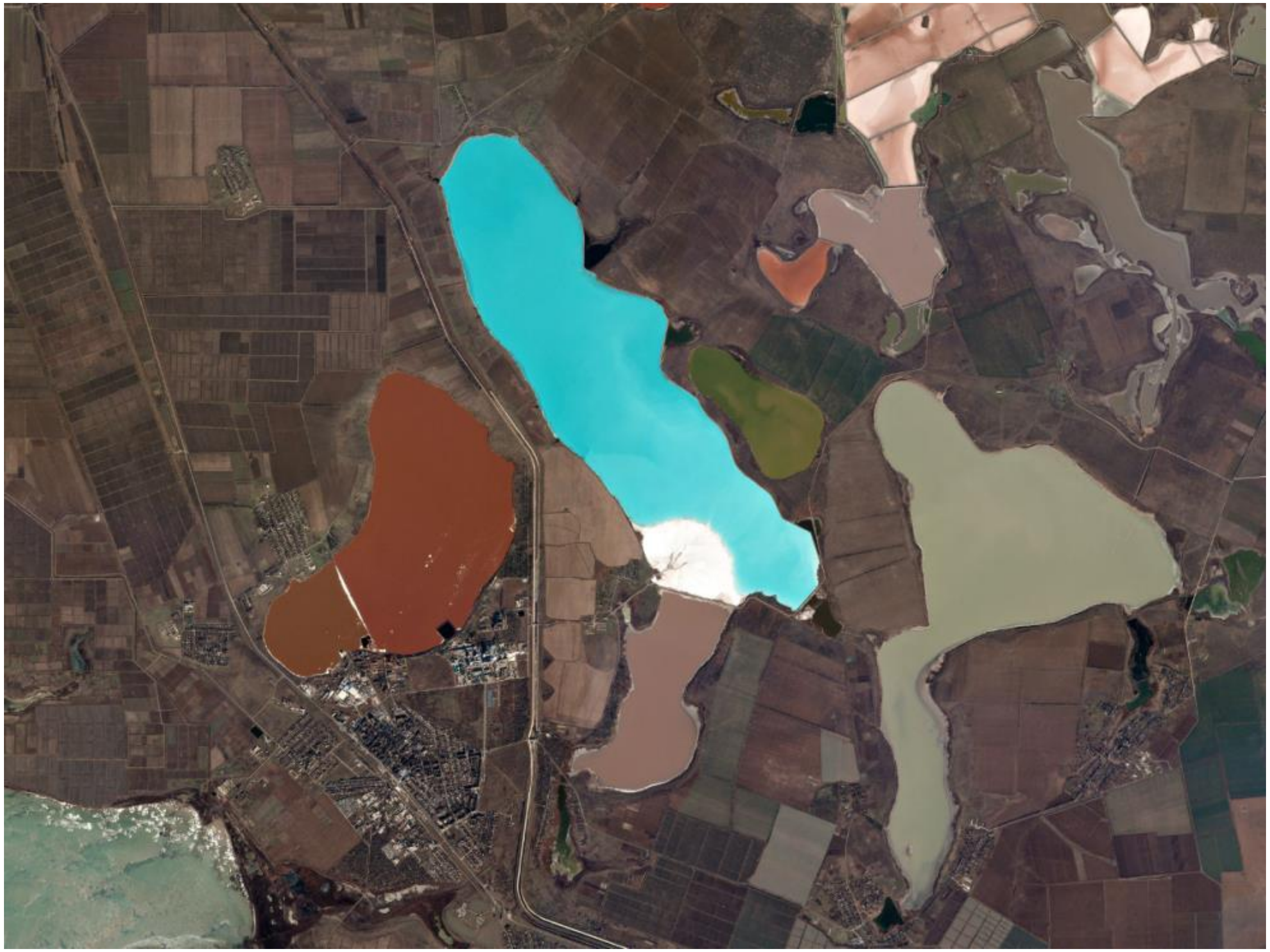
Meteosat = 3 km

Climate/environnement: low resolution



INSAT = 2,2 km

High resolution: Planetscope

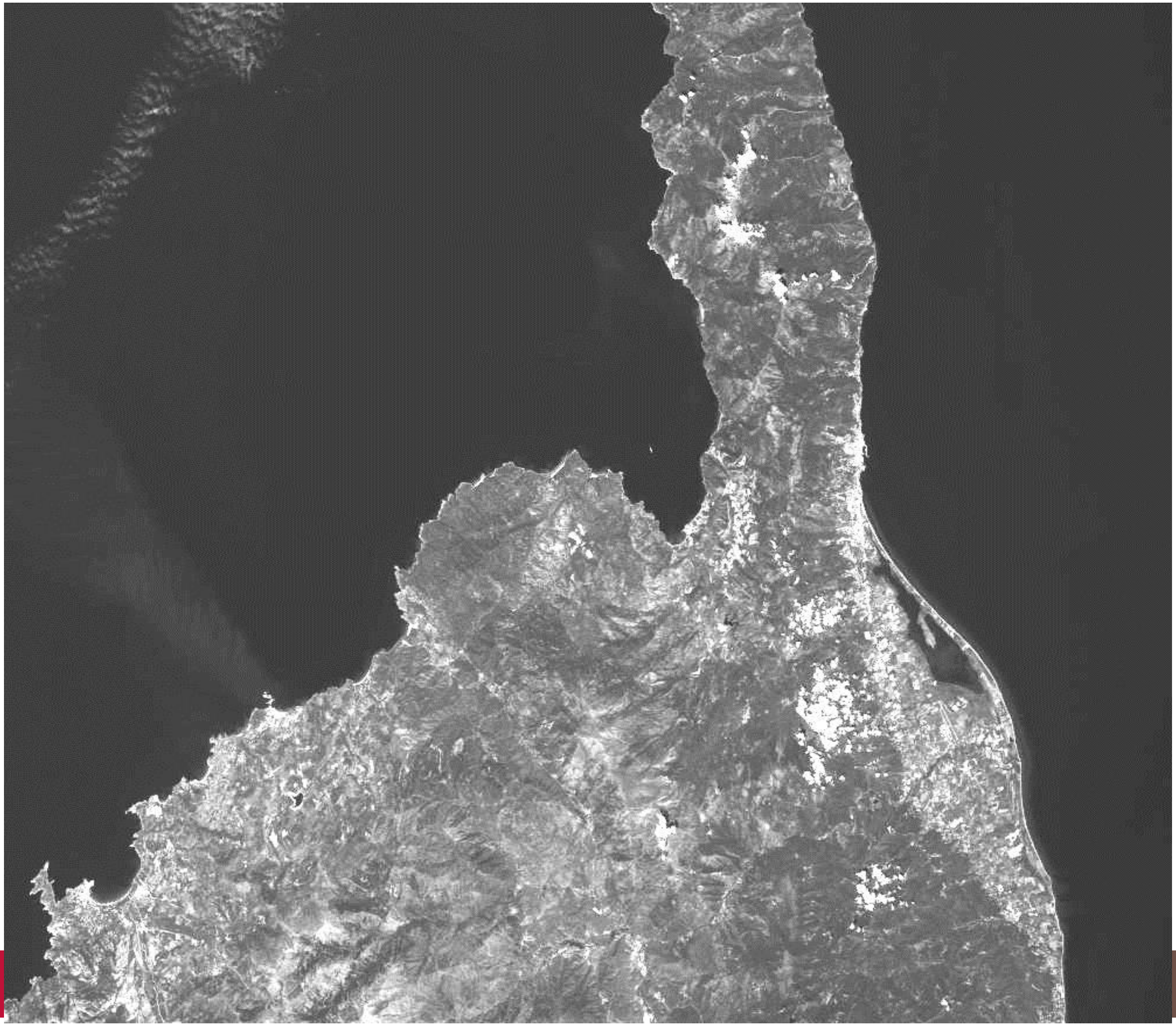


Planetscope :
Krasne Hypersalted Lake
- Crimea

Multispectral
= 4 m

175 satellites
300 Mkm^2 / day
= 2/3 Earth

SPOT 5 : high resolution ; pixel = 2,5 m



Very high resolution

Pléïades :
Bora-Bora

Panchro
= 0,70 m

Multispectral
= 2, 8 m

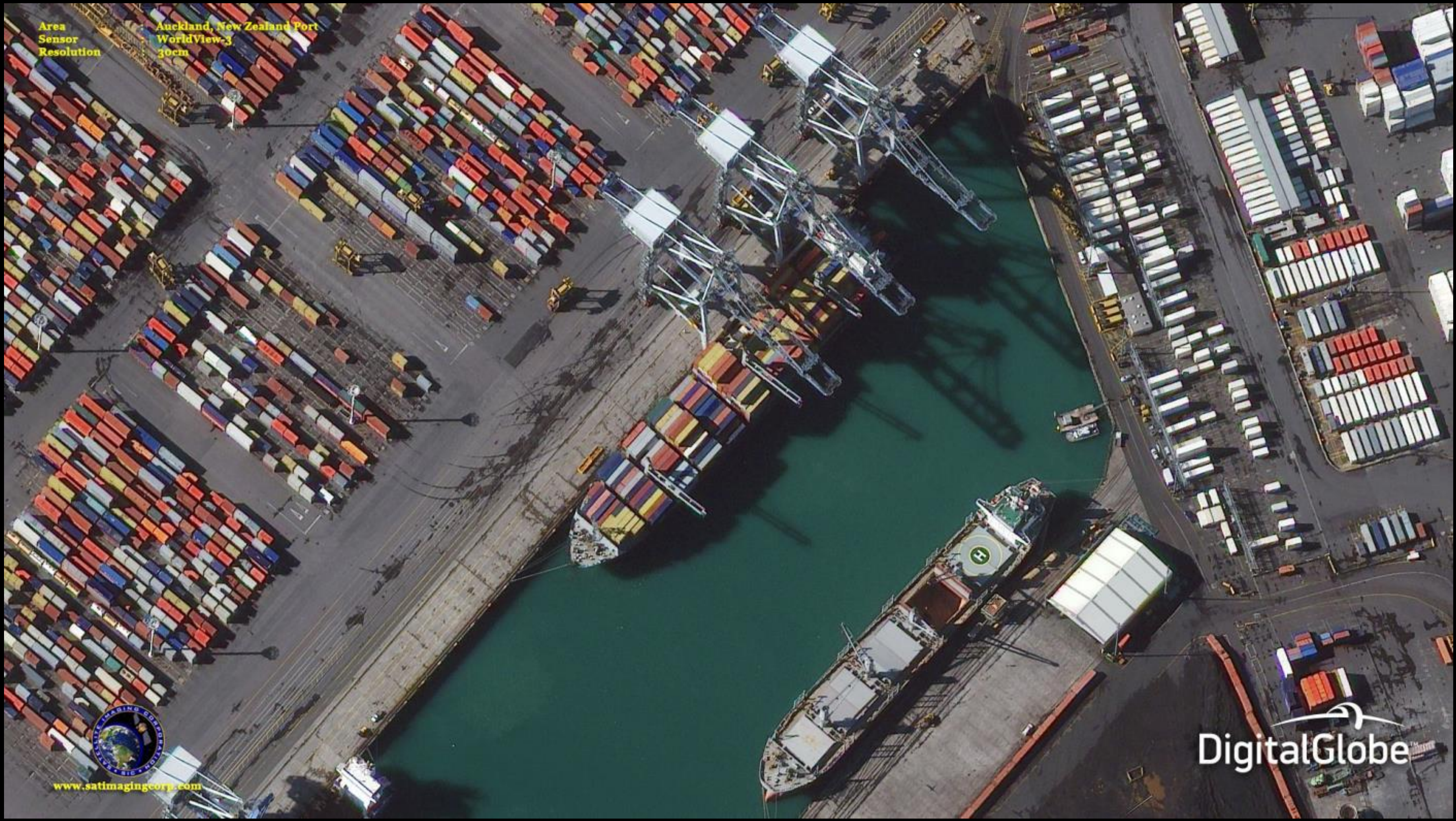
Very high resolution: Quickbird

Panchro
= 0,61 m

Multispectral
= 2, 4 m

Pléïades : Mont Saint Michel

Auckland New-Zealands



WorldView : King Abdullah Petroleum Center


Ikonos – Lebanon: agriculture

WorldView : Bayan Mines (China)

Temporal evolution: Baikal Lake with SPOT

Satellite Image archives

- How can we store millions of images?
- How can we ensure durability of storage?
- How knowing that information exists?
- How retrieving information?
- How exploiting information?

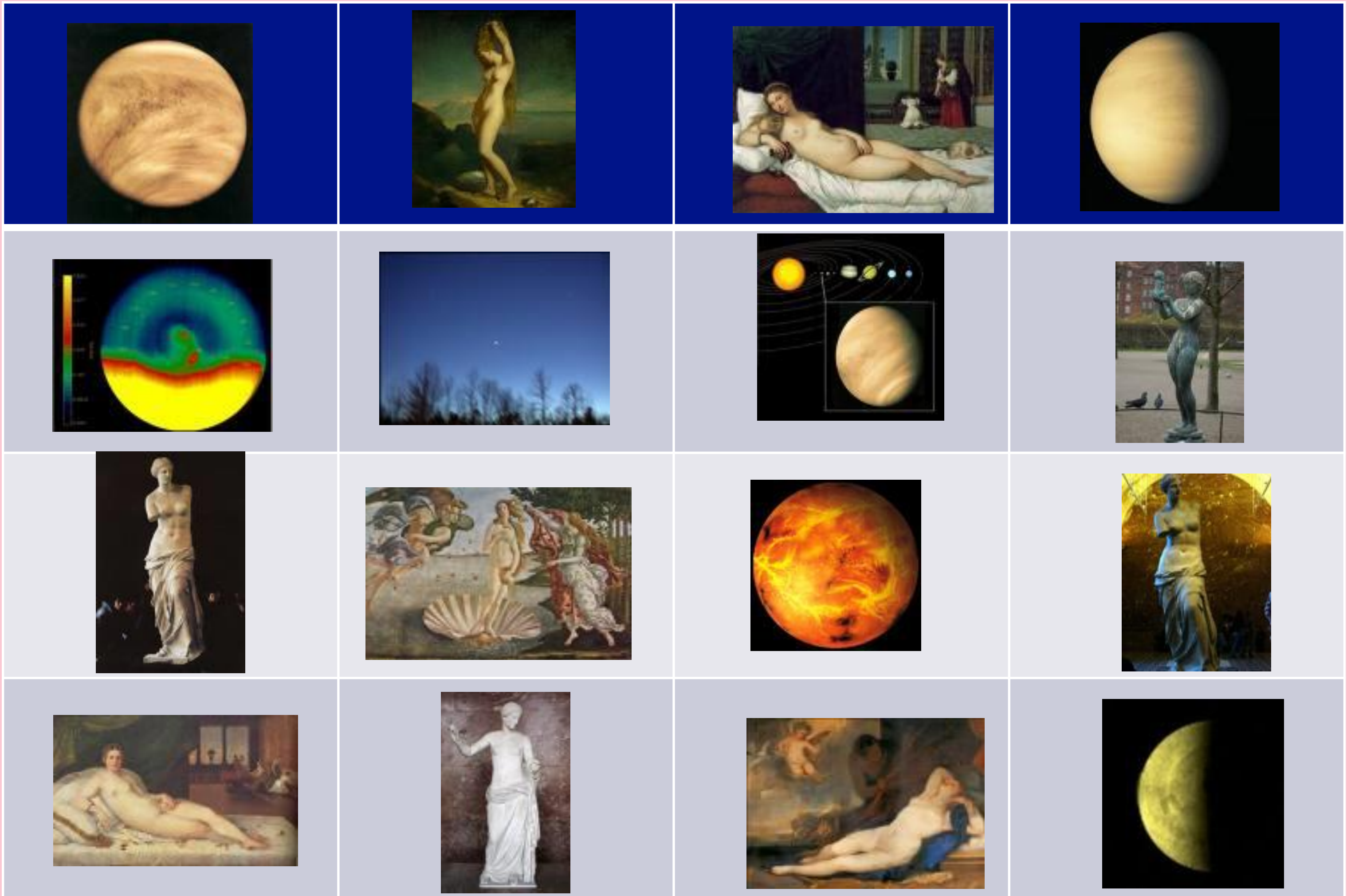
→ **Index images when receiving.**

→ **Then big-data and retrieval technics.**

Multimedia image indexing vs. Satellite image indexing

- **Multimedia information retrieval :**
  - Either from **semantic information**: name, description
    (90 % of Google-like retrieval)
  - Or from **instance** (i.e. with a reference image)
    (Face or fingerprint recognition)
  - Main choice : to be robust to some differences (scale, lighting, orientation, color, …) = **invariance**
  - Main strategy: detect invariant features
    - Histograms, area-based segmentation + color decomposition, graph description, …
    - Salient point detection: Harris, SIFT, SURF, …
  - Represent the image as a point in a *p* dimensional space $\mathbb{R}^p$

Ambiguous semantics: Venus



Textual categorisation

invariance

Salient points: SIFT

$$L(x,y,\sigma) = G(x,y,\sigma) * I(x,y),$$

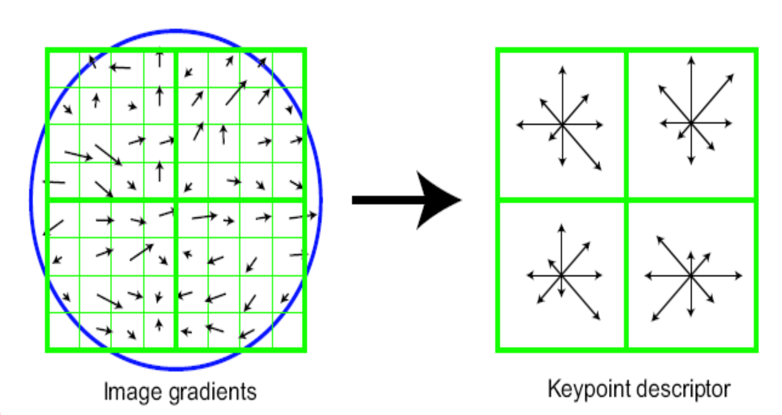
$$\frac{\partial G}{\partial \sigma} = \sigma \nabla^2 G.$$

$$\begin{aligned} D(x,y,\sigma) &= (G(x,y,k\sigma) - G(x,y,\sigma)) * I(x,y) \\ &= L(x,y,k\sigma) - L(x,y,\sigma). \end{aligned}$$

$$D(\mathbf{x}) = D + \frac{\partial D^T}{\partial \mathbf{x}}\mathbf{x} + \frac{1}{2}\mathbf{x}^T \frac{\partial^2 D}{\partial \mathbf{x}^2}\mathbf{x}$$

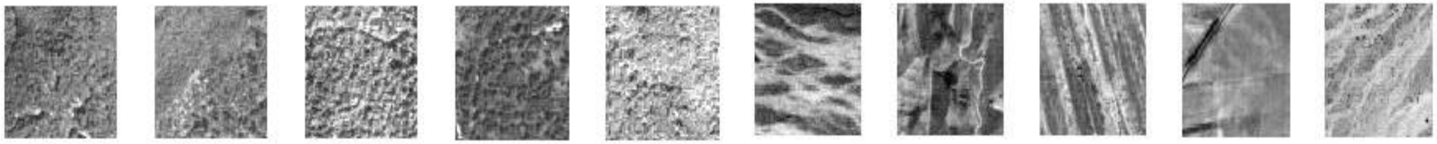
$$\hat{\mathbf{x}} = -\frac{\partial^2 D}{\partial \mathbf{x}^2}^{-1} \frac{\partial D}{\partial \mathbf{x}}.$$

$$\frac{\mathrm{Tr}(\mathbf{H})^2}{\mathrm{Det}(\mathbf{H})} < \frac{(r+1)^2}{r}$$



Category-based retrieval in specific data-bases

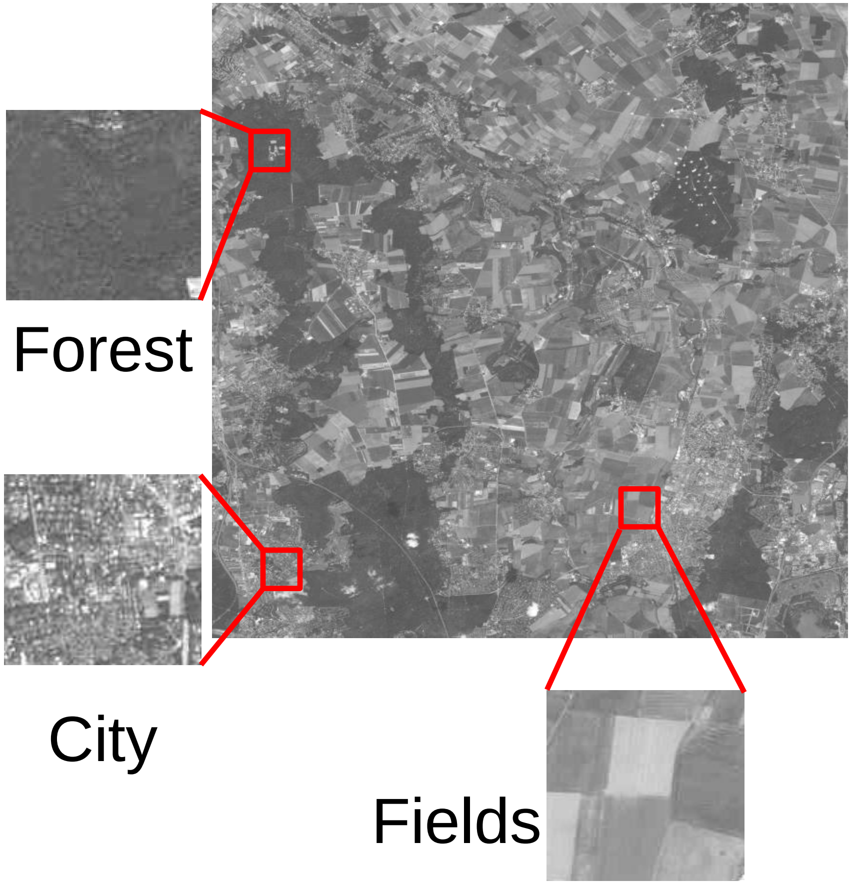
- **Examples:**
  - Biomedical
  - Biology
  - Astronomy
  - Remote sensing and satellite images
- **Goal: to retrieve images « looking the same » as a given sample in very specialized data-bases**



- **Different from : retrieving the exact object in a very broad data-base**

Satellite images

- **A very specific content**



Available information on satellite images

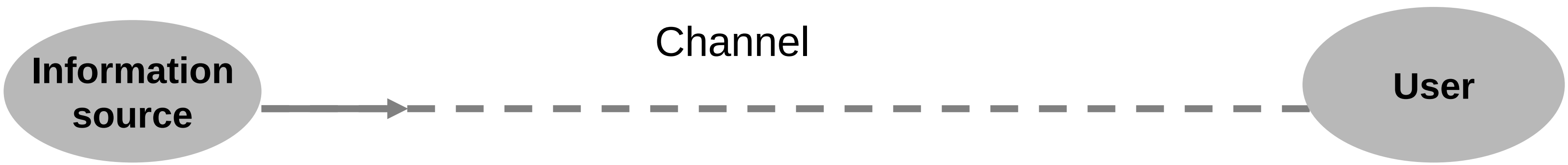
- **Accurate positionning in universal geographical references:** UTM, Mercator, Lambert, etc…
- **Precise time referencing:** seasonal variations (vegetation, insolation, agricultural production, …), sun positionning (shadows), tide effects (precise coast-line, harbours and fishering activities), meteorological conditions (snow, floods, …)
- **Satellite parameters:** resolution, spectral sensitivity, noise
- **Often:** Image quality, cloud cover

Satellite images

- **What are we looking for?** **It is not clear!**
  - Precise objects:
    - A boat — a road-crossing
    - A building — an airplane landing area
  - Generic objects:
    - A marina — a forest fire
    - Greenhouse cultures — refugee camps
    - Oil pipeline — typhoon hazards
    - A geological synclinal
  - Specific terrain configurations:
    - Conducive to: … floods, … desertification, … urban pollution, …
    - Conducive to: … build a factory, … plan a bombing, … cultivate marijuana

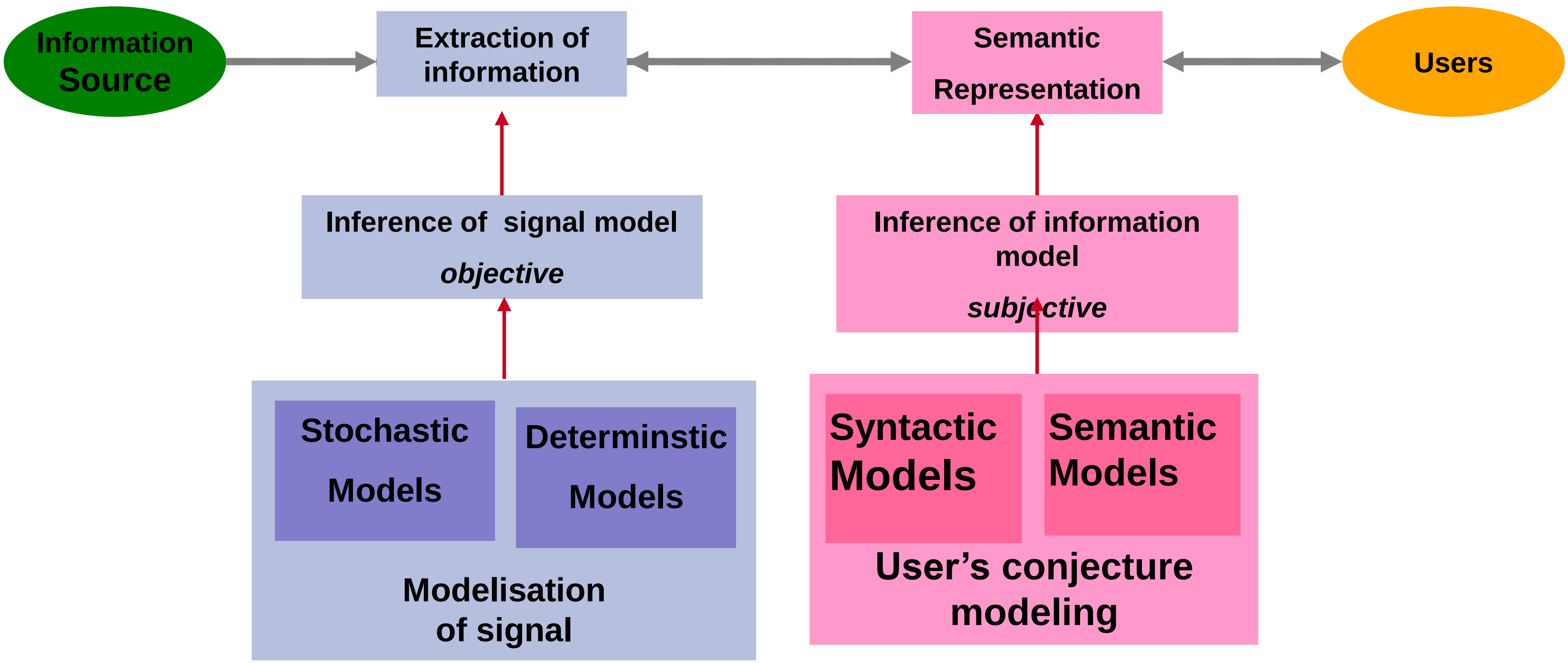
The concept of Channel coding Theory

from CNES/DLR/TelecomParisTech CoC lab



Information Theory (C. Shannon, 1948)

Channel coding concept

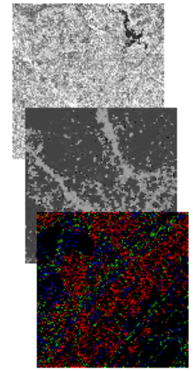


Hierarchical representation

Signal

Pixel

- Spectral
- Texture
- Line/Edge



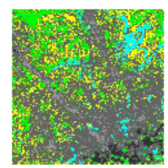
Object Components

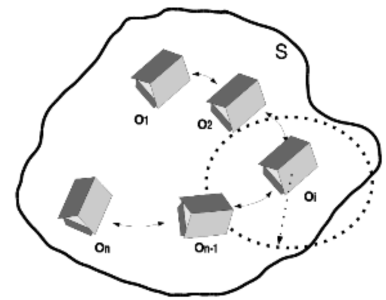
Contextual Structures

Objects and Scene Components

Region

- Classification
- Segmentation

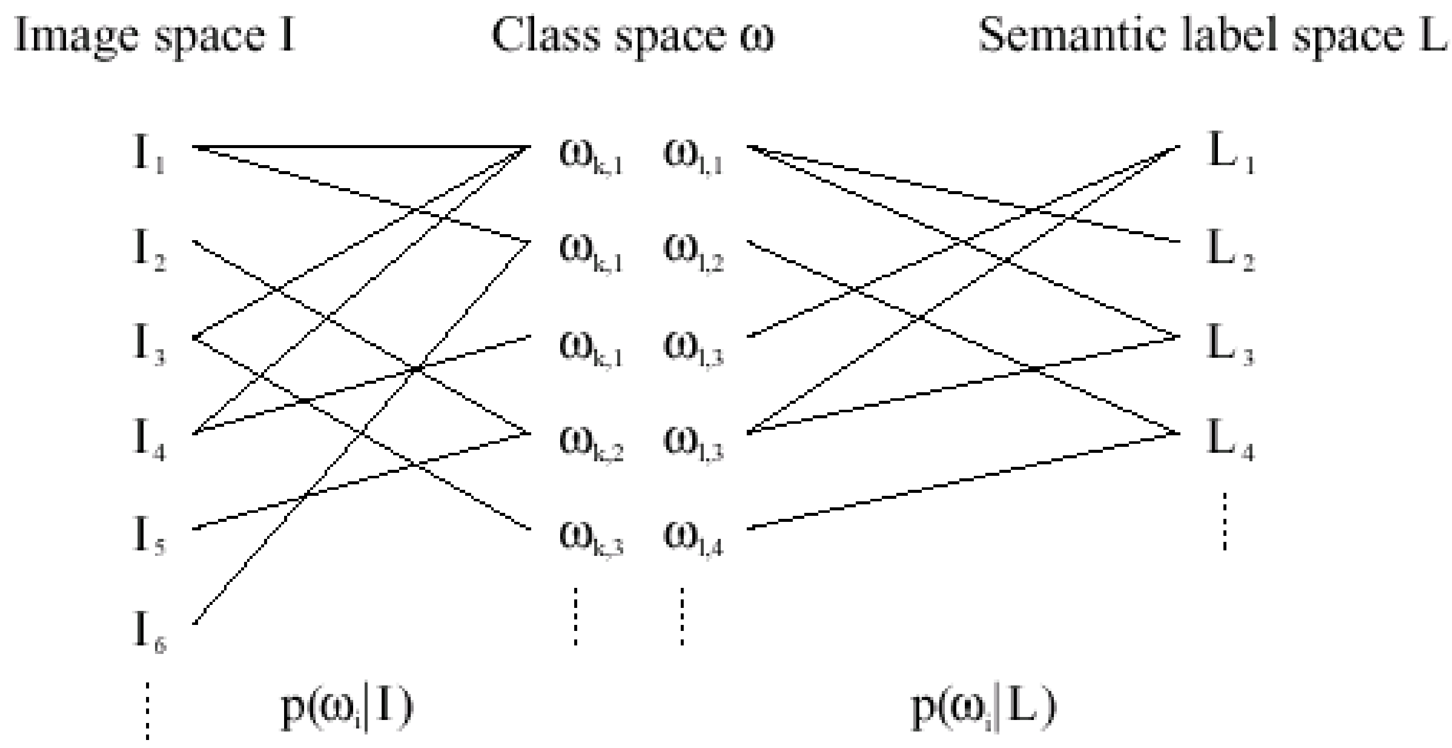




"Semantic" abstraction in signal models

"Semantic" abstraction in user ontology

Statistical coding



Optimization criterion

- **Algoritmic complexity (Rissanen) : Minimal Description Length (MDL)**
- **Rationale:** the best coding is the one that minimizes the representation as a string of characters.

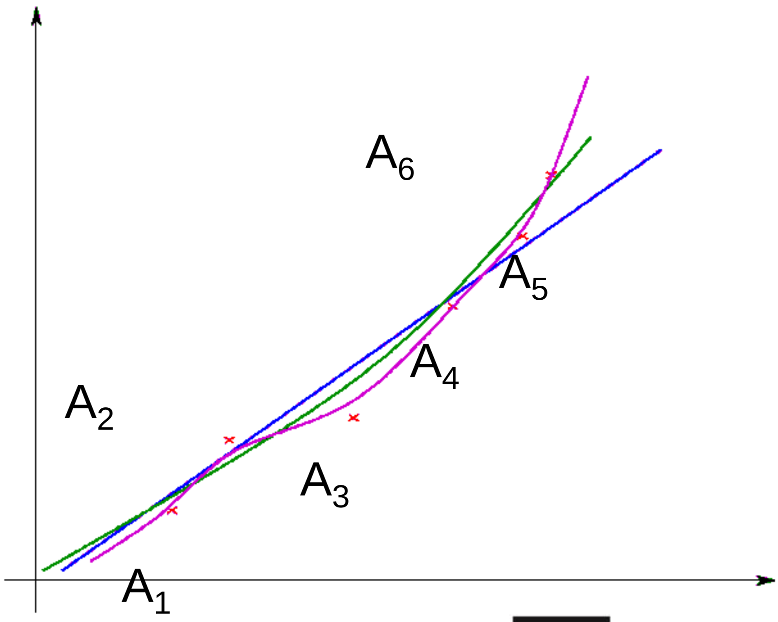
Exhaustive description ➔ 12 values { … x_i,y_i, …}

Choice of a model:

1st order : y=ax+b ➔ 2 parameters {a, b}
6 values {x_1,x_2,x_3,x_4,x_5,x_6}
large errors

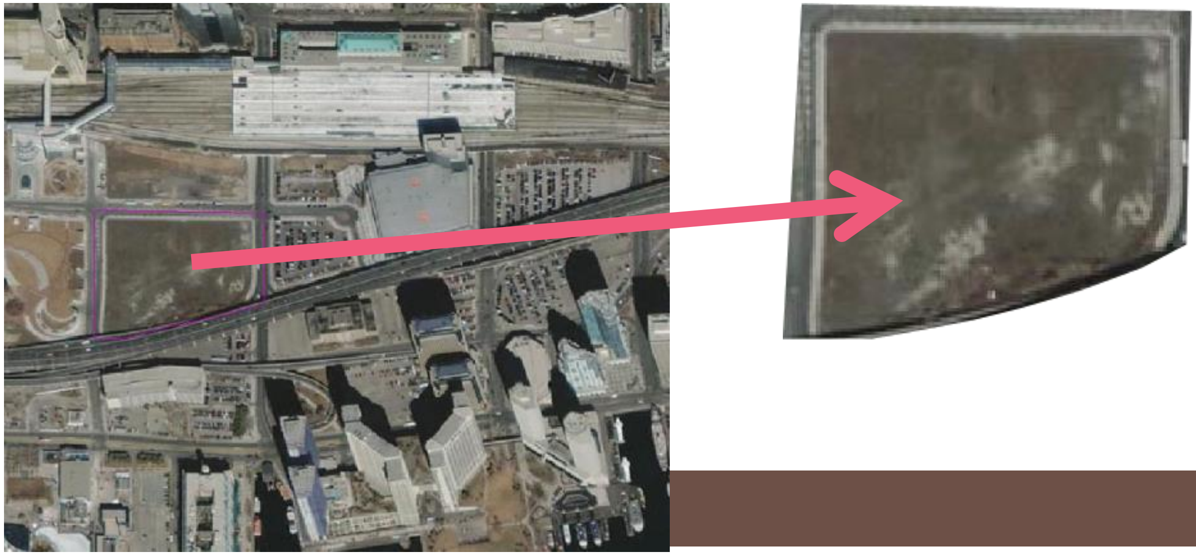
2nd order: y =ax2+bx+c ➔ 3 parameters {a, b,c}
6 values {x_1,x_2,x_3,x_4,x_5,x_6}
smaller errors

exact fit ➔ more parameters
6 values {x_1,x_2,x_3,x_4,x_5,x_6}
no error

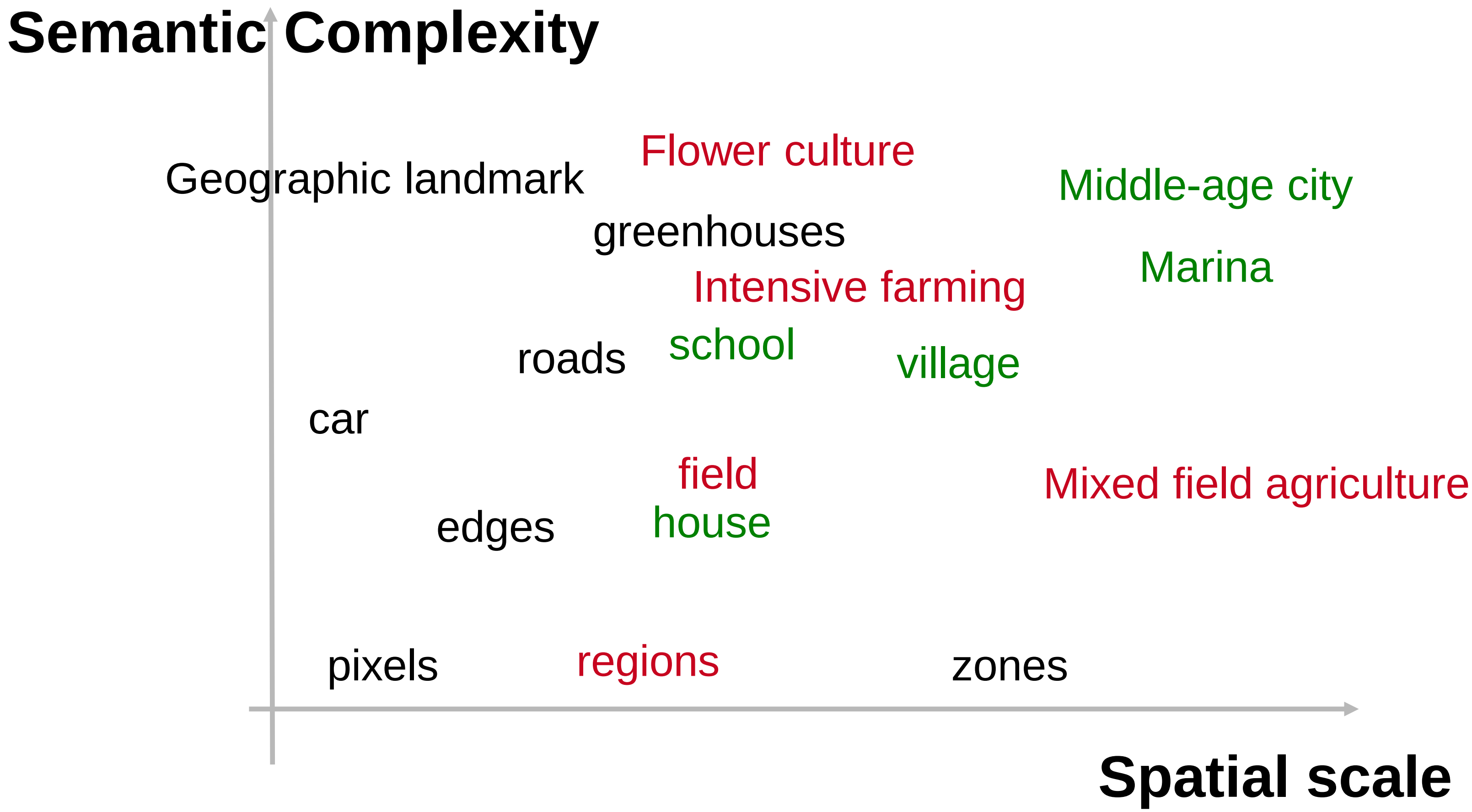


Minimal Description Length (MDL) for images

- Representation model choice (statistical or deterministic)
  - To code the shape or contour of the object
  - To code the grey level
- Representation cost of the image in the model framework
  - Shannon entropy
- Transmission cost of the model
  - For instance: Gaussian: mean + variance = *n* bytes



Spatial scale vs. Semantic complexity



Choice of features =➔ choice of indexing

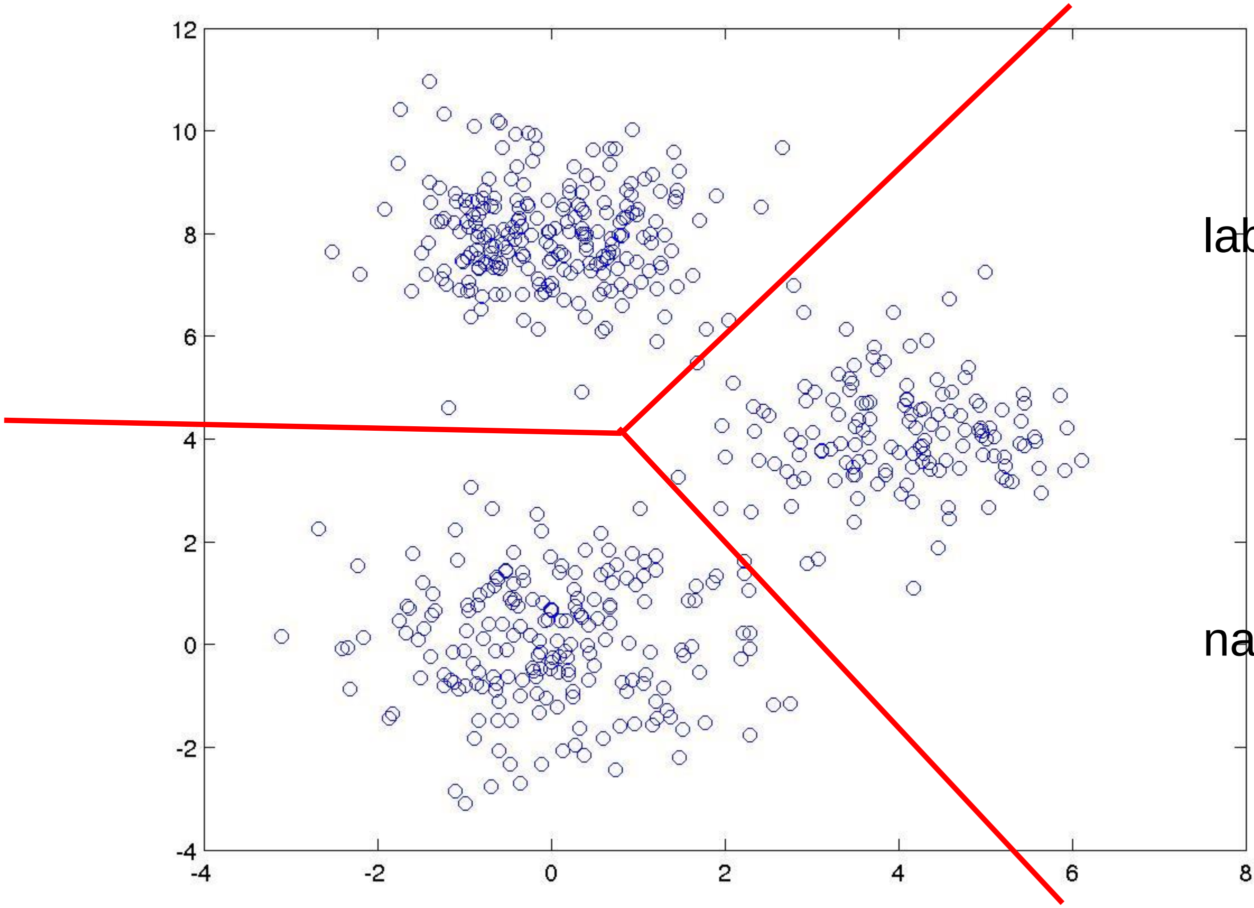
- **Radiometry**
  - Multispectral : channel + NDVI + IB + ISU
- **Textures**
  - Gabor Filters
  - Haralick cooccurrence matrices and their descriptors
  - Quadratic Mirror Filters
  - Contourlet decomposition
  - Steerable wavelets
  - Markov random fields parameters (Gaussian, Laplacian, Loglaplacien …)
- **Structures**
  - Contours, regions
  - Objects : roads, buildings, rivers, lakes
  - Road, Train or River networks

Some efficient choices

- **Indexing:** small subimages: (~ 64 x 64 pixels) = 320 m x 320 m on the ground for SPOT 5 images
- **Mixed features:**
  - Radiometry (Panchro only)
  - Structure (contours)
  - wavelets : 2 directions, 4 scales
- **Automatic feature selection (supervised:** ReliefF, Fisher FS, SVM-RFE or Unsupervised: MIC (*Max Information Compression*), k-means FS**)**

  ~ 100 features with redundancy or → 10 to 20 features without redundancy
- **Give names to classes (***from label to name***)**
  - Waste fields
  - Cultures
  - Housing
  - Road and river networks

Classification



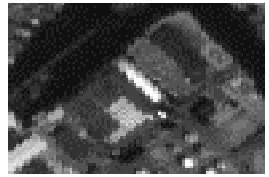
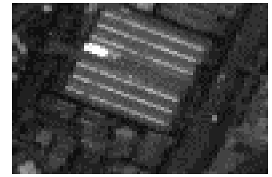
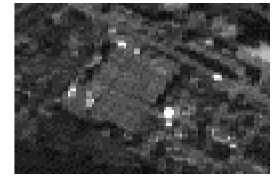
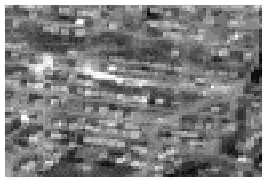
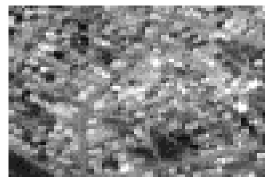
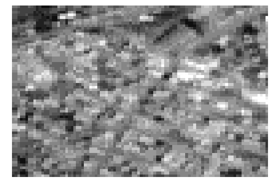
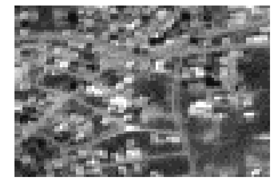
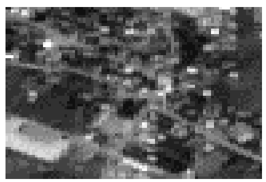
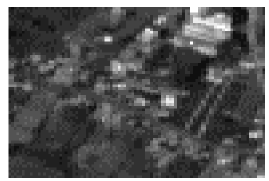
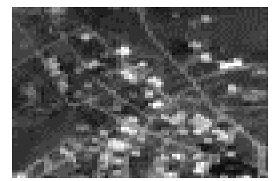
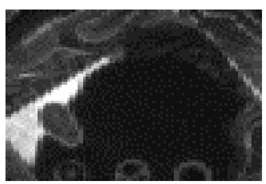
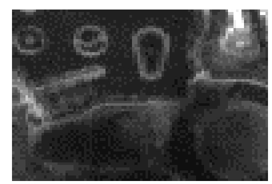
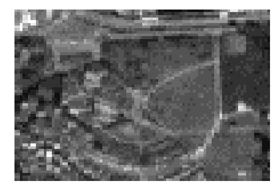
label = 24

or

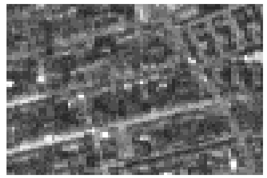
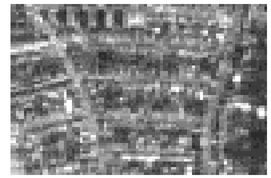
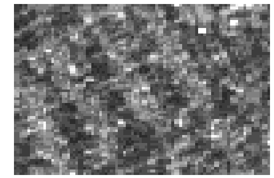
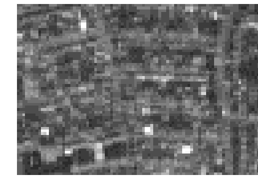
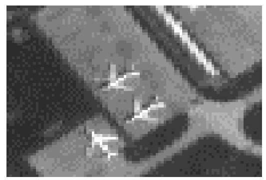
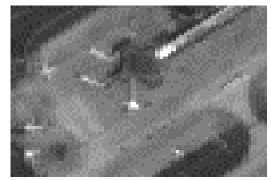
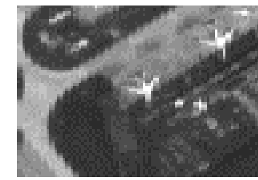
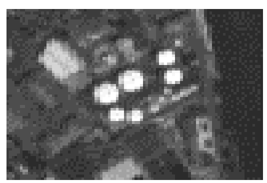
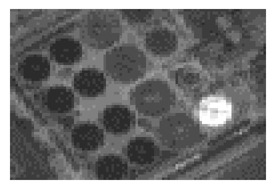
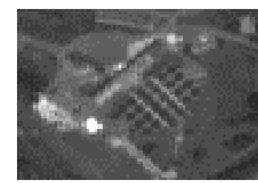
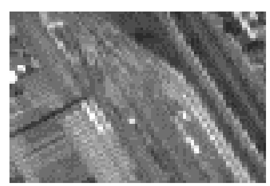
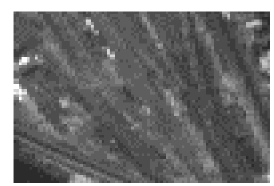
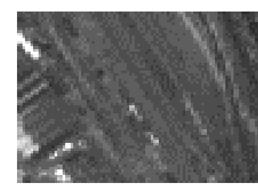
Semantic labelling

name = « Corn field »

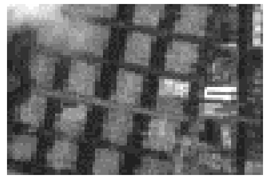
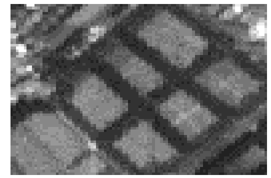
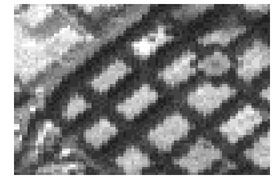
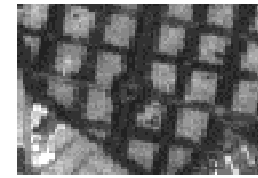
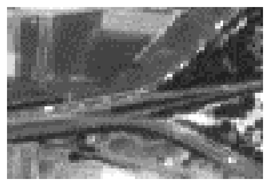
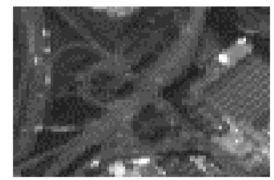
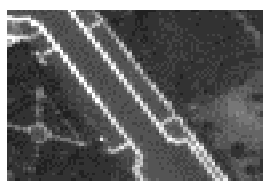
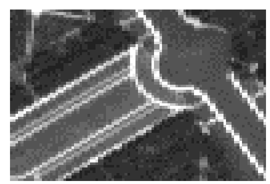
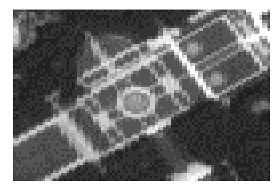
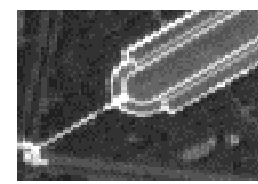
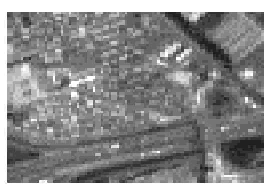
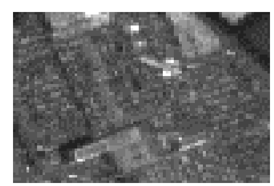
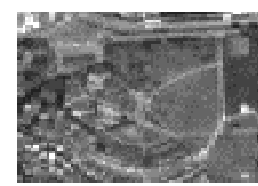
Supervised classes

factories				
Dense urban area				
villages				
Urban parks				

Supervised classes

Residential areas				
Planes				
Industrial tanks & cisterns				
Railway marshalling yard				

Supervised classes

Grave yards				
Road interchange				
Castle parks				
Parking lots				

Automatic classification

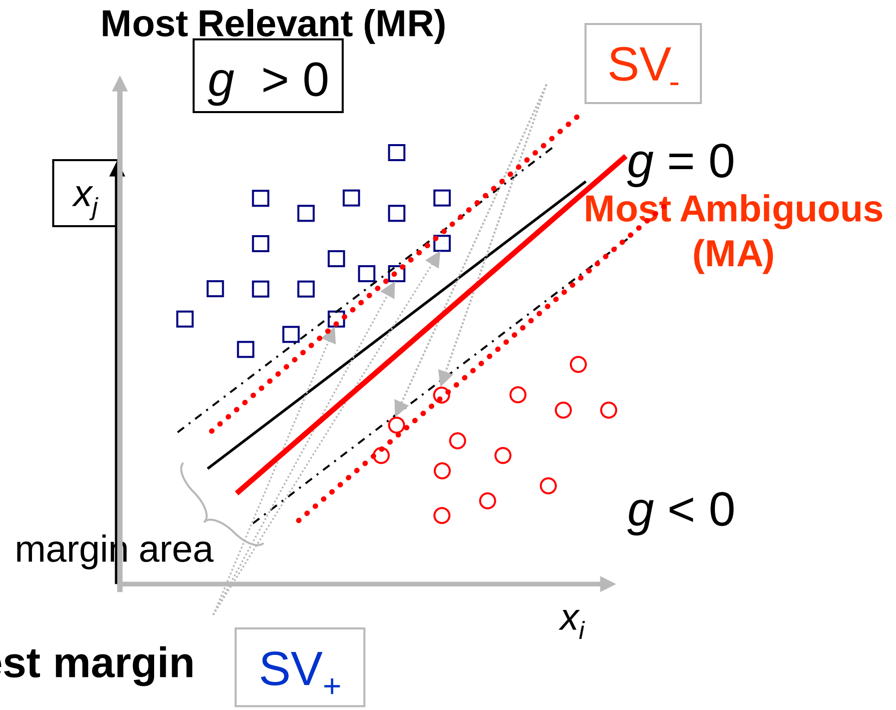
- **Supervised methods**
  - Bayesian decision
  - k- nearest neighbours
  - SVM (Support Vector Machine)
- **or unsupervised**
  - Clustering
  - Hierarchical methods
  - SVM

Support Vector Machine

- **Linear separation case**
  - Labeled data training set

    $(x_i, y_i), x_i \in F = \Re^d, y_i \in \{-1,+1\}, i = 1..N$
  - Find a separation surface

    $g(x) = w \cdot x + b = 0 \qquad y_i(w \cdot x_i + b) \geq 1$
  - Decision function *f = sign(g(x))*
  - d_+ = distance from *g* to closest {+1}
  - d_- = distance from *g* to closest {-1}
  - Margin area = $d_+ + d_- = \frac{2}{\|w\|}$

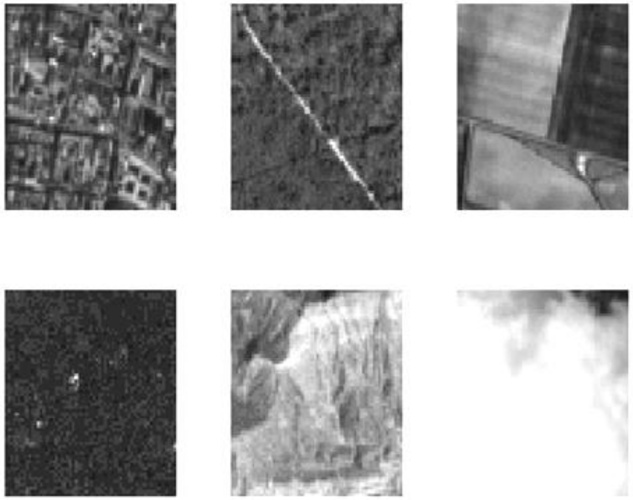


- **Find a separating hyperplane with largest margin**

$$L = \frac{1}{2}\|w\| - \sum_{i=1}^{N} \alpha_i (y_i (w \cdot x_i) - 1) \rightarrow \frac{\partial L}{\partial w} = 0 \; and \; \frac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^{N} \alpha_i y_i = 0 \; and \; w = \sum_{i=1}^{N} \alpha_i y_i x_i$$

$$\max_\alpha \left(\sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{N} \alpha_i \alpha_j y_i y_j x_i x_j\right) \rightarrow \sum_{i=1}^{N} \alpha_i y_i = 0 \,, \alpha_i \geq 0$$

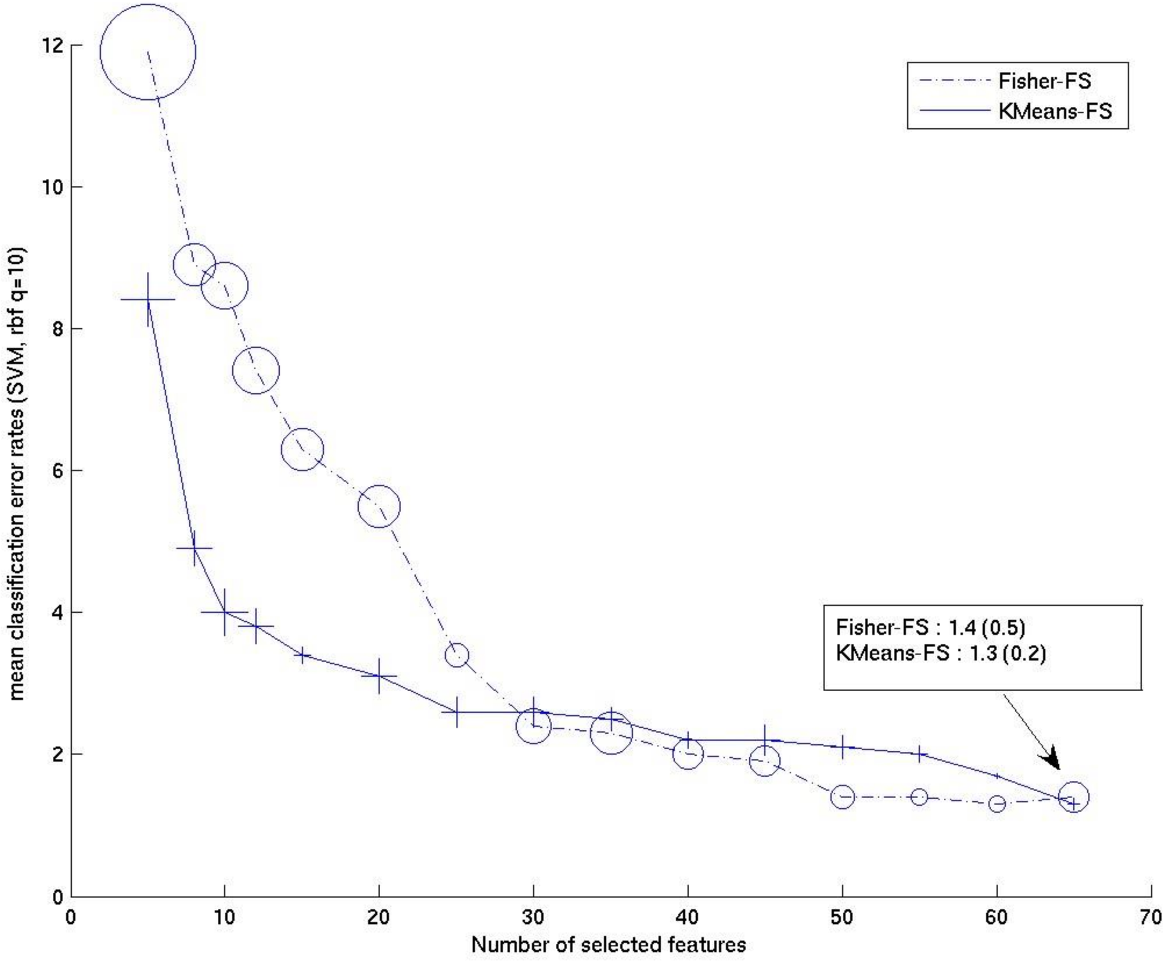
Performances of algorithms



Sub image classification (128 x 128) :
city, wood, fields, sea, desert & clouds
600 images for each class
Results: Gaussian SVM,
Mean error 1.4% ± 0.4%
(147 features, cross validated)

True\Found (%)	city	clouds	desert	fields	woods	sea
city	98.8	0	0	0.5	0	0
cloud	0	99.3	0.2	0	0	0
desert	0	0	99.0	0.3	0	0
fields	0.5	0.2	0.8	98.1	0.3	0.4
woods	0	0.2	0	0	98.0	1.4
sea	0.7	0.3	0	1.0	1.7	98.2

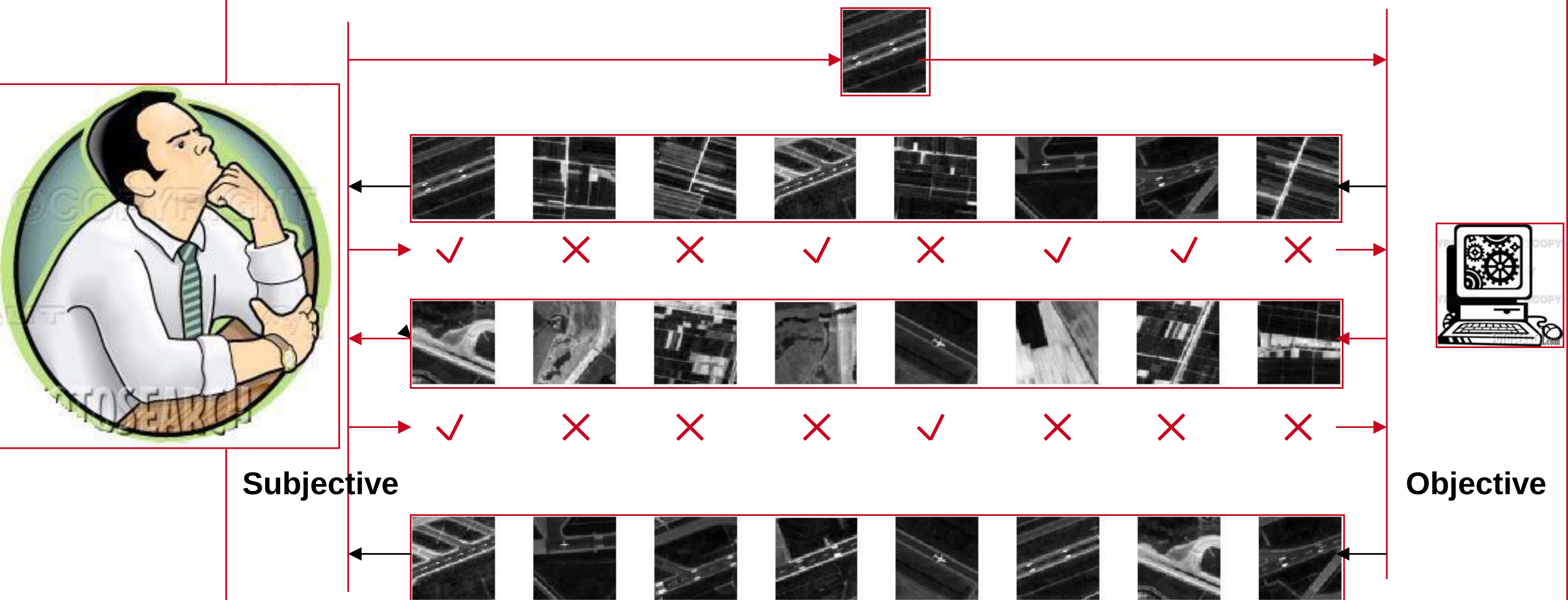
How many features?



Using a human expert

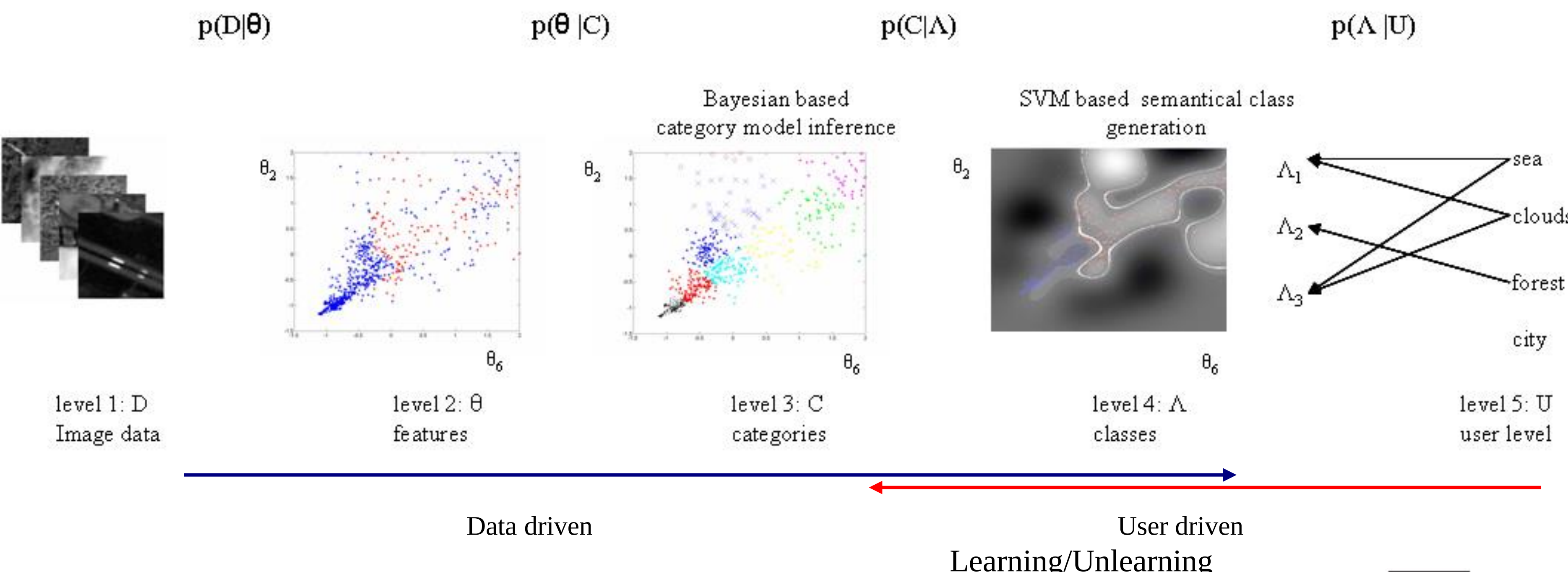
Learning with Relevance feedback

- **Man Machine dialog**

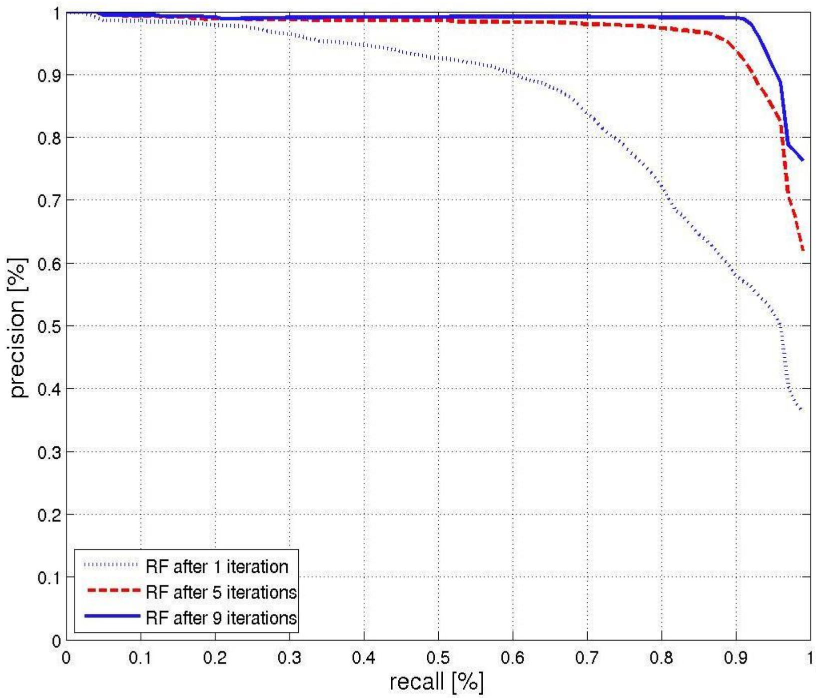


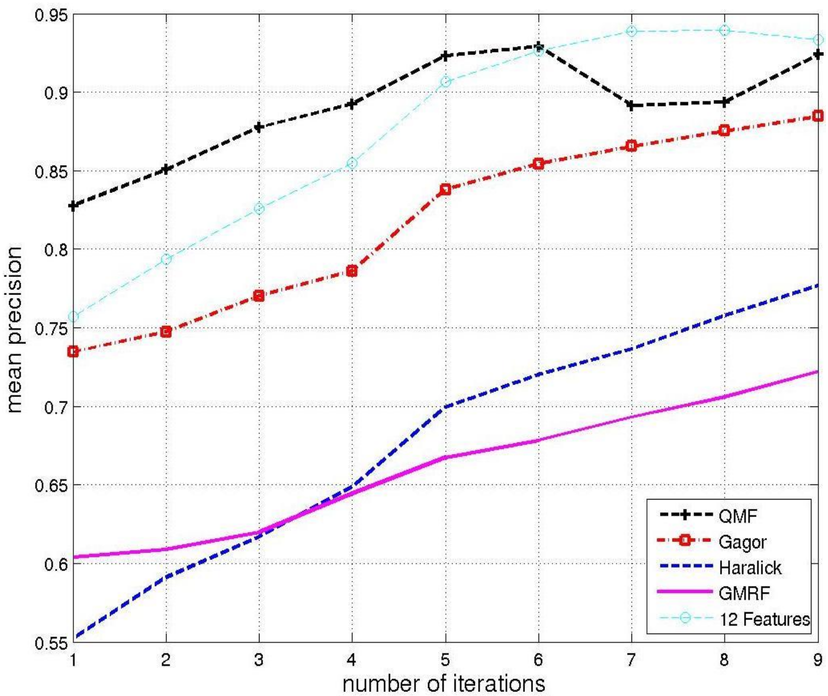
Classification hierarchy:

Pass 1 to 2 : K-means classification
Pass 2 to 3 : Bayesian decomposition of clusters with Gaussian mixtures
Pass 3 to 4 : displacement of SVM fronteer points
Pass 4 to 5 : Gaussian identification gaussiennes ➔ words



- Database composed of 600 SPOT5 images divided in 6 classes
- Used features: Gabor, Haralick, QMF and GMRF
- Gaussian Kernel
- System evaluation: Precision-Recall graphs





Typical evaluation parameters

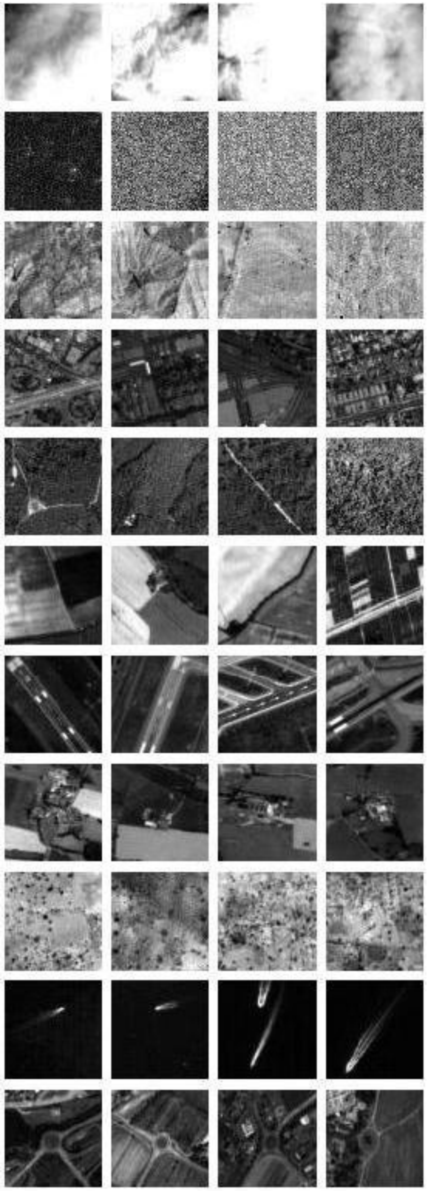
SPOT5 images ©CNES 3000x3000 pixels

- **Résolution 5m/pixel**
- **Sub-images 64x64 pixels**
- **46 images used to create a dat base with 11 classes**
- **Features : Wavelets coefficients QMF (11)**
- **100 instances per class**
- **SVM with Gaussian kernels:**

$$K(\theta_i,\theta_j) = \exp(-2\|\theta_i - \theta_j\|^2)$$

- **8 images in the relevance feedback loop**
- **15 RF loops**
- **Evaluation made on the 60 most likely images**

1. clouds
2. sea
3. desert
4. city
5. Forest
6. Fields
7. airport
8. Village
9. Savana
10. Boat
11. Road crossing



From Low to High Level
Changing the scale

Complexity of images

Analysis window : real size
128 x 128 pixels

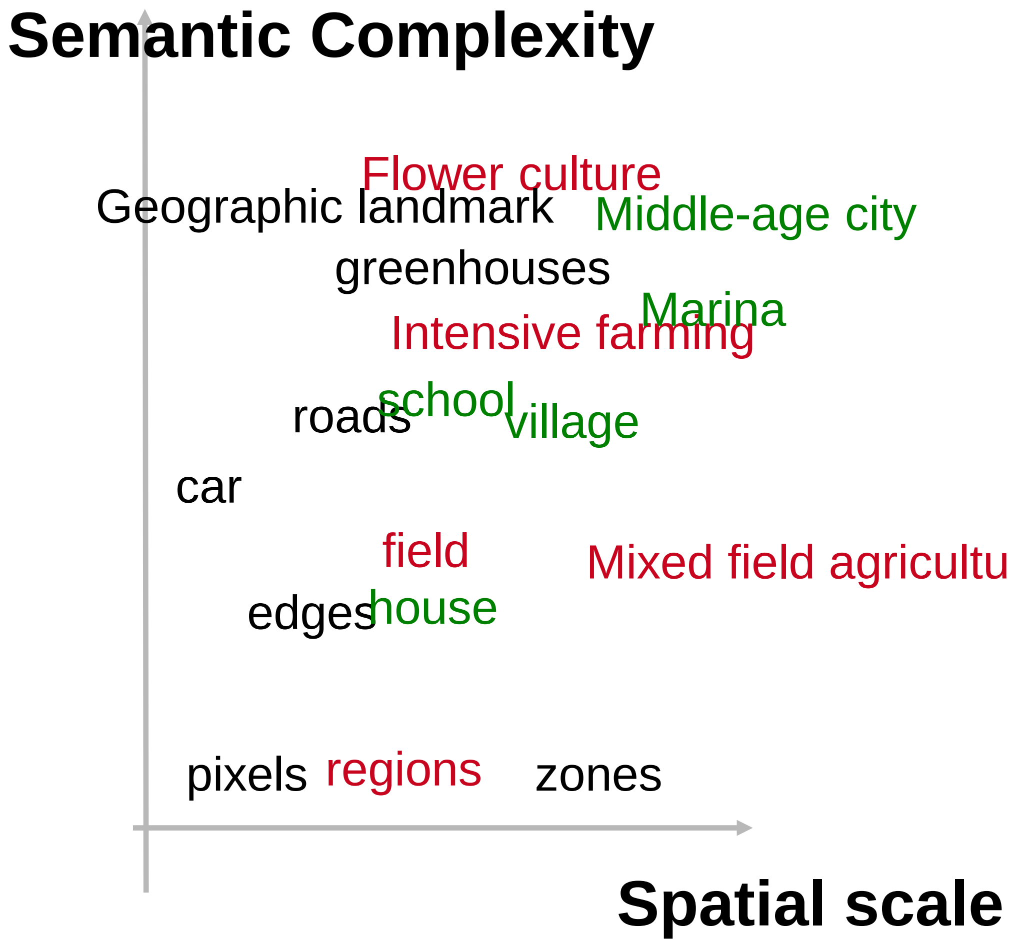
Analysis window : enlarged

Hierarchical representation

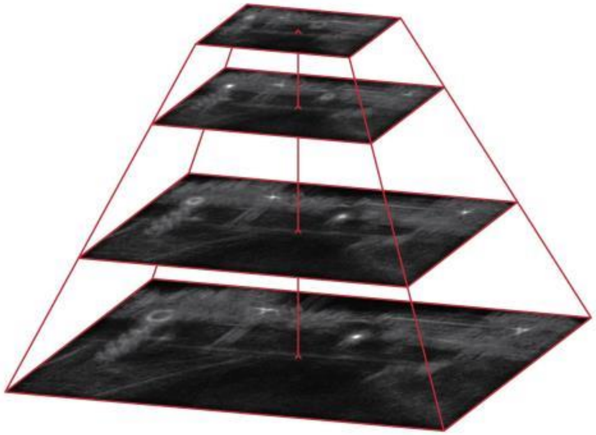
Two goals:
- Enlarge the field of view
- Increase the semantic level

Grouping strategy:
- Sliding window
- Pyramid
- Growing and Merging

Decision strategy:
- Bag of Visual Words (BOVW)



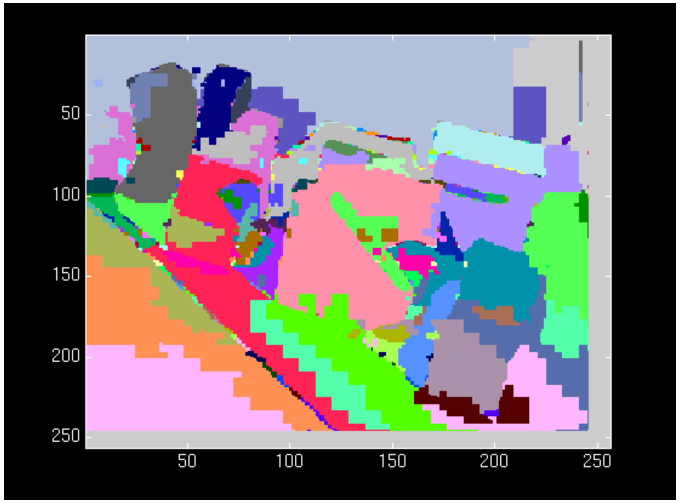
Scale enlargement strategy



Pyramid

Sliding window

Growing and Merging



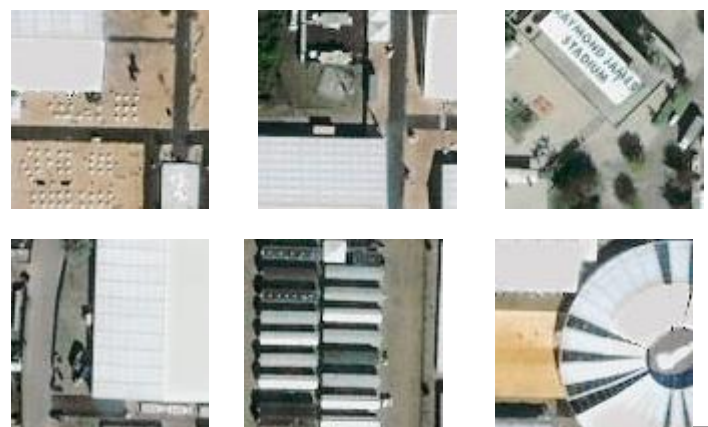
Increasing the semantics

Park = {trees+fields+tracks}

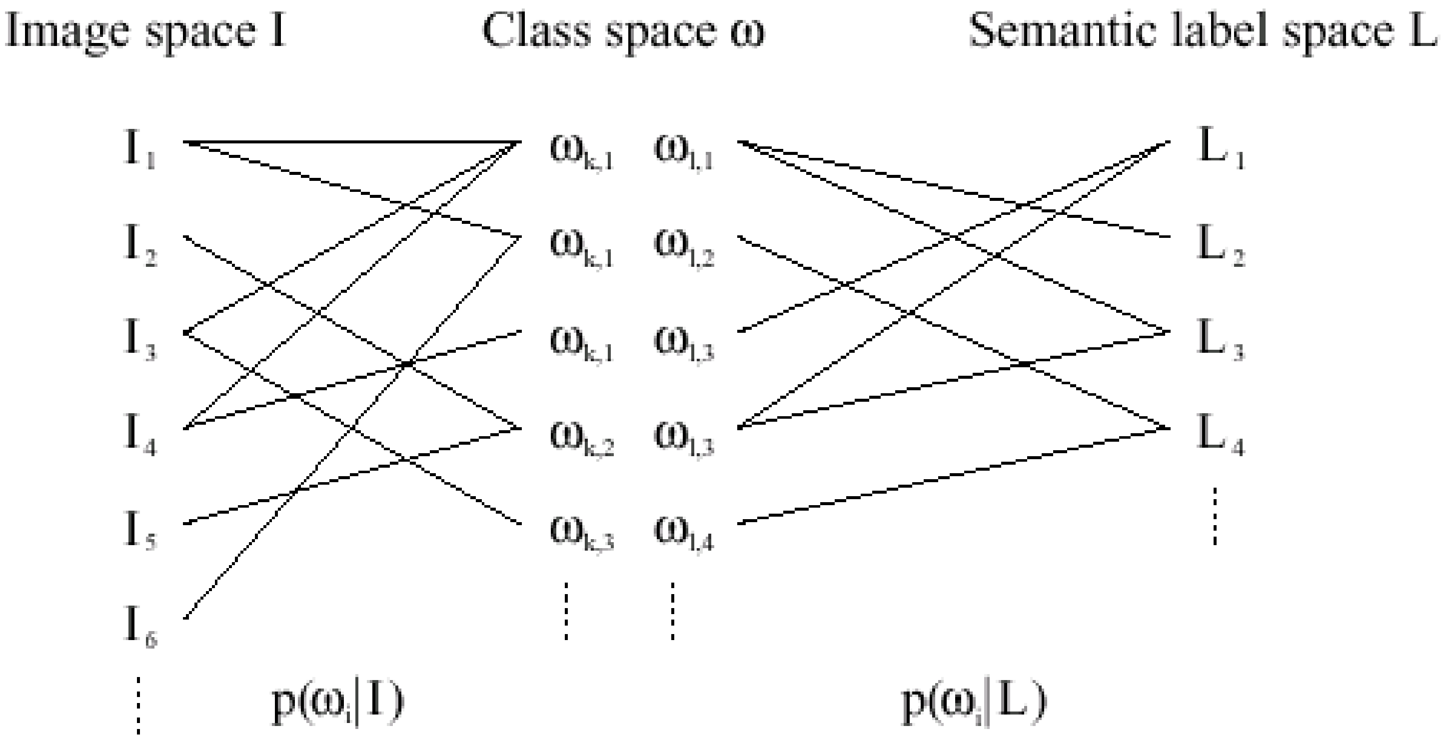
Residential area = {houses + lawns + pools + roads}

Waste area ={waste+lawns+trees+roads}

Commercial area = {buildings+houses+parking lots+ waste



Statistical coding



Decision making: Bag of Words

- **2 levels: L=Low, H=High**
- **Ordered list of *N* classes at L = $\{c_1, c_2, \ldots c_N\}$**
- **At H: 1 super-region with *n* objects, each with 1 class = *n* labels**
  - The super-region is described by the ordered list of the probability (or the occurrence) of each class:

$$R_k=\{p_1, p_2, \ldots p_n\}$$

- **At H, super-regions are classified into super-classes C_j by partitionning *n*-Dimensional space according to R_k**

Machine Learning

Unsupervised learning

Many different algorithm using different criteria

- **K-means**
- **Kernel K-means algorithm (J.Shawe-Taylor, 2004)**
- **Spectral K-means algorithm (A.Ng, 2002)**
- **Hierarchical algorithm (Ward, G.Karypis, 1998)**
- **EM algorithm for MAP estimation : Autoclass (P. Cheeseman, 1996)**

Which one to choose?

How to use them?

Détermination of the optimal number of clusters : MDL

Minimum Description Length (Rissanen, 1978)

$$\min_{\kappa,\Theta} \; \Lambda = -\log(P(\mathrm{X}\,|\,\Theta)) + \frac{1}{2}\kappa \log(M) \qquad (1)$$

P(X|Θ) of Gaussians and a "hard clustering" (1) is:

$$\Lambda = -\sum_i n_i \log\left(\frac{n_i}{|\Sigma_i|}\right) + \frac{1}{2}\kappa \log(M) + const \qquad (2)$$

n_i **= number of samples in i[th] cluster,**

$|\Sigma_i|$ **= determinant of covariance matrix**

$|\Sigma_i|$ **calculated both for spectral and kernel clustering algorithms**

Coassociation Matrix

Partition membership is a binary matrix:

$$D_{ij}^{n} = \begin{cases} 1, \text{if i and j are in the same cluster for n-th partition;} \\ 0, \text{otherwise.} \end{cases}$$

Co-association is a probability that two elements i and j are in the same cluster:

$$C = \frac{1}{N}\sum_{n=1}^{N} D^{n},\ N \text{ number of clustering}$$

Coassociation Matrix

Coded partitions

$$X^1 = \begin{vmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{vmatrix}$$

$$X^2 = \begin{vmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{vmatrix}$$

Partition membership:

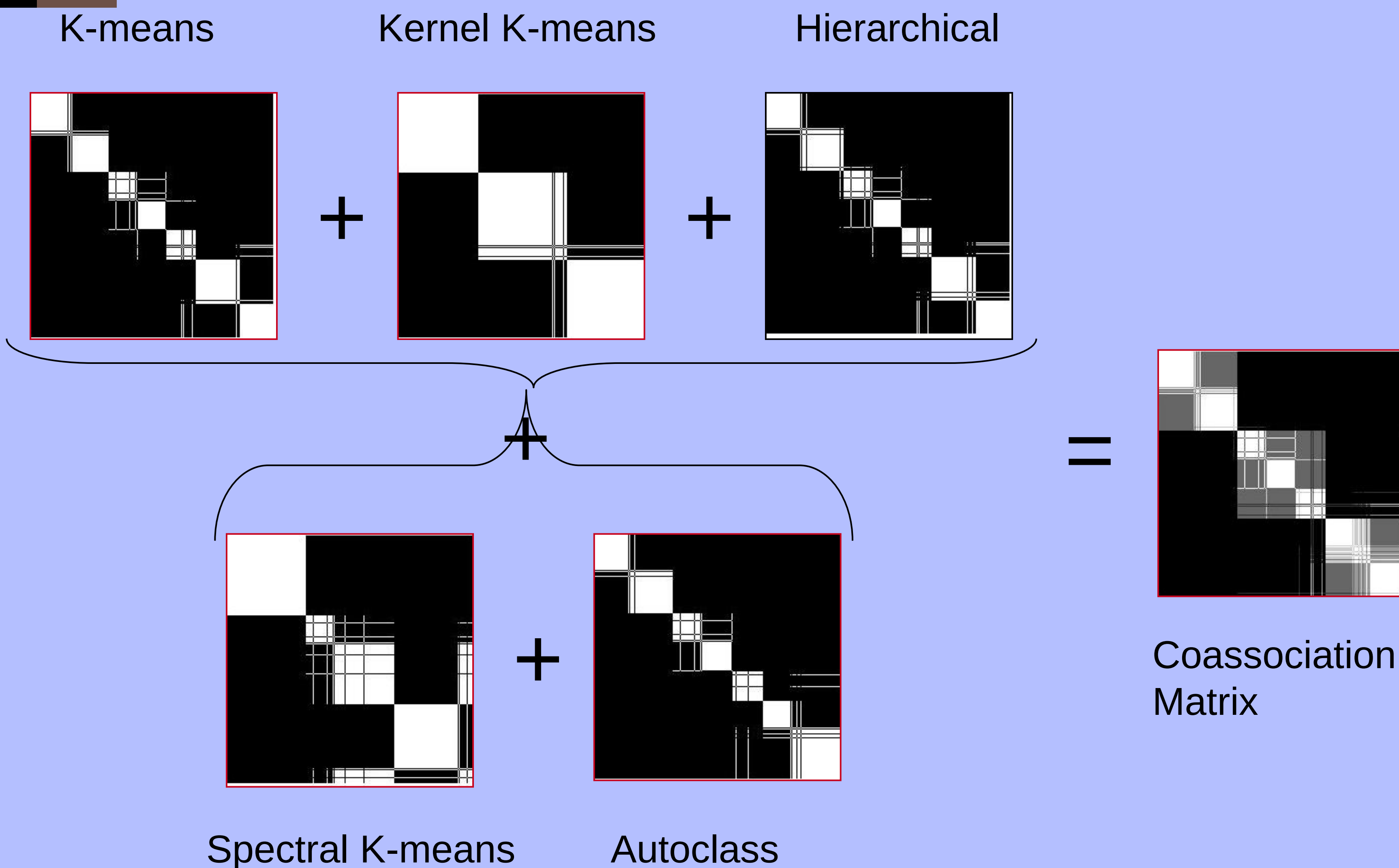
$$D^1 = X^1 (X^1)^T = \begin{vmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{vmatrix}$$

$$D^2 = X^2 (X^2)^T = \begin{vmatrix} 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{vmatrix}$$

Co-association matrix

$$C = \frac{1}{2} \sum_{n=1}^{2} D^n = \begin{vmatrix} 1 & 1 & 0.5 & 0.5 & 0 & 0 \\ 1 & 1 & 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0.5 & 1 & 1 & 0 & 0 \\ 0.5 & 0.5 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{vmatrix}$$

Exemple


K-means
Kernel K-means
Hierarchical
+
+
+
=
Spectral K-means
+
Autoclass
Coassociation
Matrix

Un supervised combination of clustering

$$D_{ij}^{n} = \begin{cases} 1, \text{ if } \ i \text{ and } \ j \text{ are in the same cluster for } n^{\text{th}} \text{ clustering;} \\ 0, \text{otherwise.} \end{cases}$$

Co-association matrix $C = \frac{1}{N}\sum_{n=1}^{N} D^{n}$, N number of clustering **Fred & Jain (PAMI 2005)**

Final combination = X **, such that:**

$$X = \arg\min \left\| C - XX^{T} \right\|^{2} \qquad \textit{subject to } \ X^{T}X = I, X \in \{0,1\} \qquad (3)$$

$$X_{iq} = \begin{cases} 1, \text{ if } \ i \in q, \forall\, i, q = 1,\ldots,Q; \\ 0, \text{otherwise.} \end{cases}$$

Combination algorithm: single-link based clustering (1974, Hubert) to minimize (3)

Images: SPOT5, 5 m/pixel

Barcelona

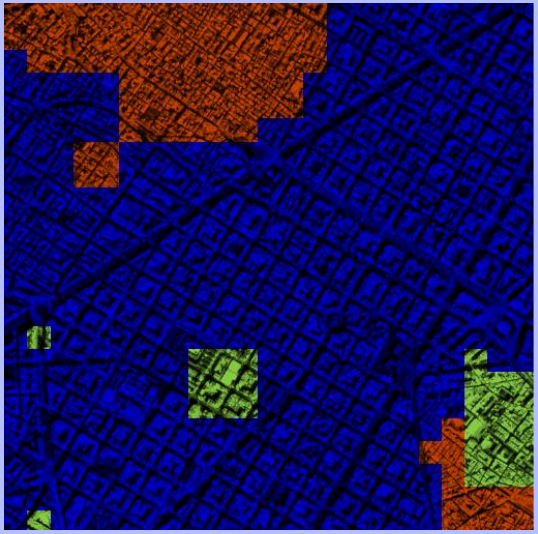
Istanbul

Los Angeles

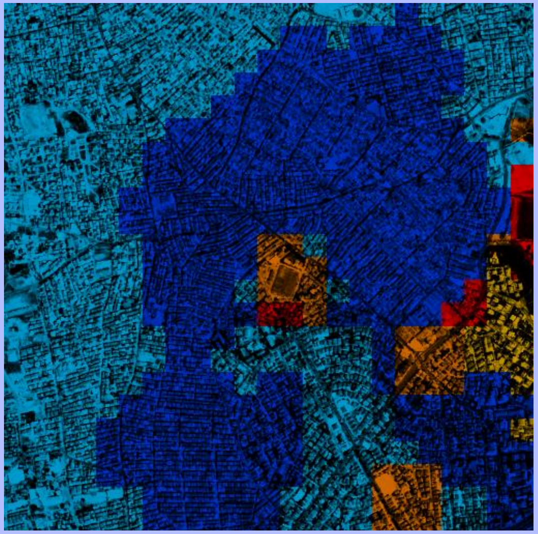
Madrid

Results. Clustering by Autoclass

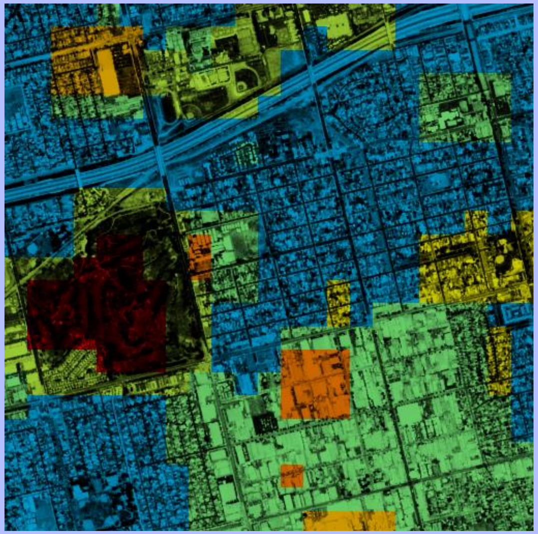
Optimal number of clusters: 14



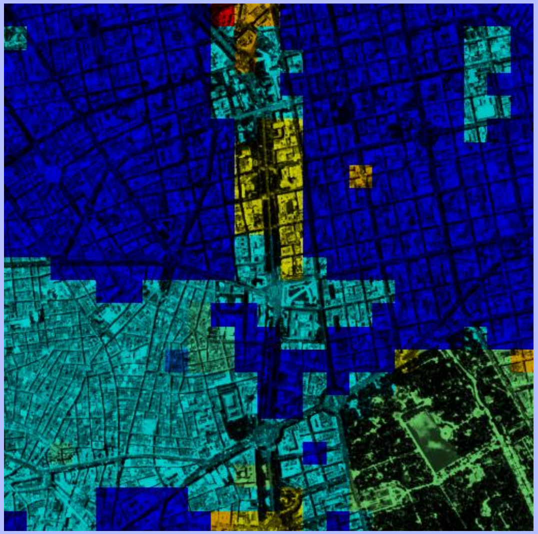
Barcelona



Istanbul



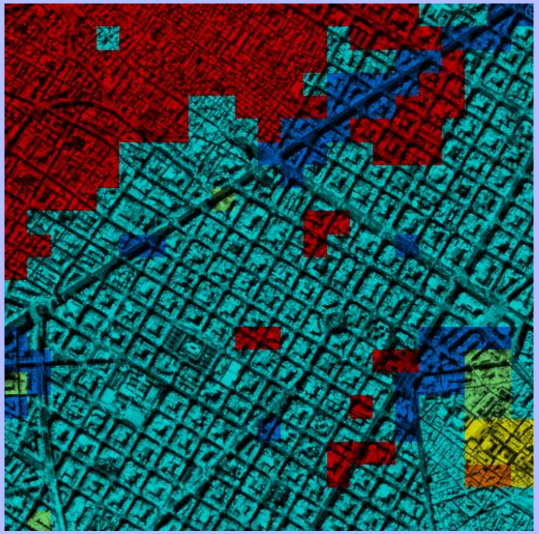
Los Angeles



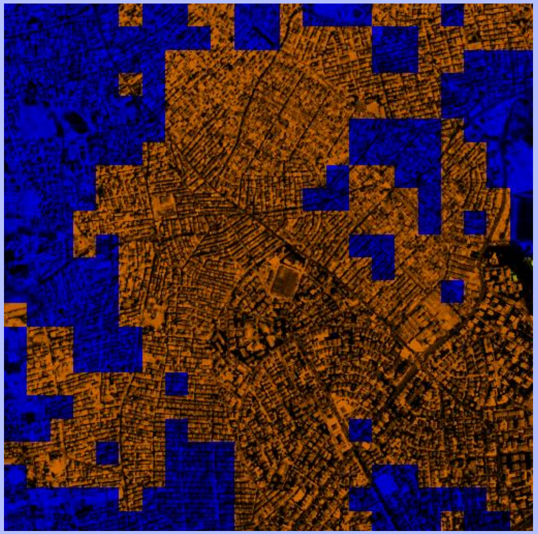
Madrid

Clustering by Kernel K-means

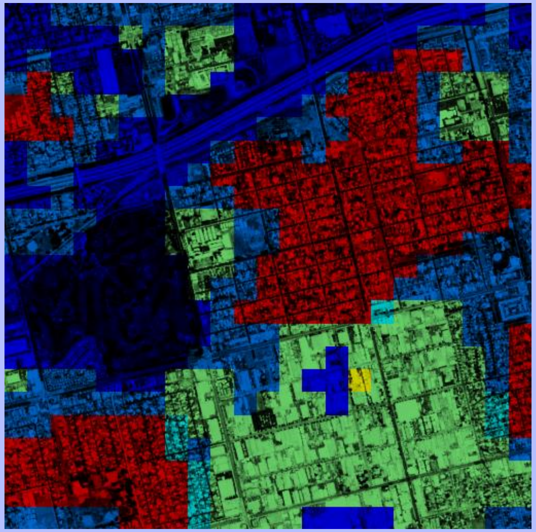
Optimal number of clusters: 7



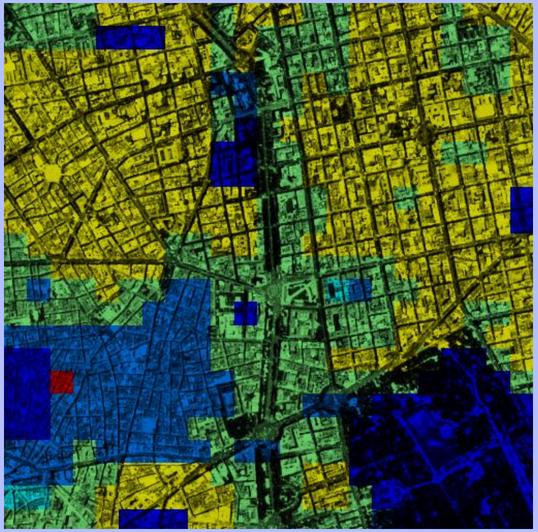
Barcelona



Istanbul



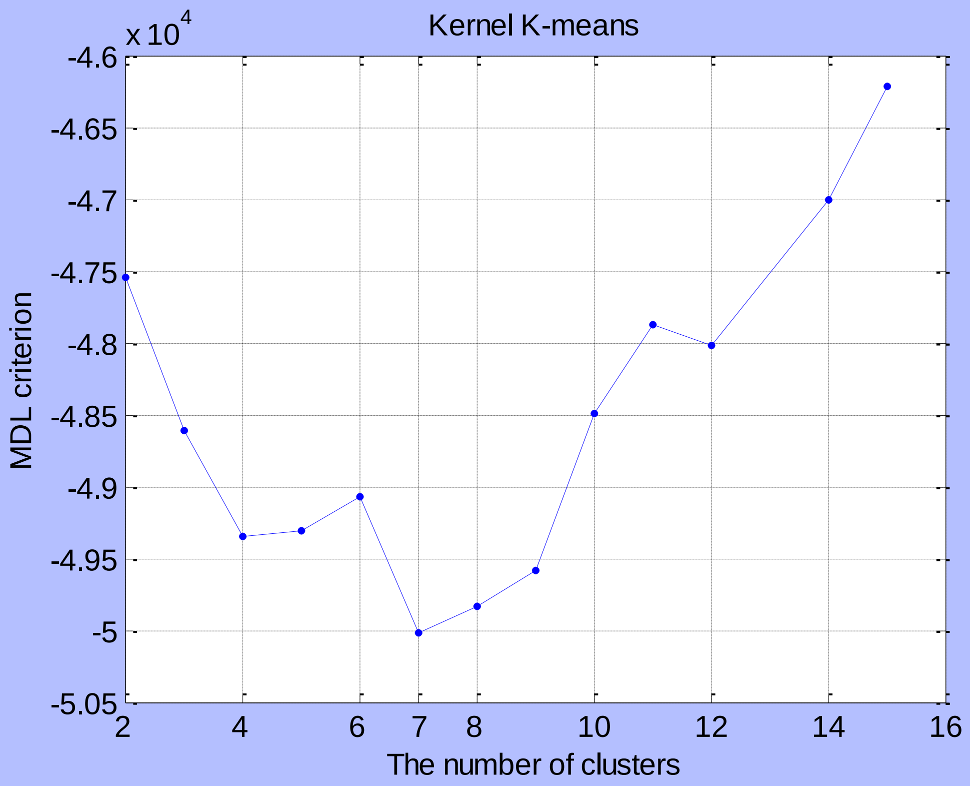
Los Angeles



Madrid

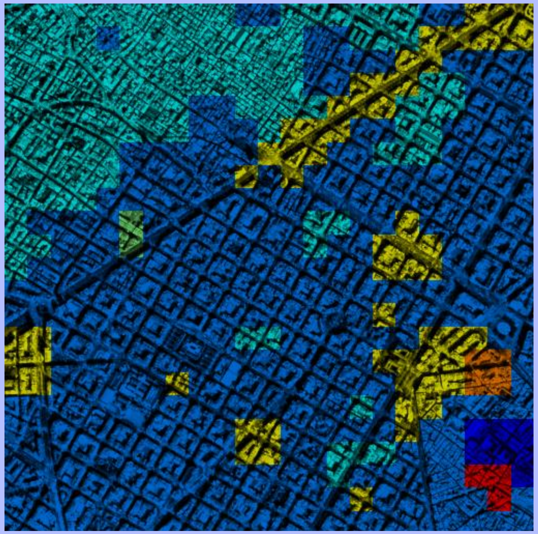
Clustering by Kernel K-means

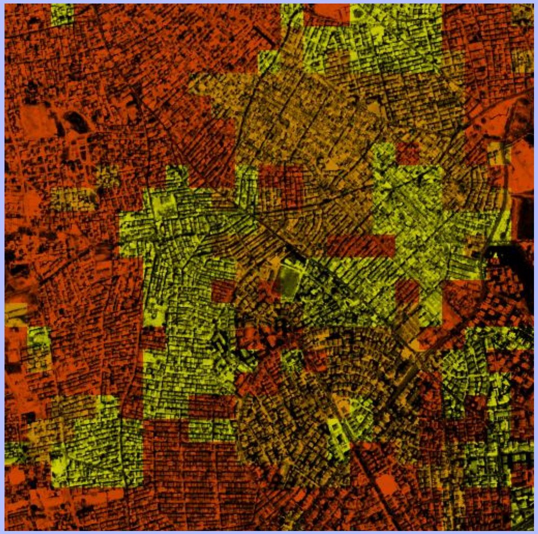
Optimal number of clusters: 7

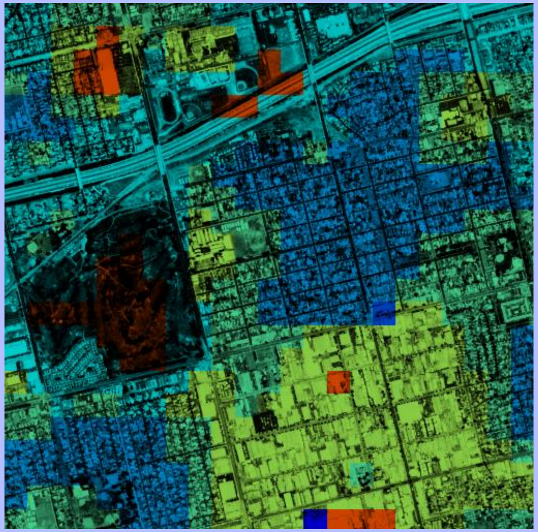


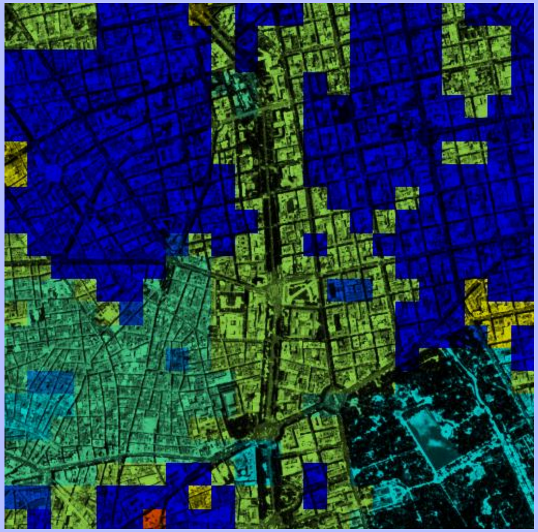
Clustering by Spectral K-means

Optimal number of clusters: 9


Barcelona

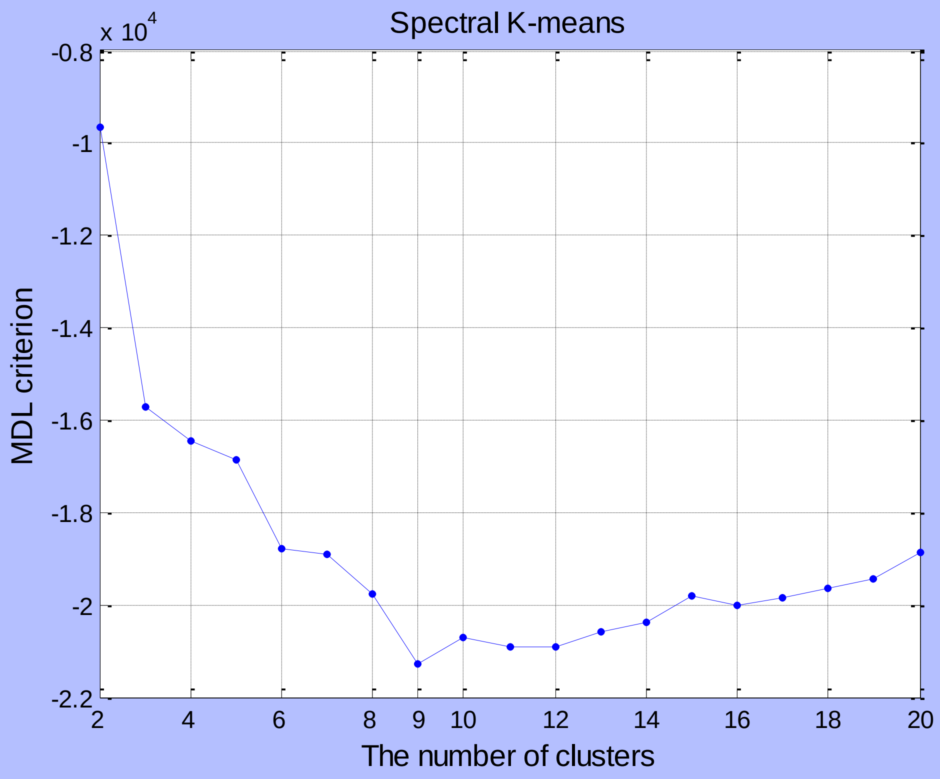

Istanbul


Los Angeles


Madrid

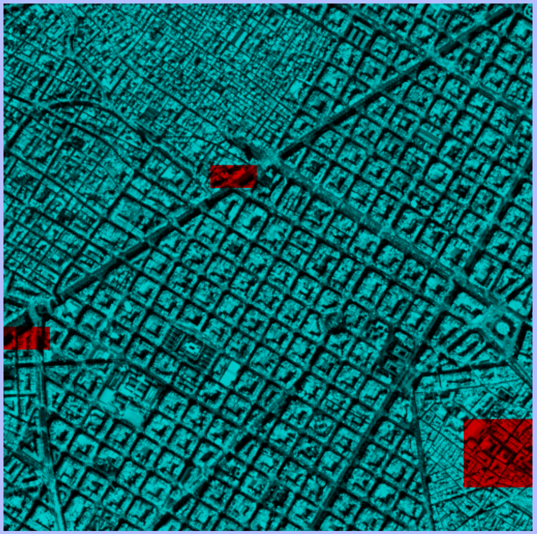
Clustering by Spectral K-means

Optimal number of clusters: 9

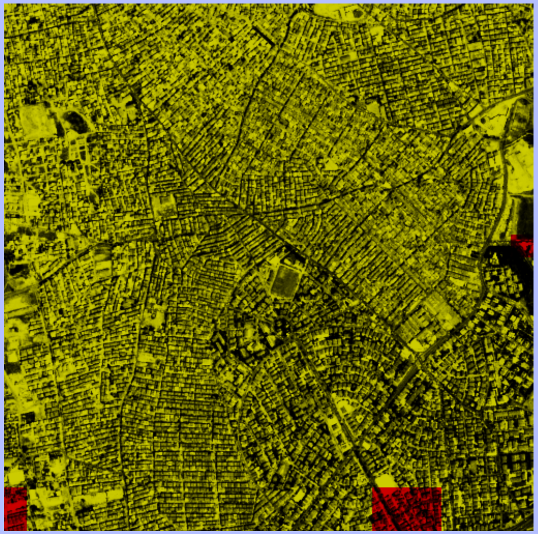


Clustering by Ward algorithm

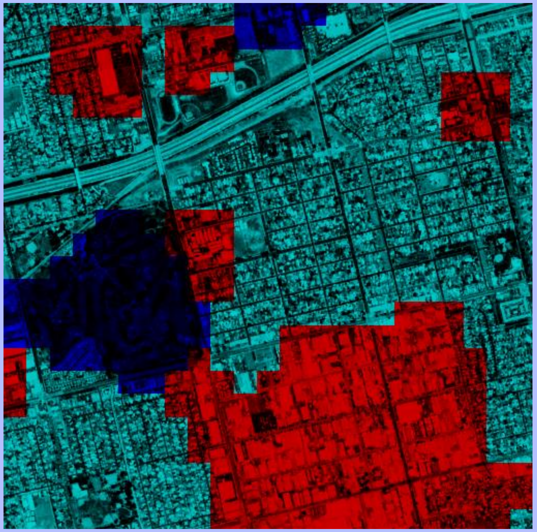
Optimal number of clusters: 4



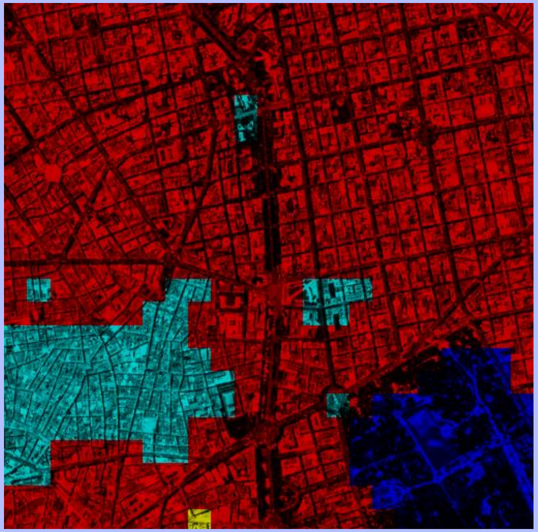
Barcelona



Istanbul



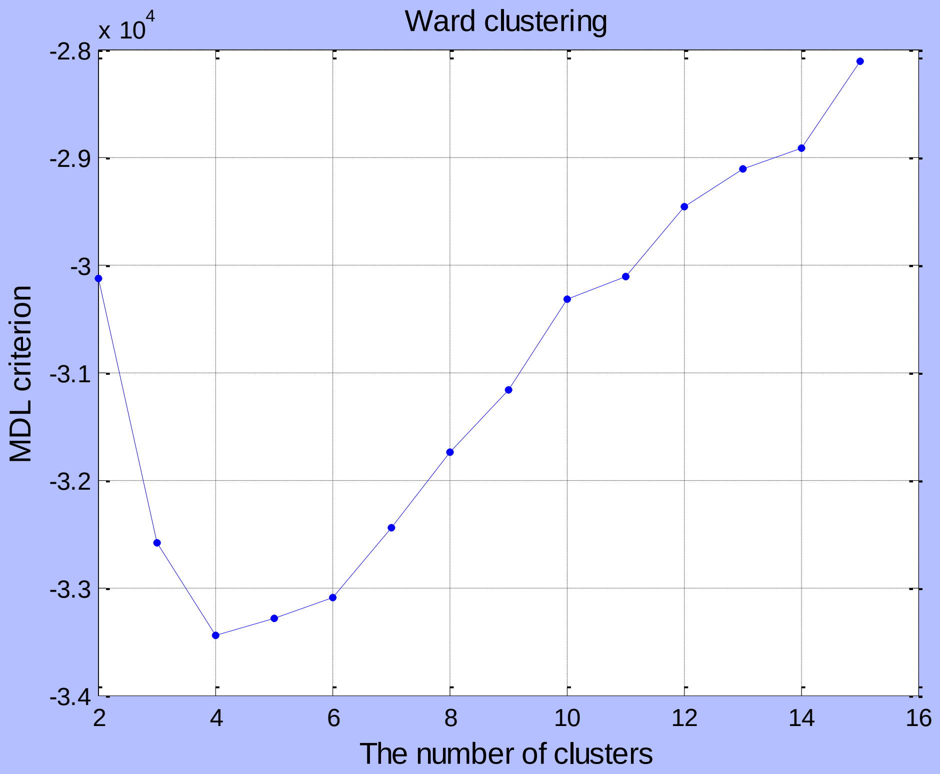
Los Angeles



Madrid

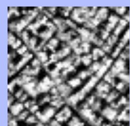
Clustering by Ward algorithm

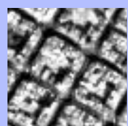
Optimal number of clusters: 4

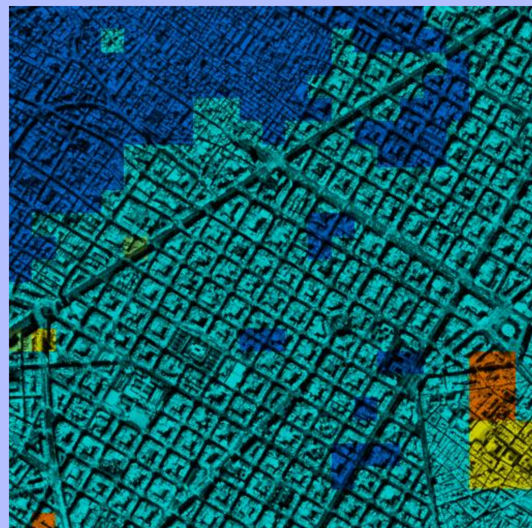


Unsupervised combination

Optimal number of clusters: 11







Barcelona

Istanbul

Los Angeles

Madrid

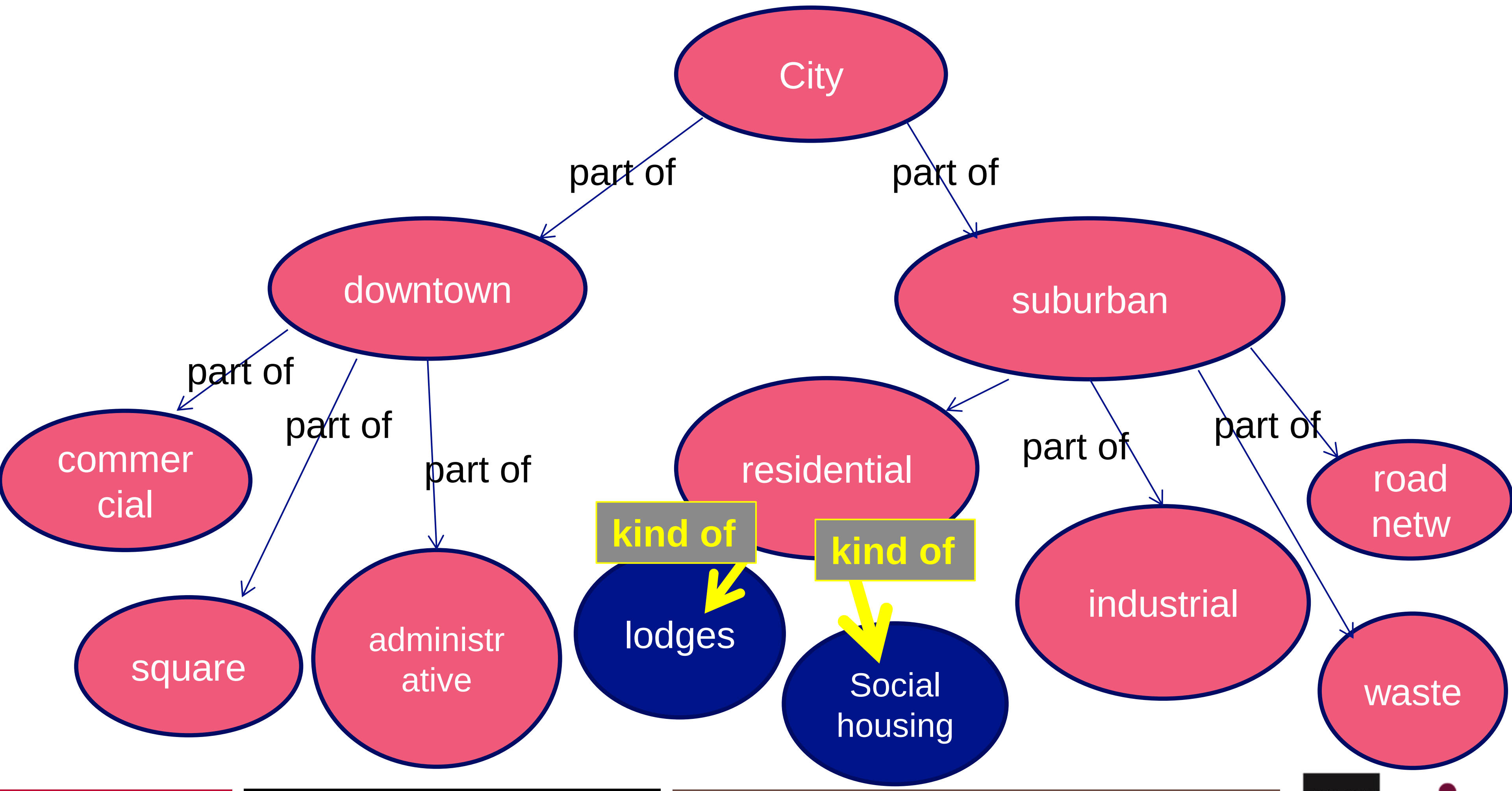
Towards an ontology

Ontology

- **Ontology is the philosophical study of the nature of being, …, as well as the basic categories of being and their relations**

- **What entities exist or may be said to exist and how such entities may be grouped, related within a hierarchy, and subdivided according to similarities and differences.**

Ontology


City
part of
part of
downtown
suburban
part of
part of
part of
commer
cial
square
administr
ative
residential
kind of
kind of
lodges
Social
housing
part of
part of
industrial
road
netw
waste

How to introduce semantics? Where are words coming from?

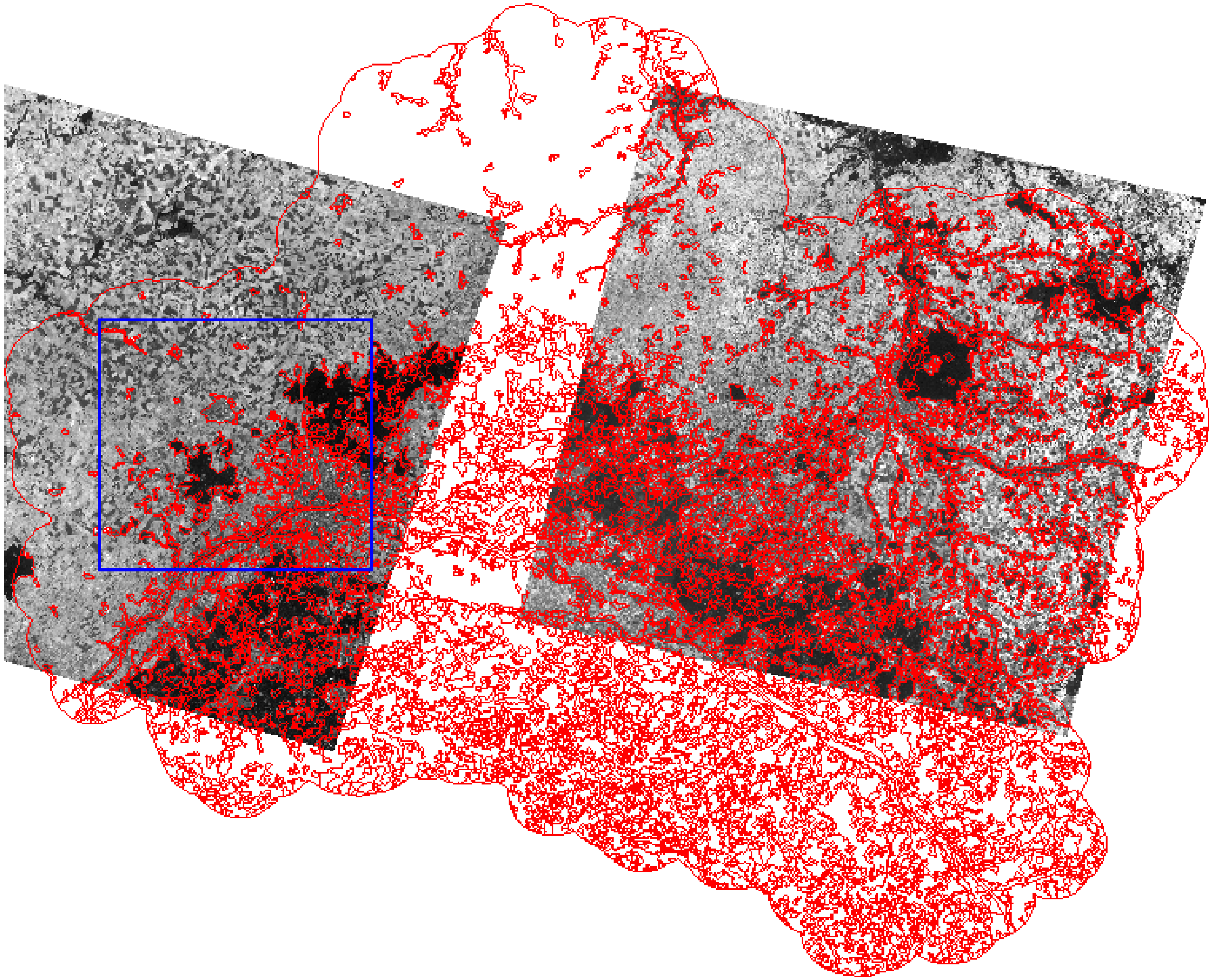
- **Supervised methods**
  - Fully manual indexing (experts or crowd sourcing)
  - Partly: learning (relevance feedback)
- **Contextual analysis of the document**
  - Tittle, caption, text, web site
- **Use of external data-bases**
  - Corine Land Cover (to learn classes and categories)
  - Maps and GIS (annotation)
- **Semantics inference**
  - Bayesian Modelling
  - Latent Models = Dirichlet, Blei & Jordan
  - « Ontological » deduction
  - Spatial reasonning

Example : CorineLandCover ontology

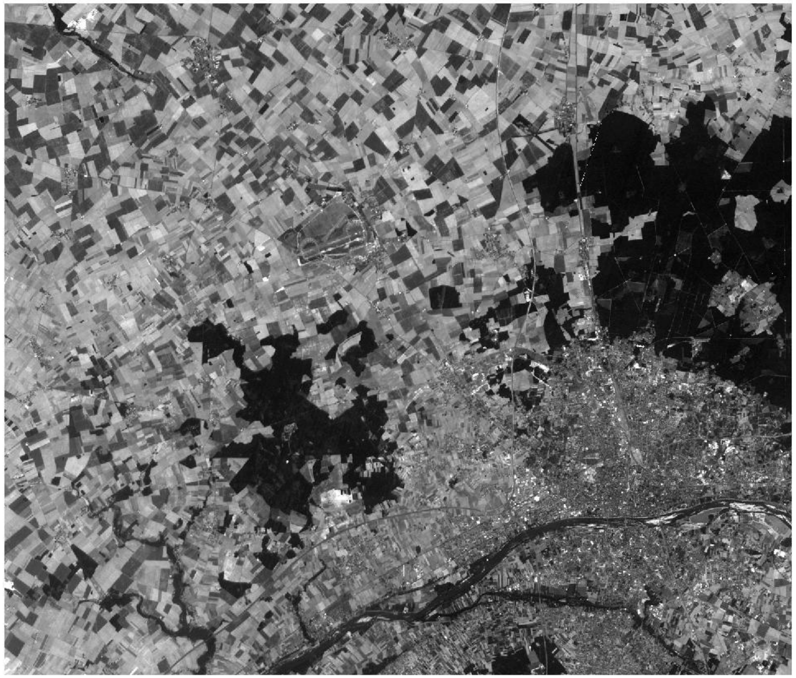
- 111: Continuous urban fabric
- 112: Discontinuous urban fabric
- 121: Industrial or commercial units
- 122: Road and rail networks and associated land
- …
- 211: Non-irrigated arable land
- 221: Vineyards
- 222: Fruit trees and berry plantations
- …
- 231: Pastures
- 242: Complex cultivation patterns
- 243: Land principally occupied by agriculture with significant areas of natural vegetation
- 311: Broad-leaved forests
- 312: Coniferous forests
- 313: Mixed forests
- …
- 411: Inland marshes
- …
- 511: Water courses
- …

Description of the test area

- Part of 39-253 SPOT scene, georeferenced and radiometrically corrected
- 1 760 536 pixels (1226 x 1436), 22 classes, unequally distributed

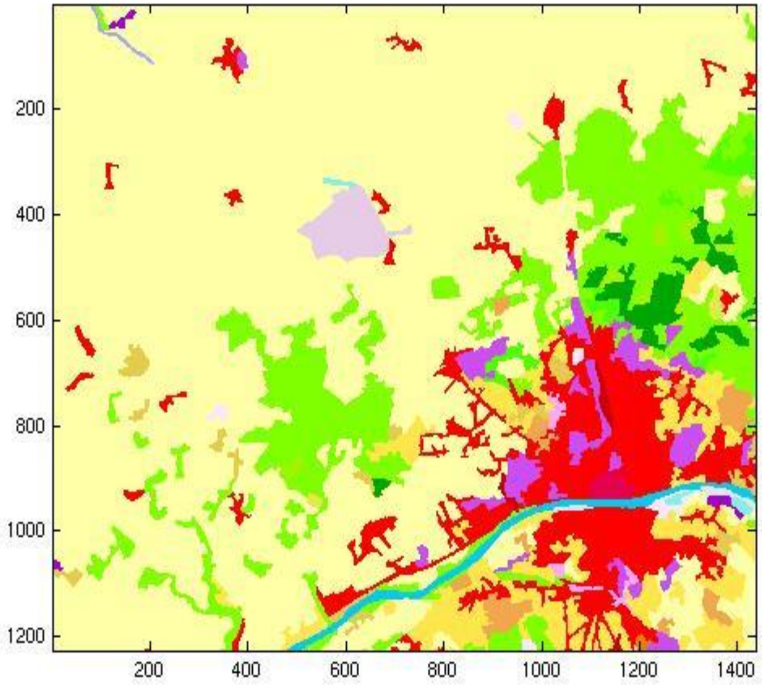


Description of the test area



SPOT image, band 3

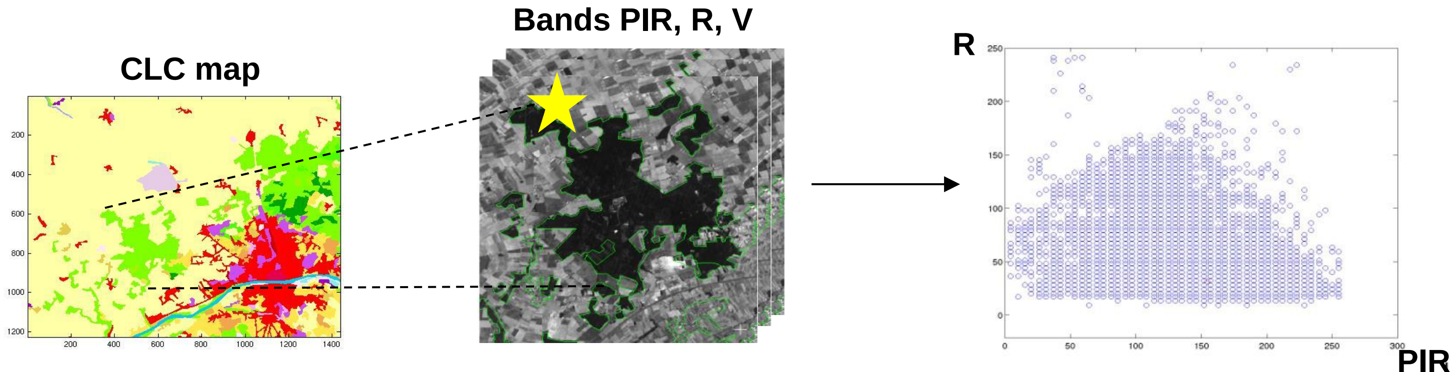
The corresponding CLC map



Learning and classification methodology

- **Assumption**
  - **Stationnarity, pixels independently distributed**
  - **Gaussian model**

- **Computation of the mean (m_i) and the covariance matrix (Σ_i) for each CLC class C_i.**



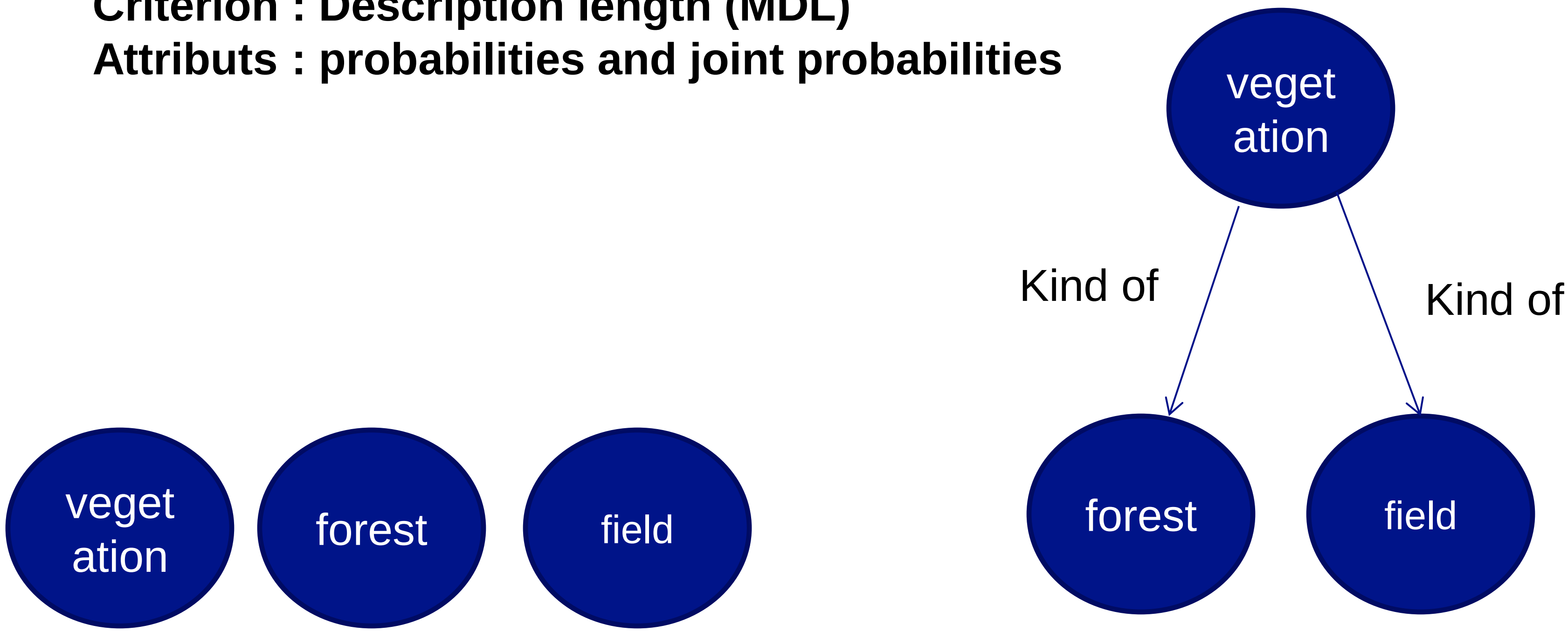
Automatic ontology building

- **Structural semantics between objects (inspired from linguistic methods)**
  - Sinonymy
  - Antonymy
  - Hyperonymy/hyponymy *kind of*
  - Meronymy/holonymy *part of*

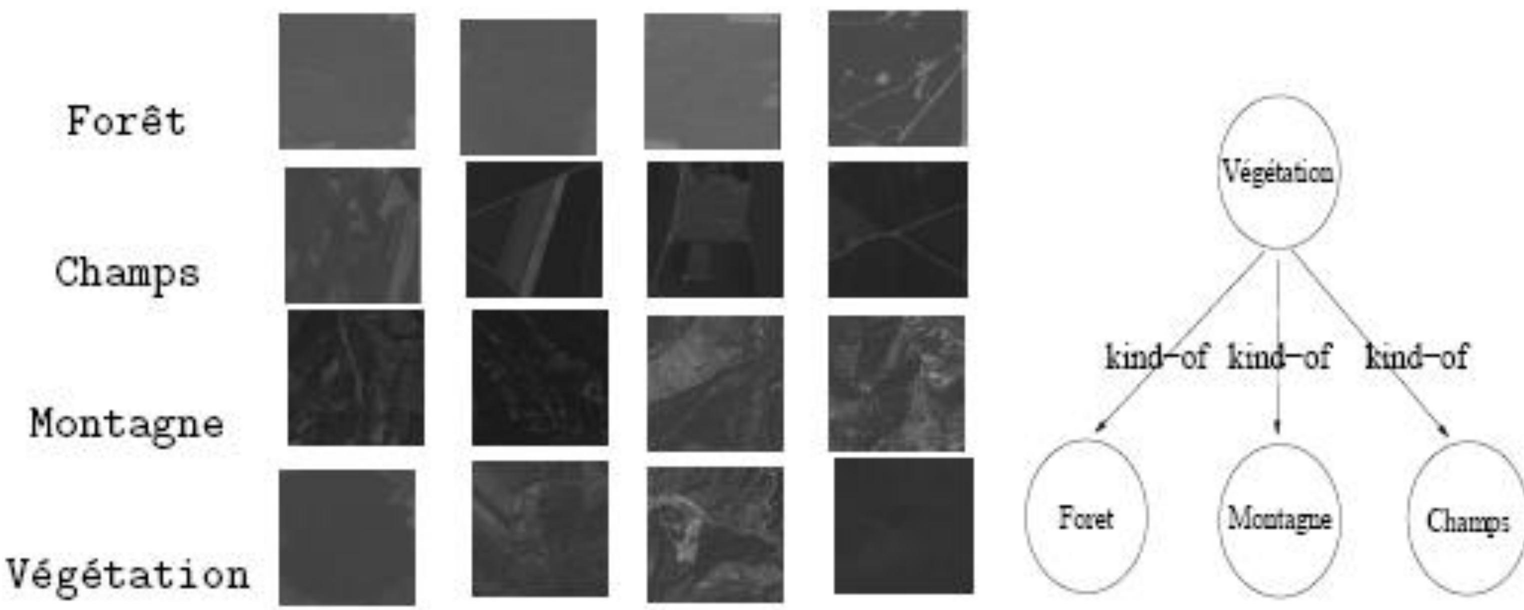
Methodology

Which model is the less expensive?

Criterion : Description length (MDL)
Attributs : probabilities and joint probabilities

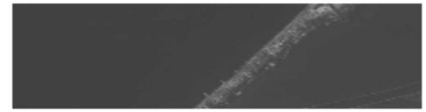


Hierarchy with semantic lincs

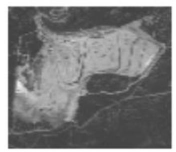


Using bag of words (BOWs) to learn hierarchies

Zone résidentielle



mer



carrière

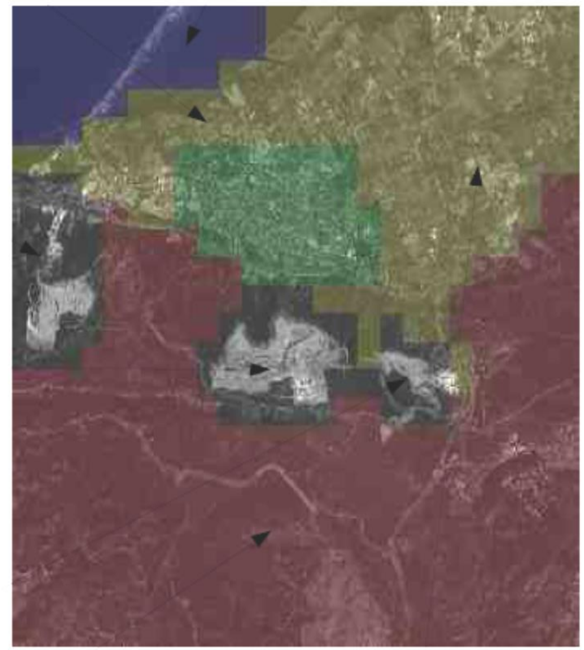
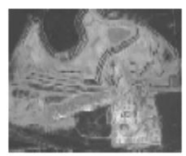
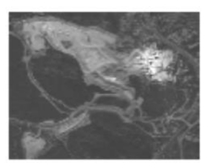


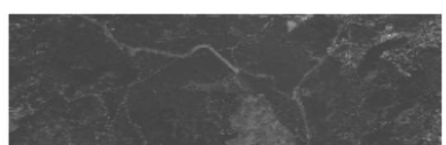
image test



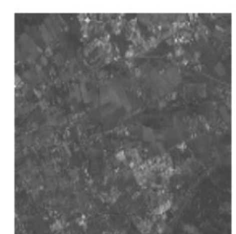
carrière



carrière



Zone montagneuse

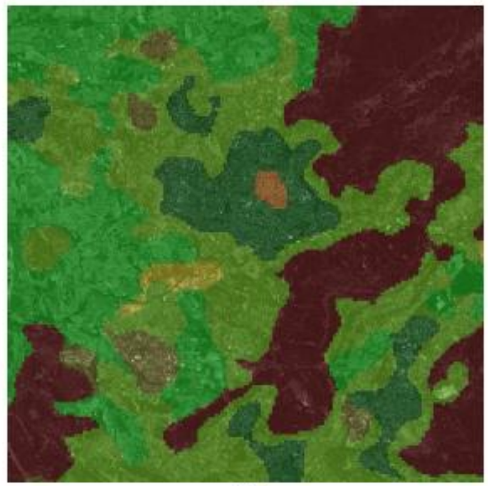


zone rurale

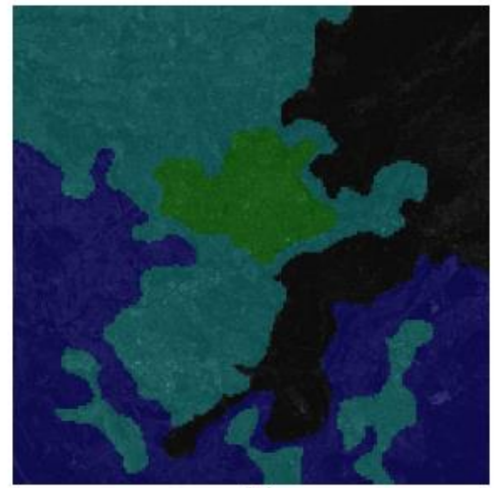
To learn a hierarchy

Low Level

High Level



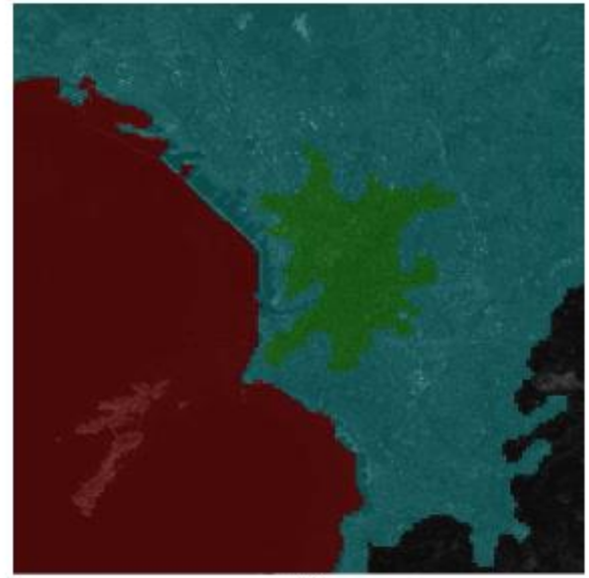
(a)



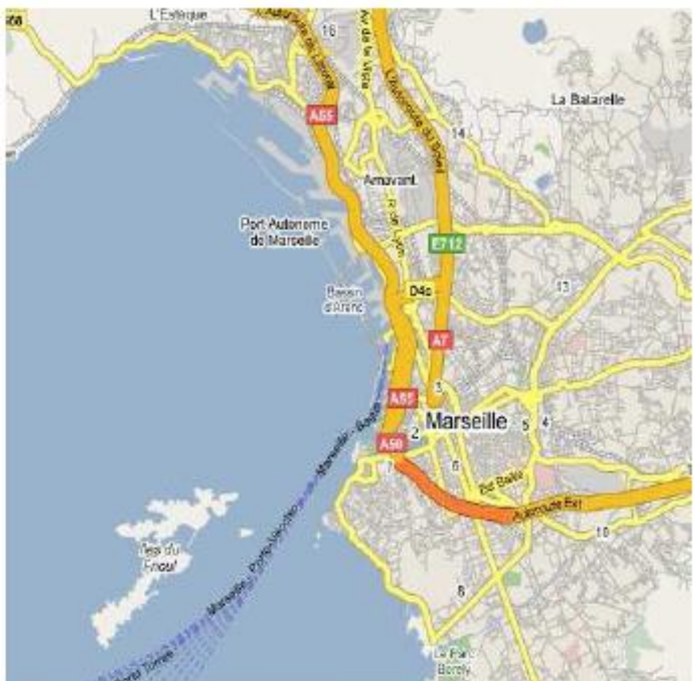
(b)

(c)

(a)



(b)


L'Estaque
La Batarelle
Port Autonome de Marseille
A55
E712
D4a
A7
A50
Marseille

(c)


Bois
Zone résidentielle
Champs
Aéroport
Zone industrielle
Habitations éparses
Mer
Centre ville
Carrière
Cimetière
Zone montagneuse
Zone rurale
Agglomération
Banlieue industrielle
Zone maritime