Classifications

- **Automatic**
- **Supervised**
- **Principal Components Analysis (PCA)**

Pierre-Louis Frison *pierre-louis.frison@u-pem.fr*

Université Gustave Eiffel

LaSTI
Laboratoire des Sciences et Technologies de l'Information Géographique

Références

- Image analysis, Classification, and change detection in Remote sensing

  Canty, ed. CRC Press, 2010

- Introduction to remote sensing, 5ième édition

  Campbell and Wynne, ed. the guilfor Press, 2011

- Remote Sensng and Image Interpretation, 6ième édition

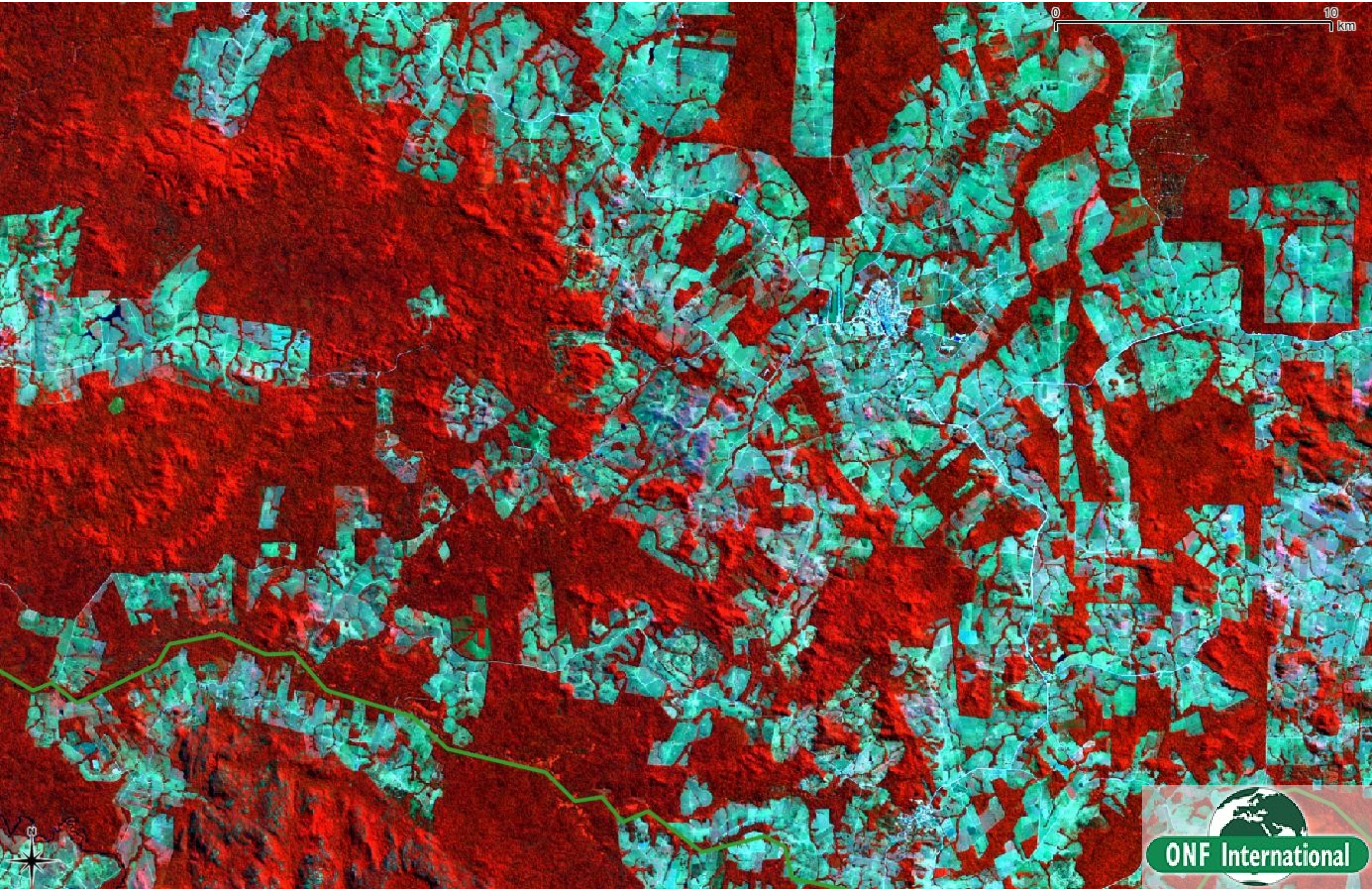
  Lillesand, Kieffer and Chipman, ed J. Wiley & sons, 2007

CLASSIFICATION

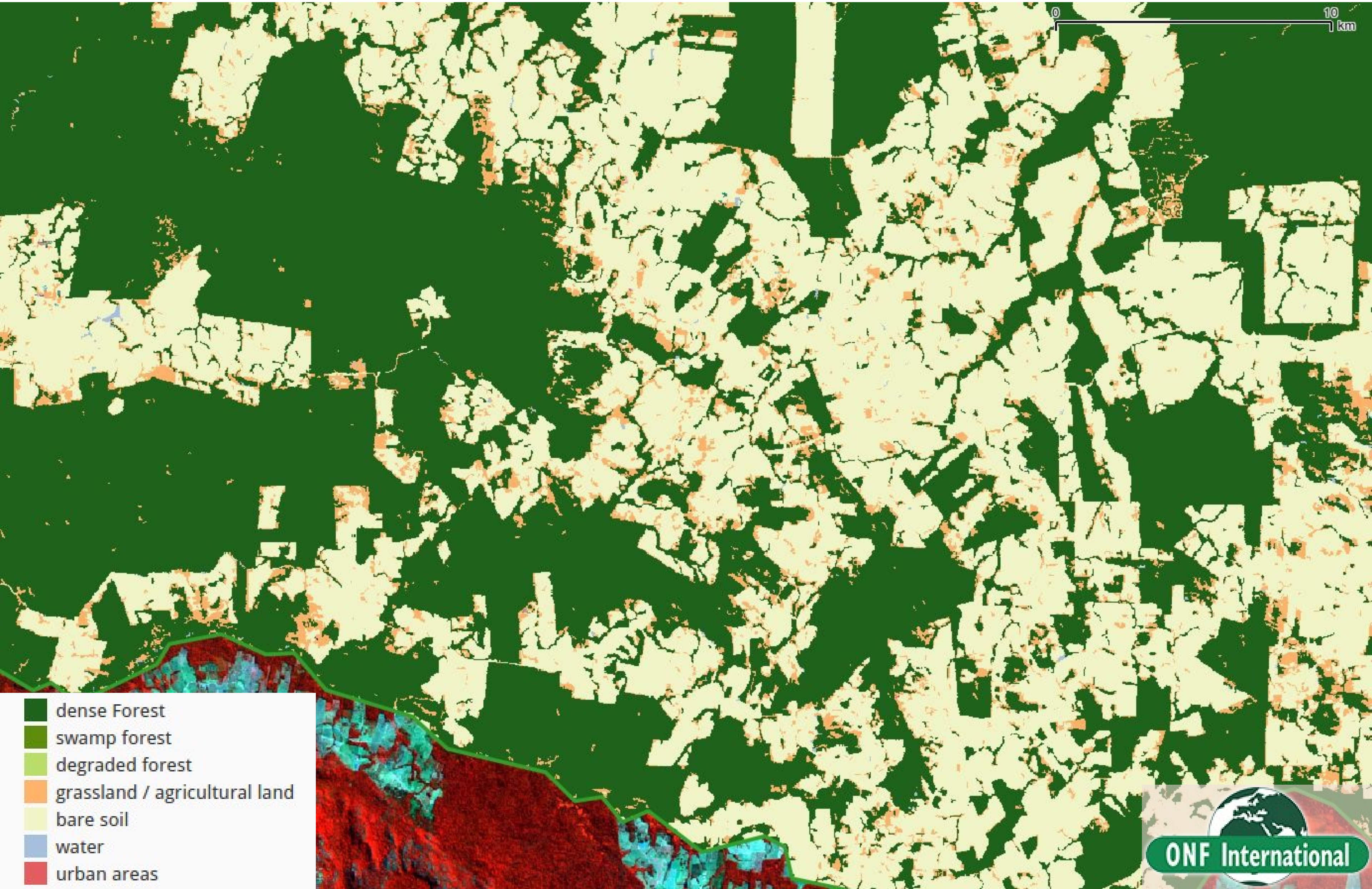
Remote Sensing Image

Thematical Map
(Land Cover, land use, forest types, cloud types,….)

LANDSAT TM image



Classification result



QUELQUES RÉSULTATS

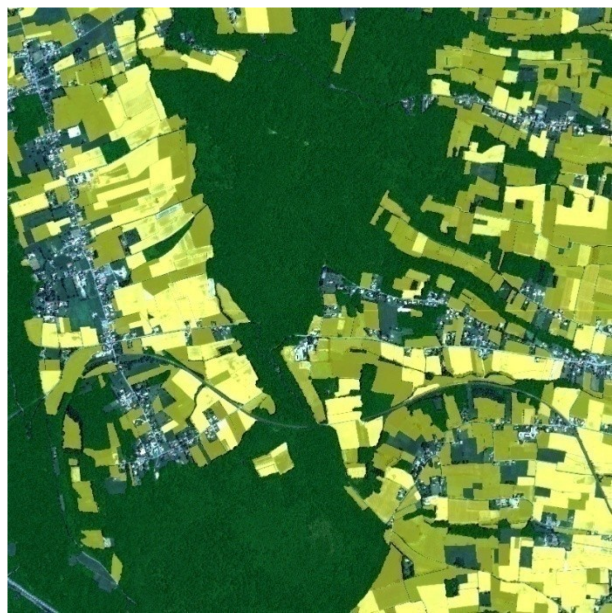
- **Détection de changements entre une BD 2D topographique et une image aérienne ou satellite**

BD OCS | Classification

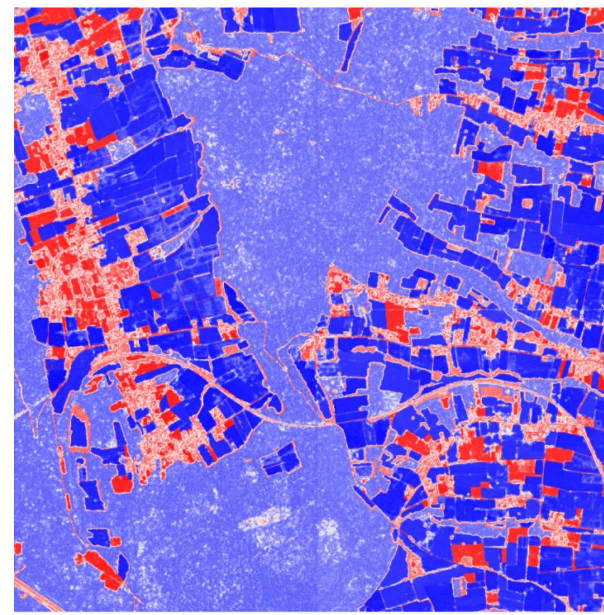
Image(s) → OCS → Carte de changements

BD OCS

Image Pléiades (0,5m)



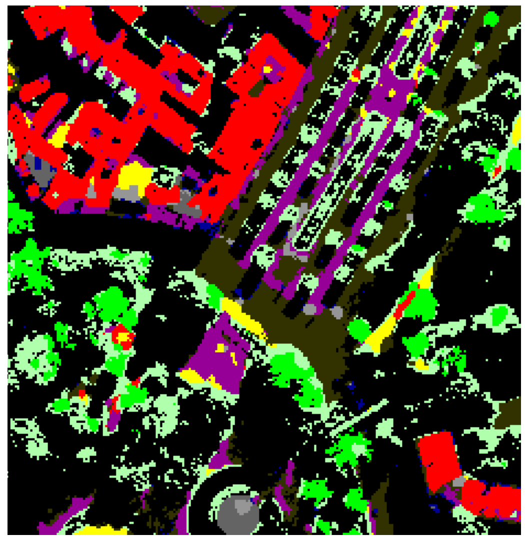
RGFor - RPG

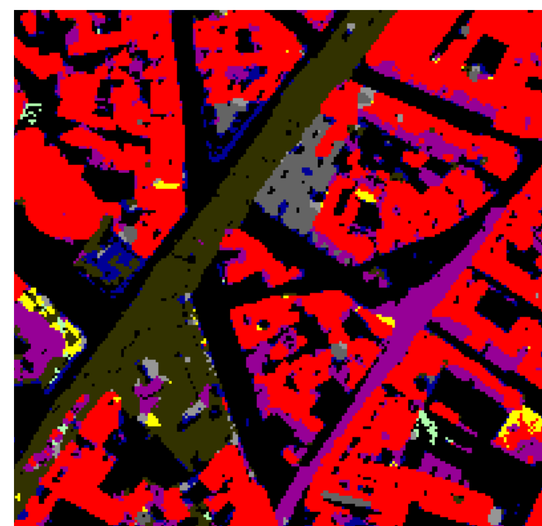


Non changement/changement

QUELQUES RÉSULTATS

- **OCS enrichie par imagerie hyperspectrale**

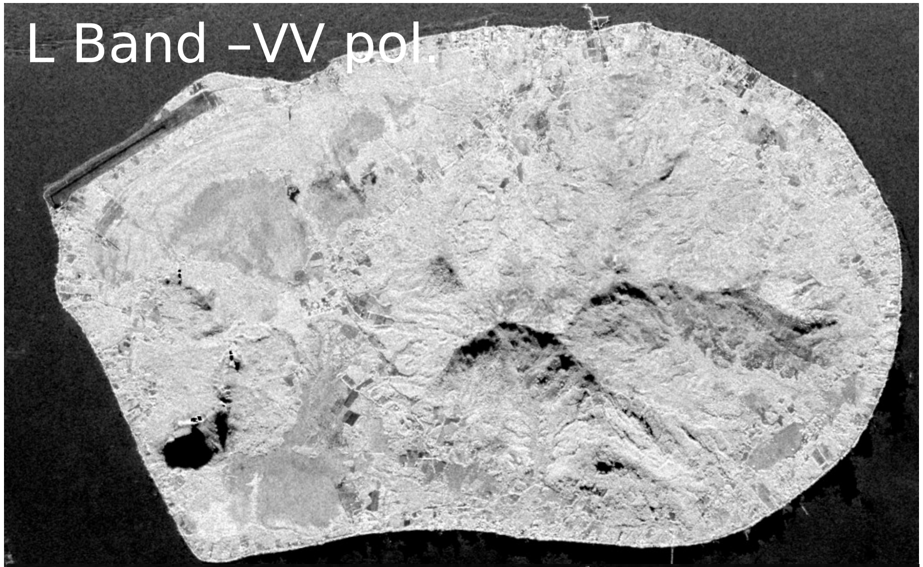


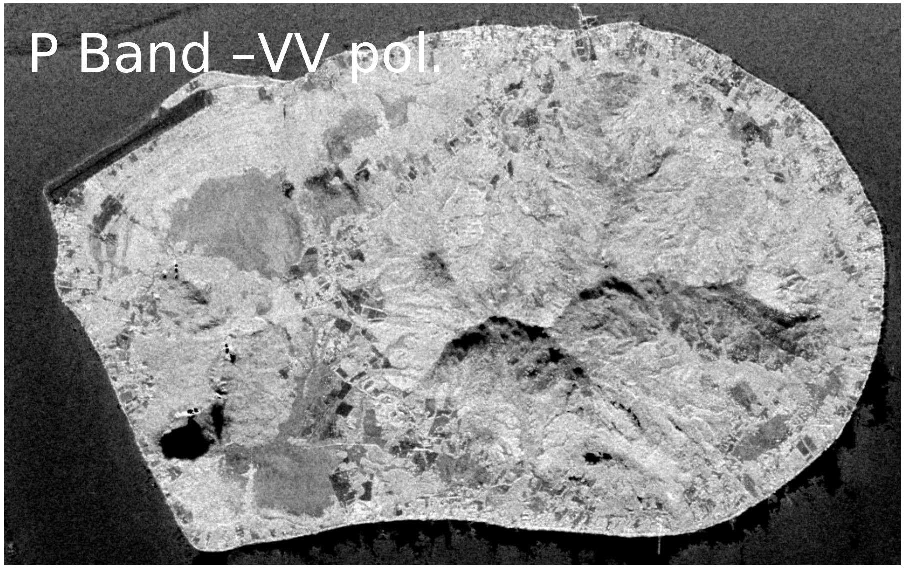


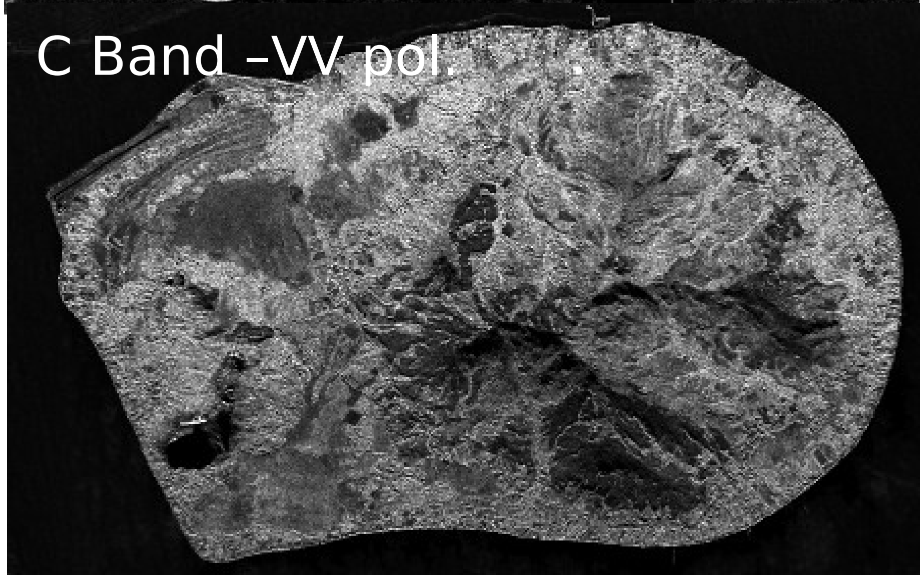
Campagne Umbra
(IGN-ONERA)
0,8m → 1,6m

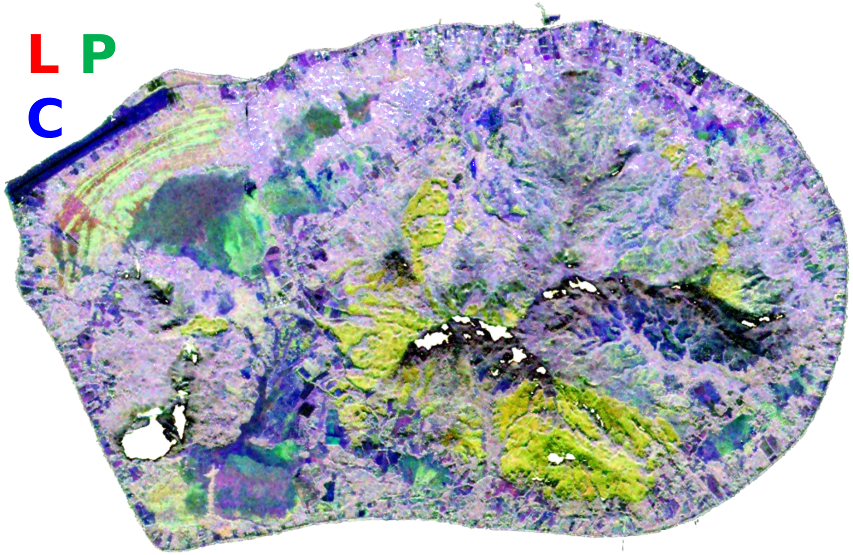
Tuiles rouges
Métal 1
Métal 2
Ardoises
Pavés
Asphalte
Sol nu
Végétation haute
Végétation basse

Frequency diversity – Wavelength



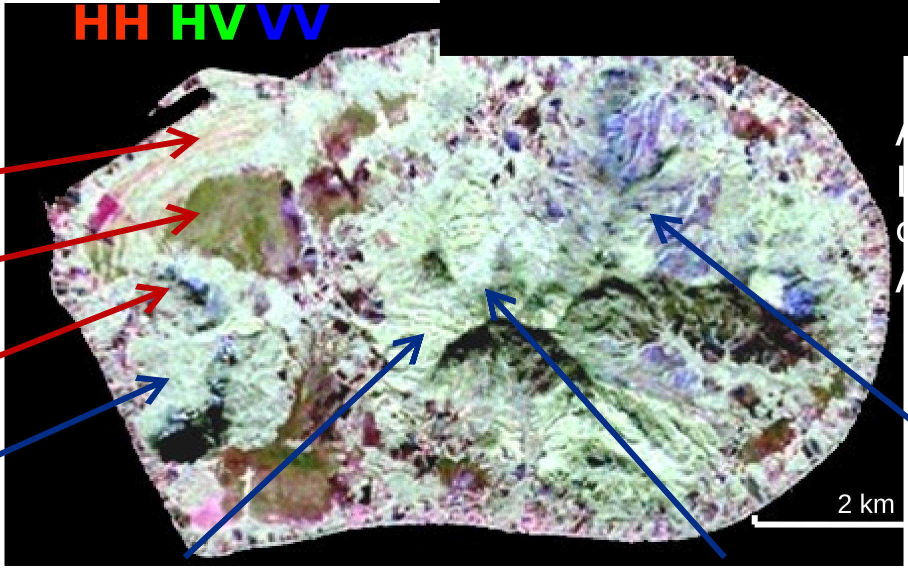


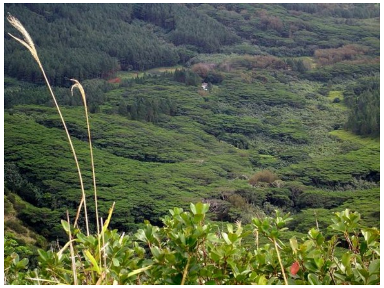




radar polarimetry fot forest cartogrphy

Tubuai Island, French Polynesia

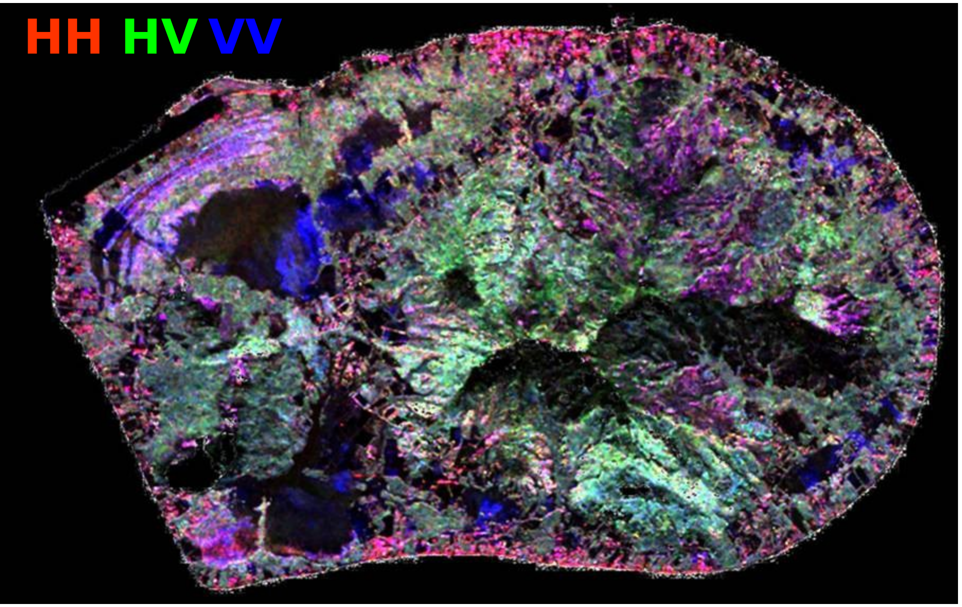




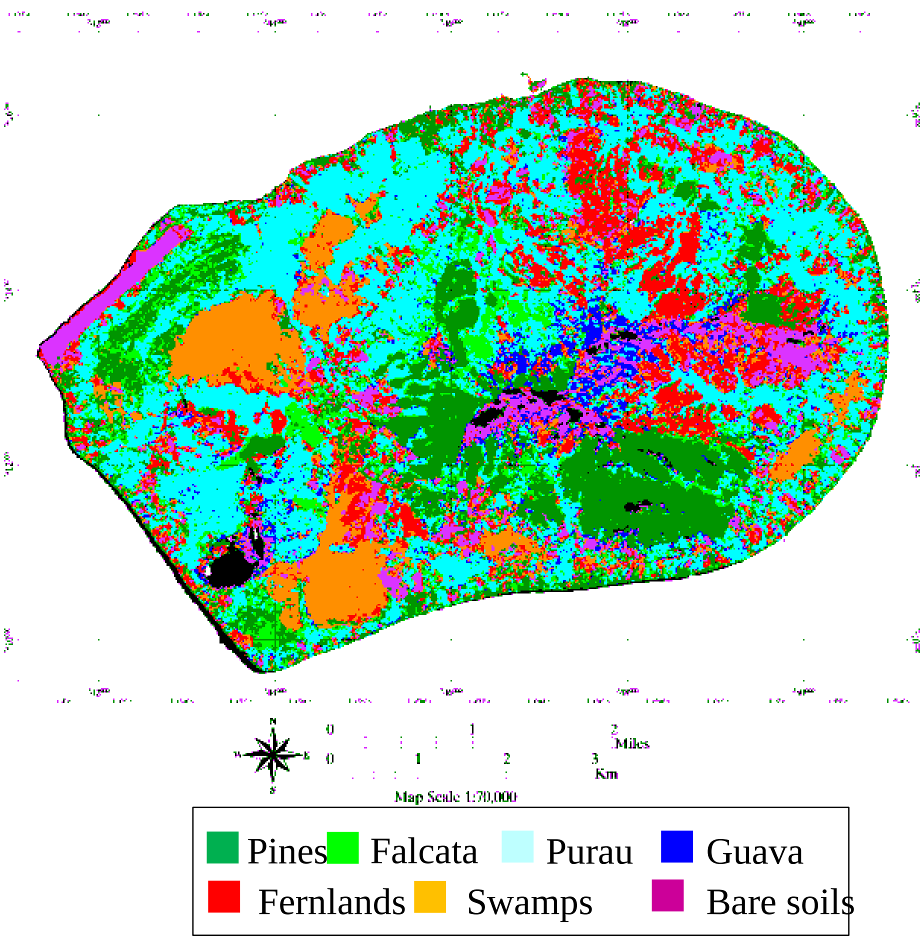
tation classification with radar polarimetry

Tubuai Island (French Polynesia)

AIRSAR data
August 2000



Classification



(4 Forest types)

CLASSIFICATION

Based on information:

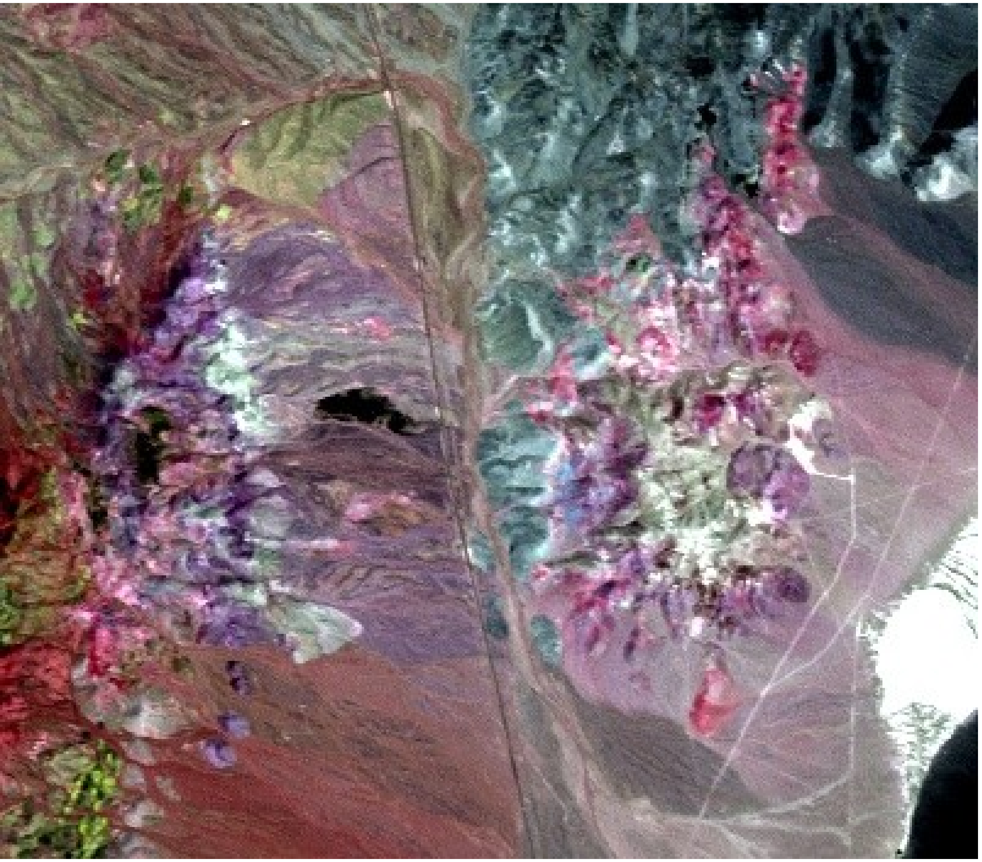
spectral

And/or ***spatial*** (neighborhood, adjacent pixels, ***texture***)

And/or shape (object ***oriented***)

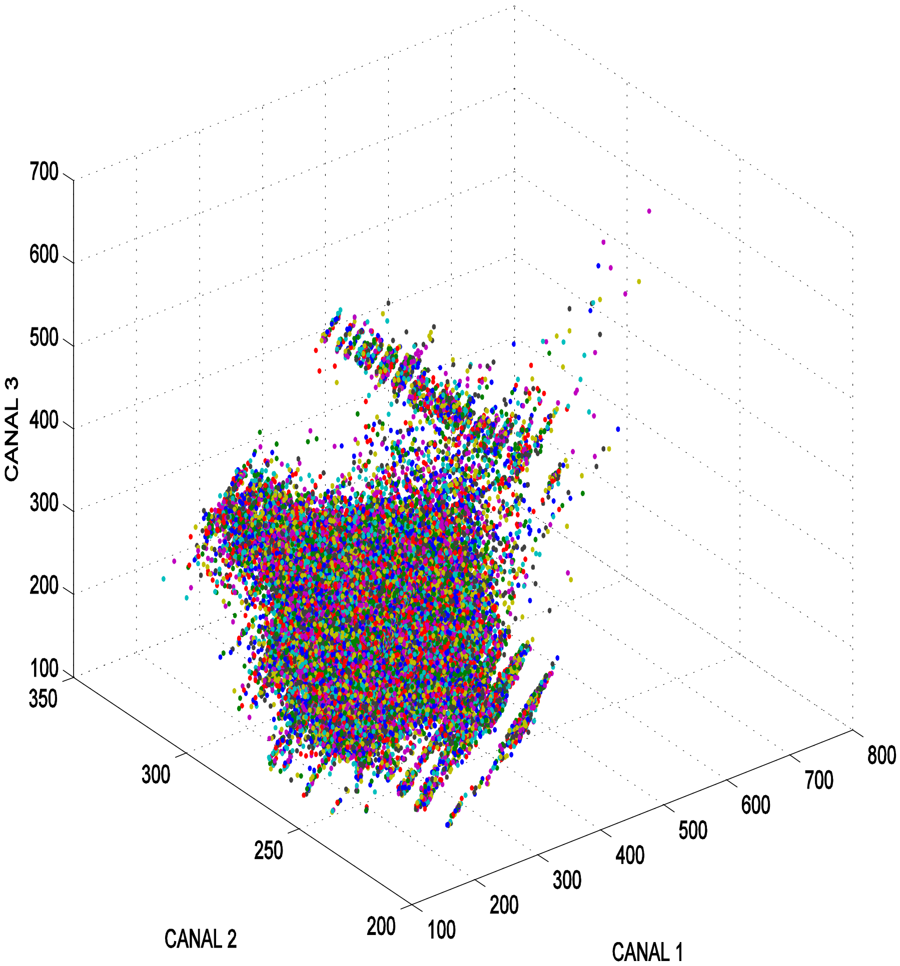
SPECTRAL INFORMATION

Spatial domain



RIS airborne image, Cuprite, USA

Spectral domain



Example of automatic classification

The k-means algorithm

Based on minimum distance criteria in the spectral domain

1) Segmentation of the spectral space into k sub-spaces

(*k* set *a priori*)

Next iteration:

2) Computes the barycentre of each sub-space
3) Pixel affectation to the class with closest barycentre

Repeat 2) and 3) until

a maximal number of iterations (set *a priori*) t

or differences between barycentre position < ε ($\mathcal{E}$ threshold set *a priori*)

CLASSIFICATION

Basée sur l'exploitation de l'information:

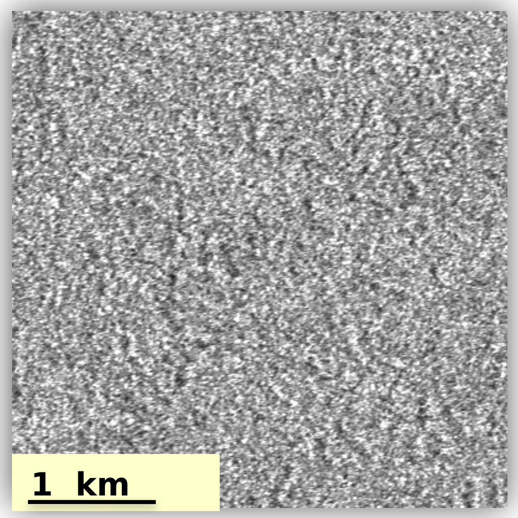
spectrale

et/ou ***spatiale*** (notion de voisinage, pixels contigus, ***texture***)

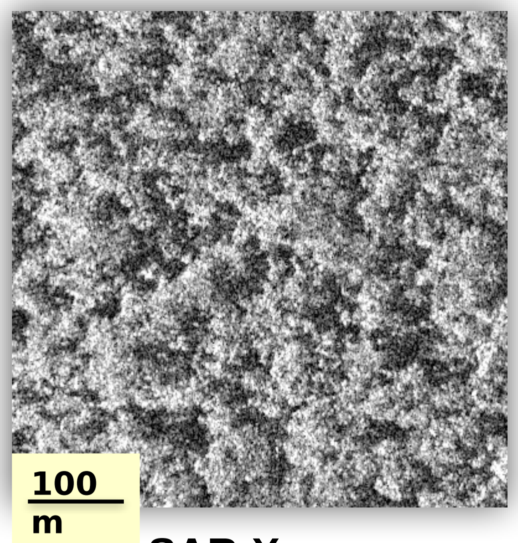
et/ou forme (***orientée objet***)

INFORMATION TEXTURALE

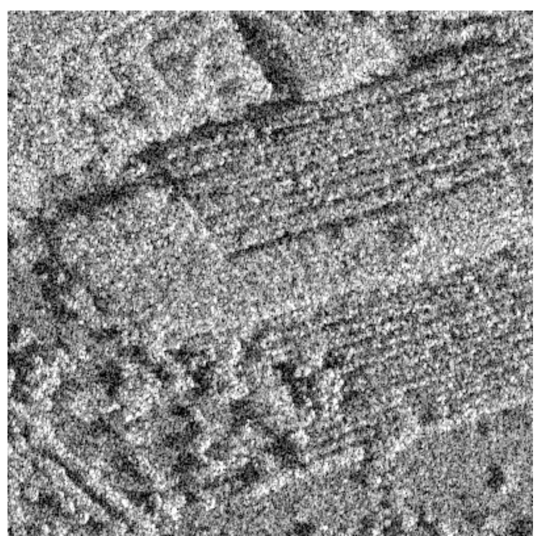
Données RADAR



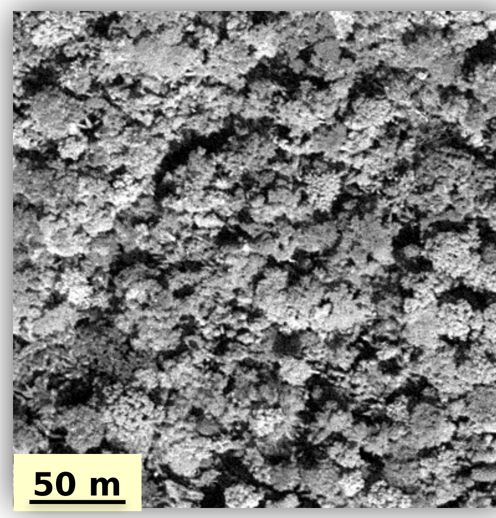
PALSAR, HH, mode FBD,
Résolution **: 30 m**



TerraSAR-X, SpotLight-HH,
Résolution **: 1 m**



Donnée Optique



Geoeye, Panchromatique**,**
Résolution **: 0.5 m**

CLASSIFICATION

Basée sur l'exploitation de l'information:

spectrale

et/ou ***spatiale*** (notion de voisinage, pixels contigus, ***texture***)

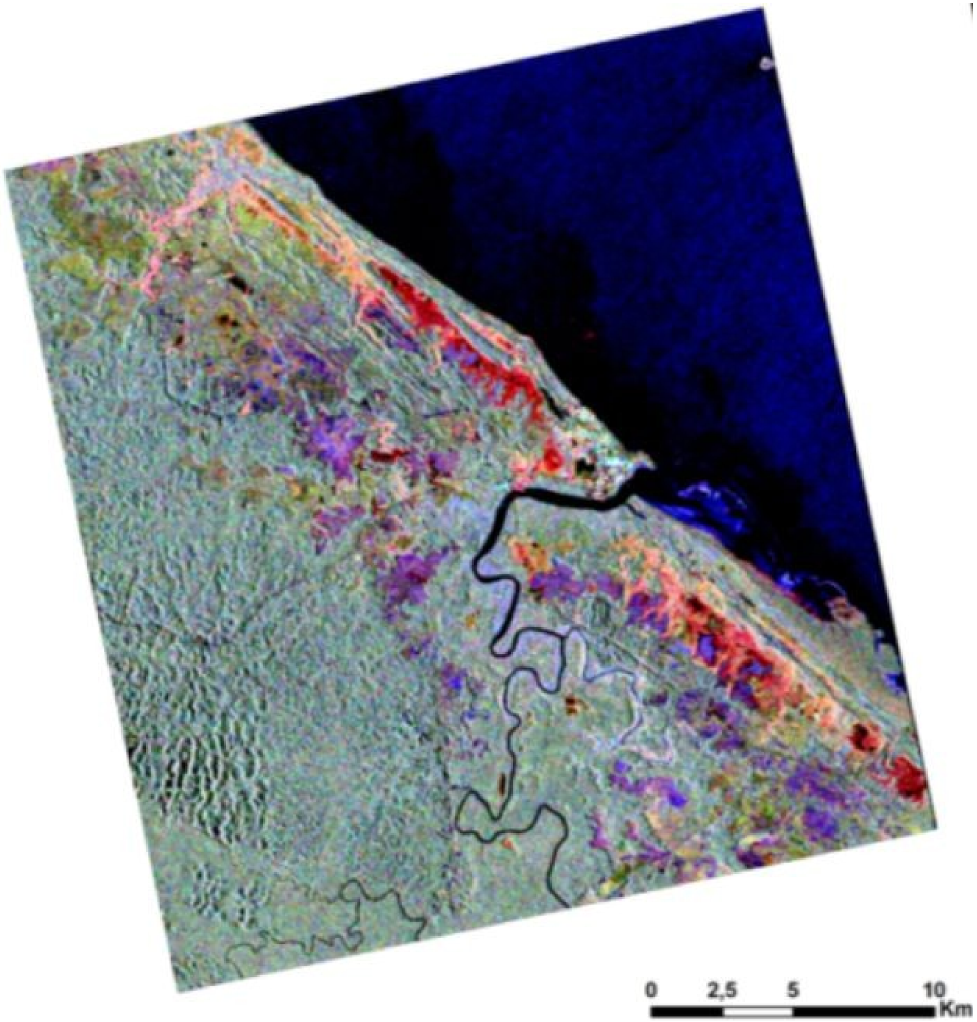
et/ou forme (***orientée objet***)

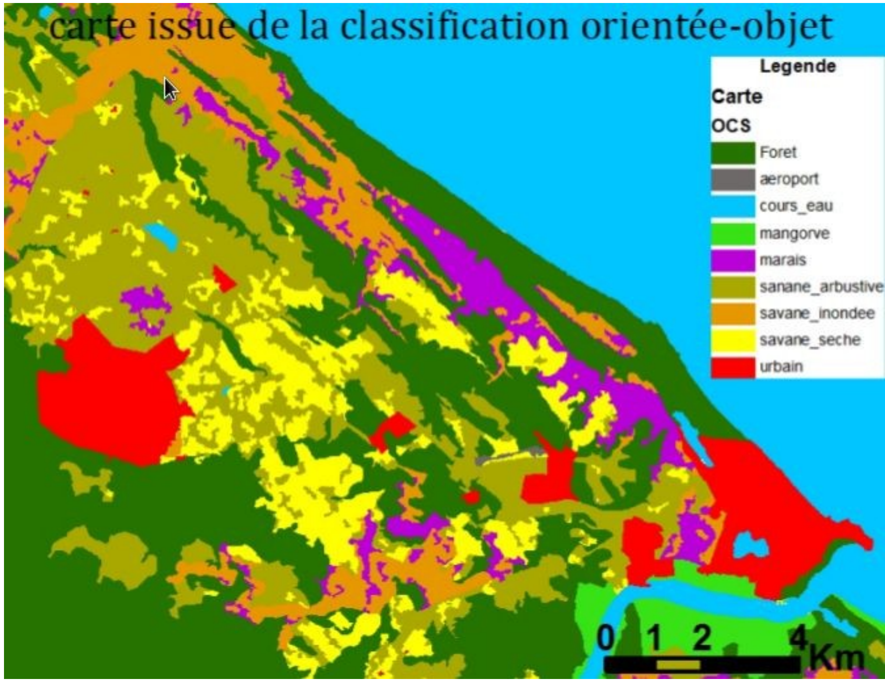
CLASSIFICATION ORIENTEE OBJET

Source: Geosystems

CLASSIFICATION ORIENTEE OBJET

Image RADARSAT2





CLASSIFICATION

2 types d’algorithmes

Classifications automatiques (non supervisées)

L’algorithme détermine des classes automatiquement

Classifications supervisées

Des zones d’intérêt sont définies par l’utilisateur

(relevés terrain, photo-interprétation)

L’algorithme calcule les propriétés de chaque classe d’apprenti

Il généralise la classification à l’ensemble de l’image à partir de

CLASSIFICATION

2 types d'algorithmes

Classifications automatiques (non supervisées)

L'algorithme détermine des classes automatiquement

Classifications supervisées

Des zones d'intérêt sont définies par l'utilisateur

(relevés terrain, photo-interprétation)

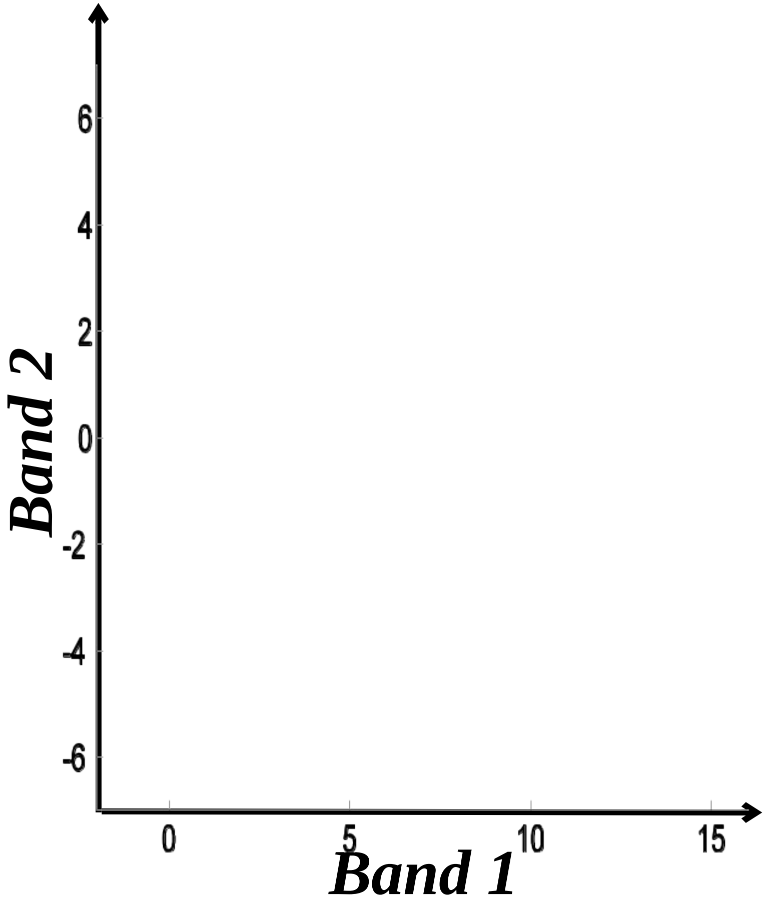
L'algorithme calcule les propriétés de chaque classe d'apprenti

Il généralise la classification à l'ensemble de l'image à partir de

THE K-MEANS ALGORITHM

Example for an image with 2 bands and 3 classes

Data to process



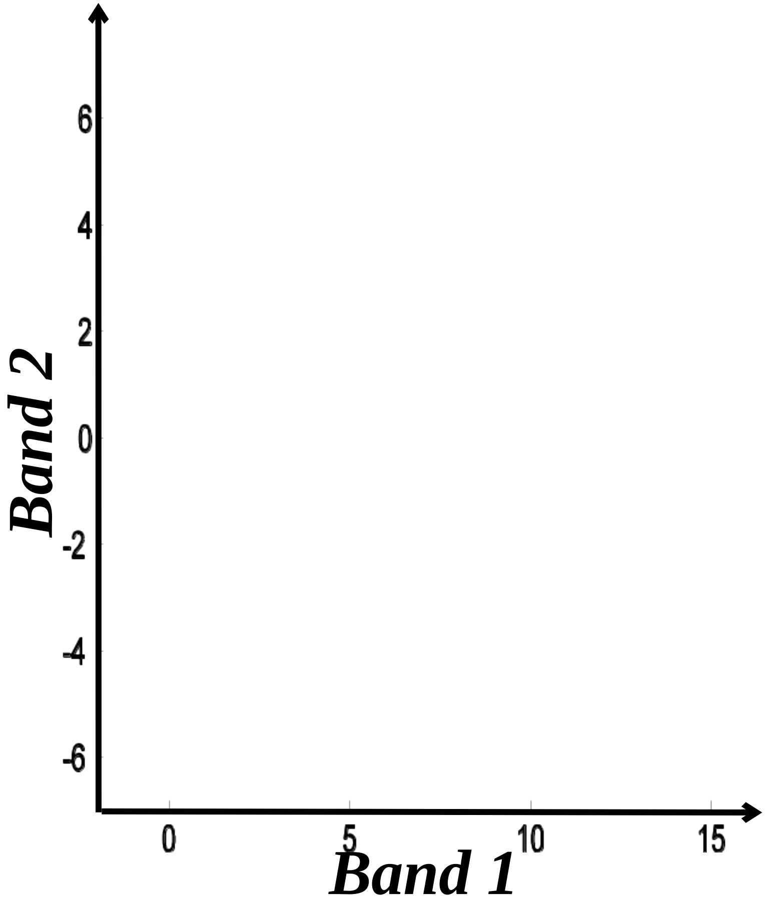
Spectral domain

Class 1
Class 2
Class 3

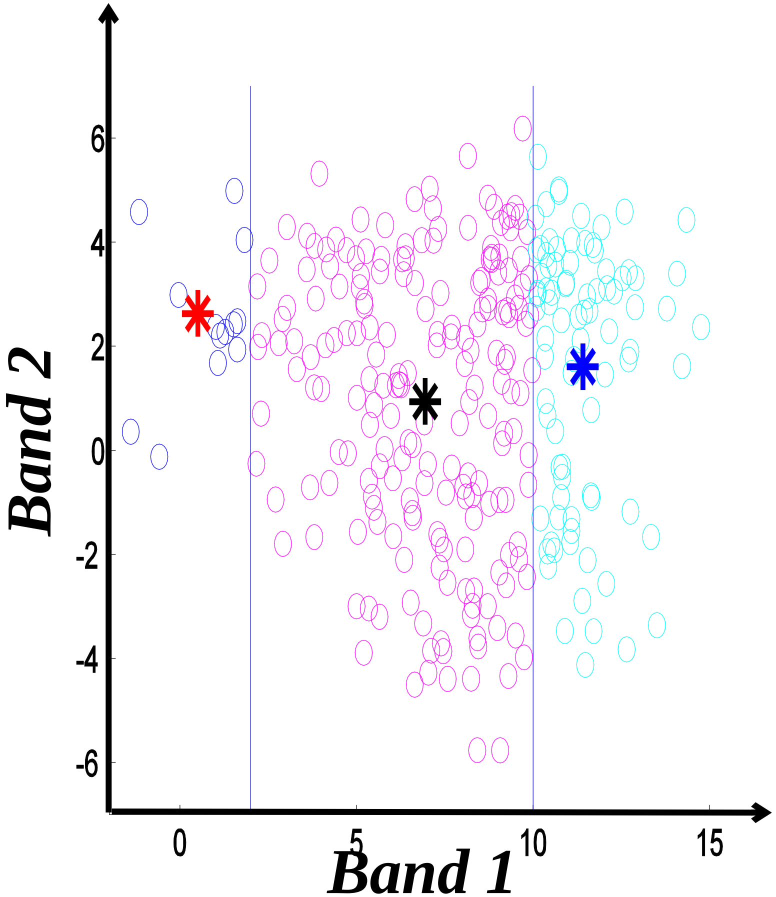
THE K-MEANS ALGORITHM

Example for an image with 2 bands and 3 classes

Data to process



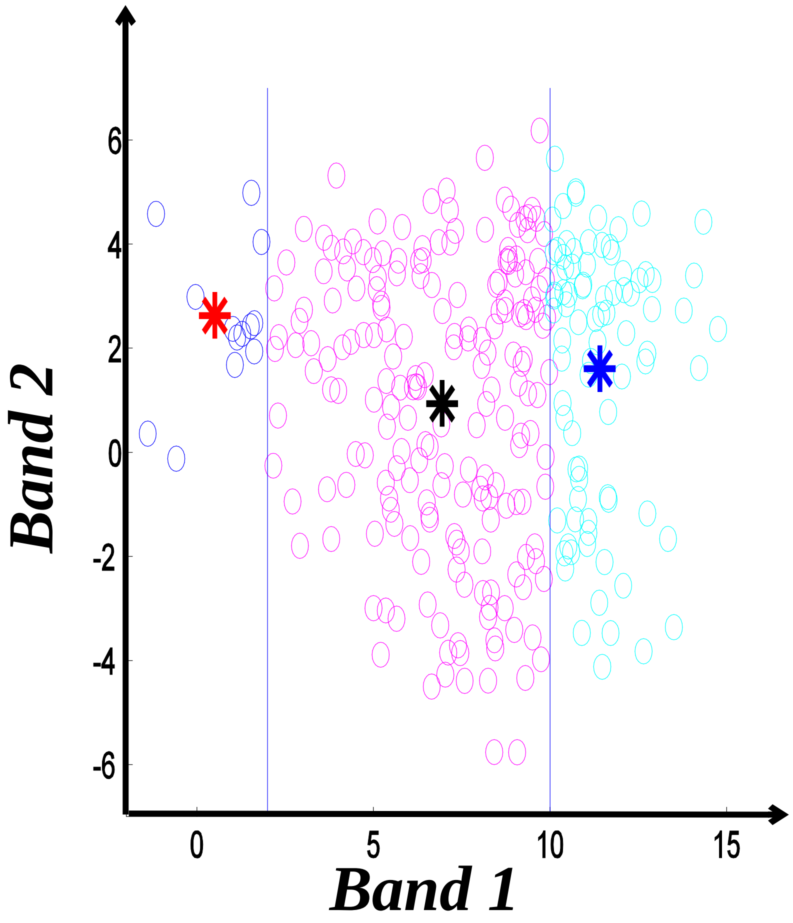
Initialisation



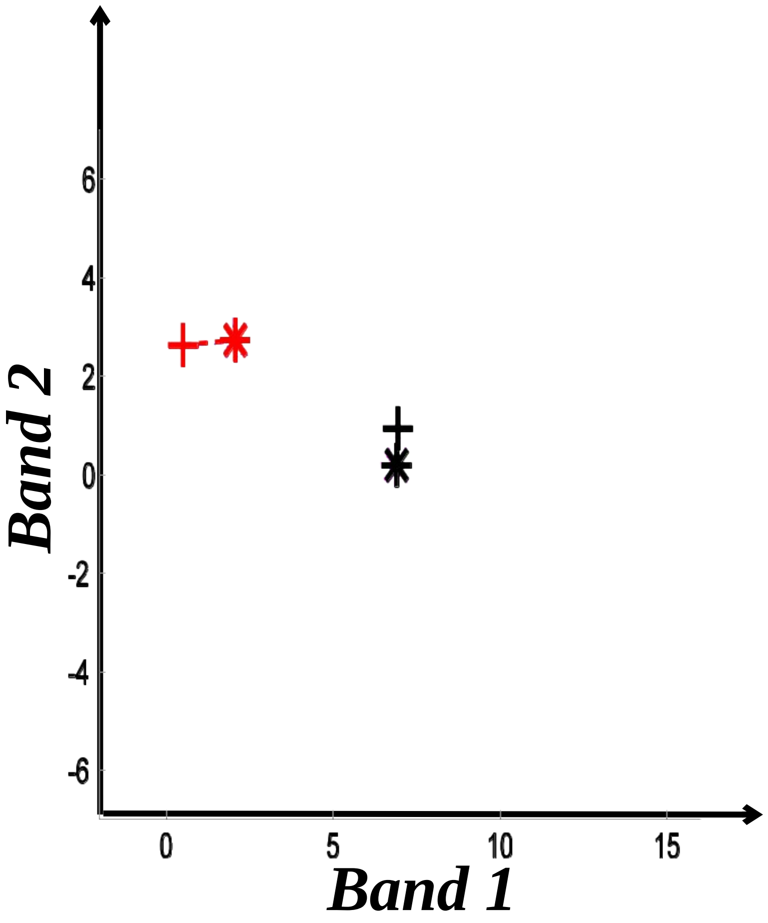
THE K-MEANS ALGORITHM

Example for an image with 2 bands and 3 classes

Initialisation



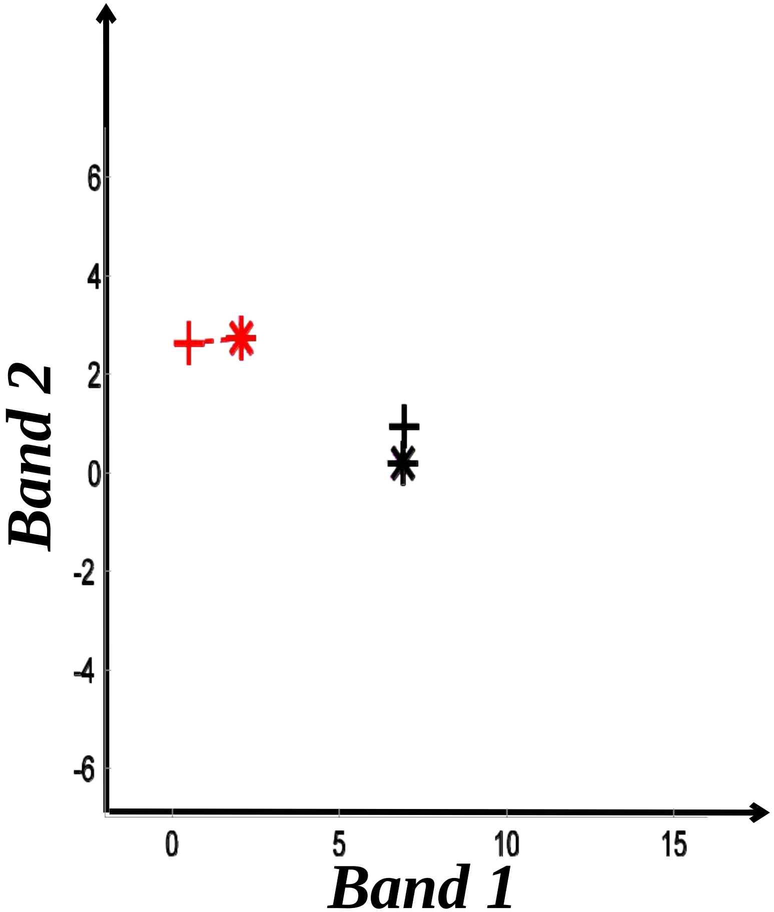
1rst iteration



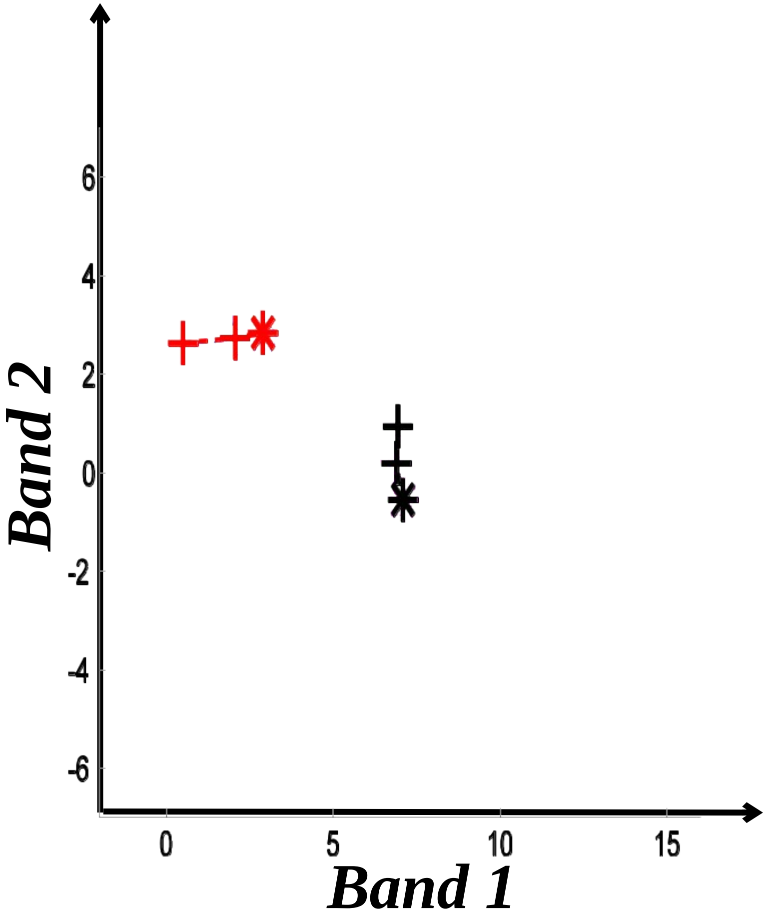
THE K-MEANS ALGORITHM

Example for an image with 2 bands and 3 classes

1rst itération



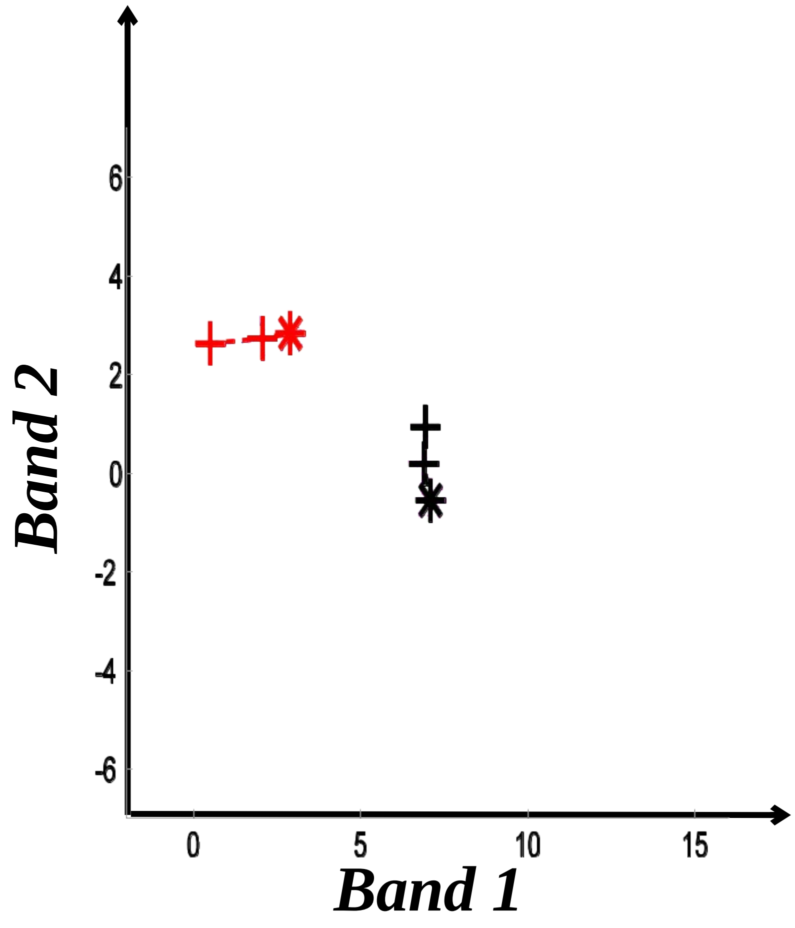
2nd itération



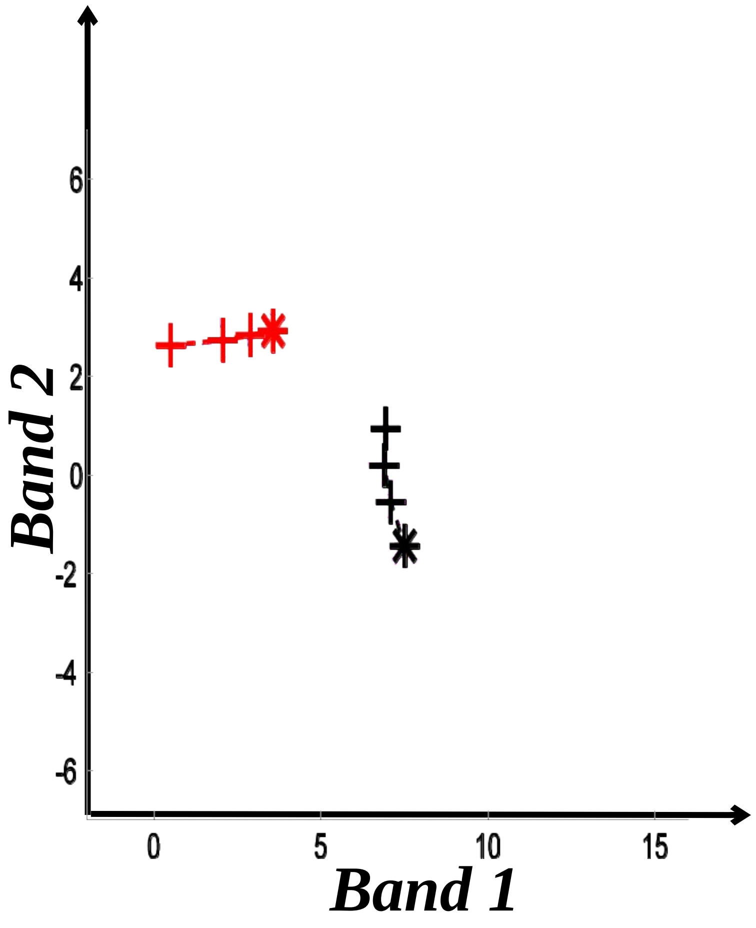
THE K-MEANS ALGORITHM

Example for an image with 2 bands and 3 classes

2nd itération



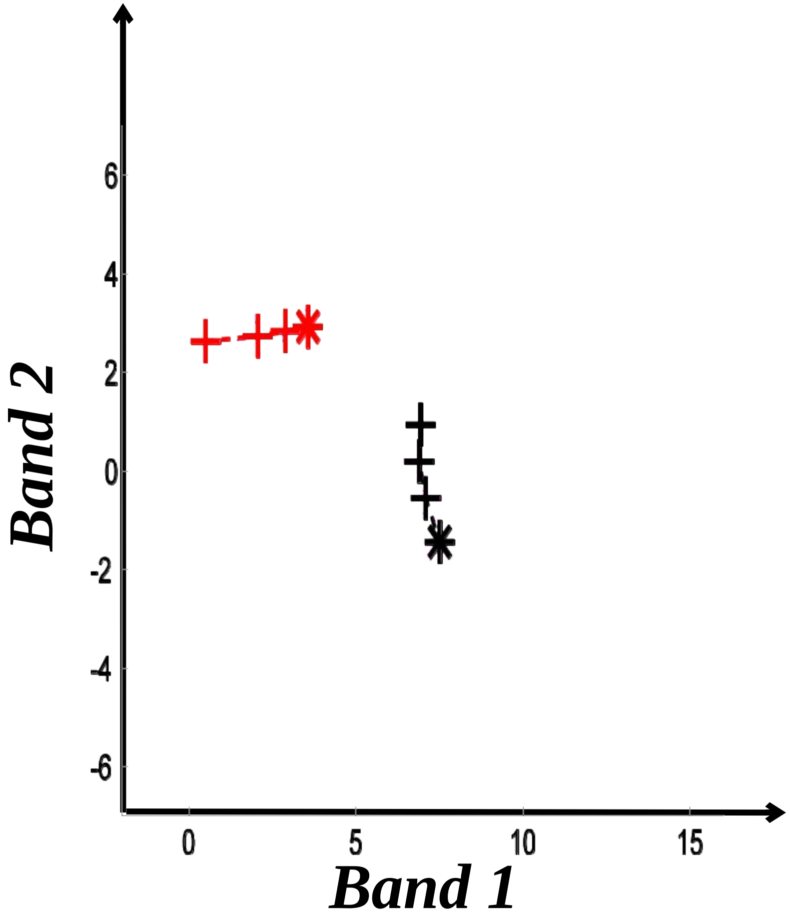
3rd itération



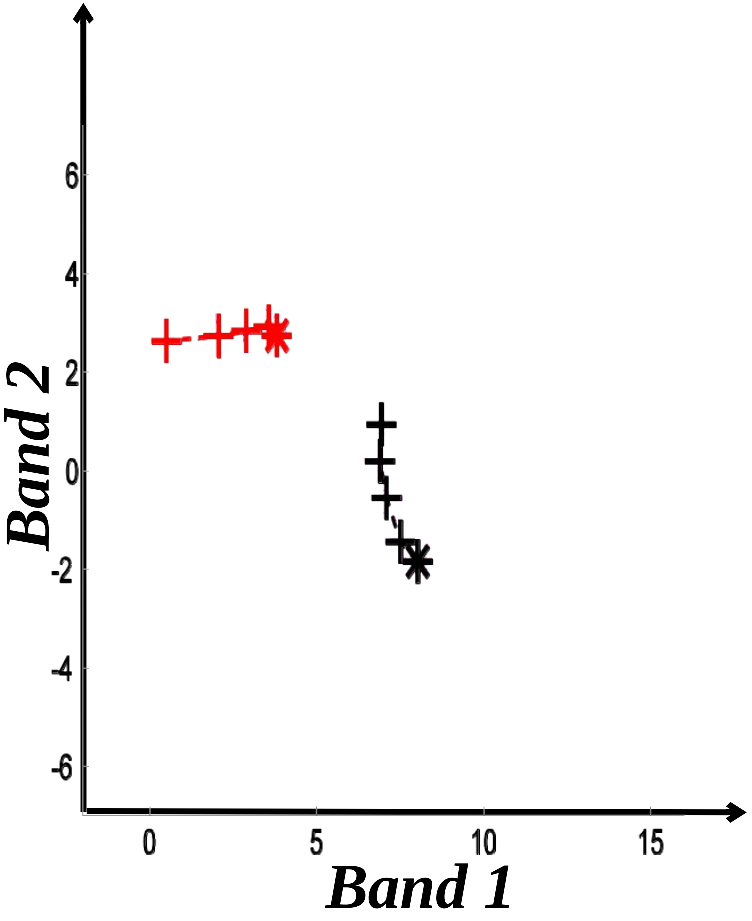
THE K-MEANS ALGORITHM

Example for an image with 2 bands and 3 classes

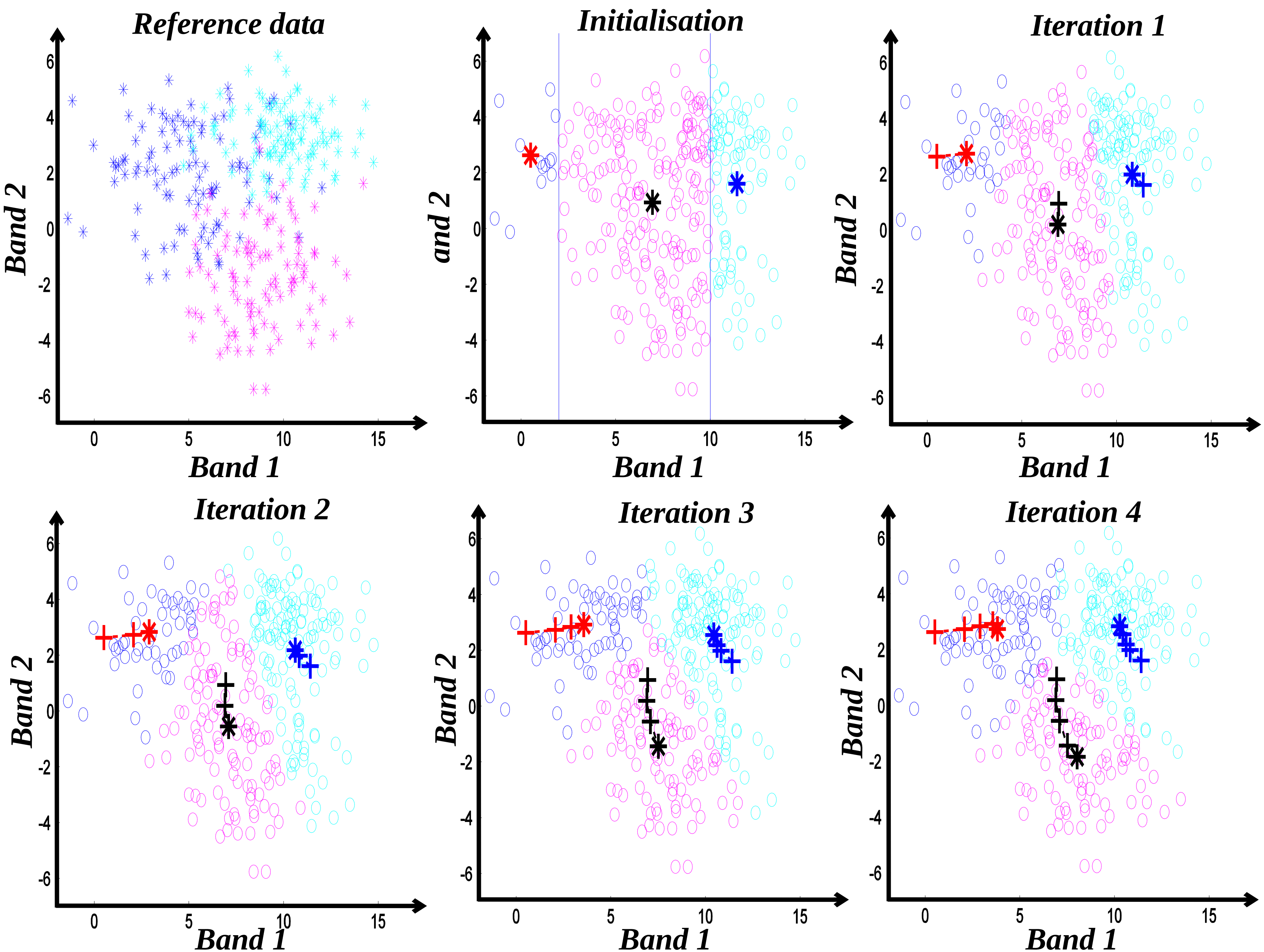
3^{rd} itération



4^{th} itération



THE K-MEANS ALGORITHM



SUPERVISED CLASSIFICATION

*1) The user defines **polygons** on the image*

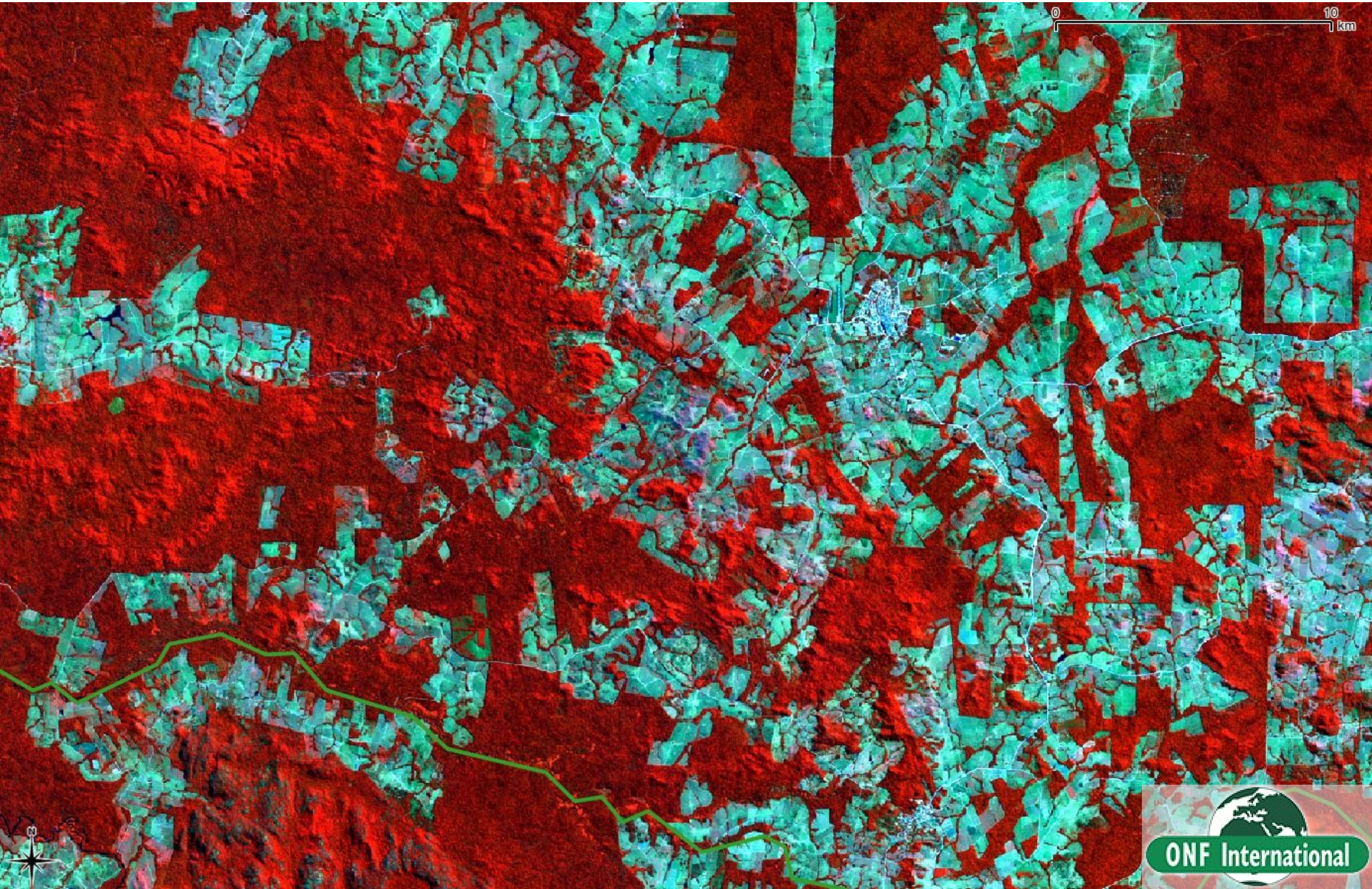
Several polygons representative of each classes to be determined

(photo-interprétation, ground survey,....)

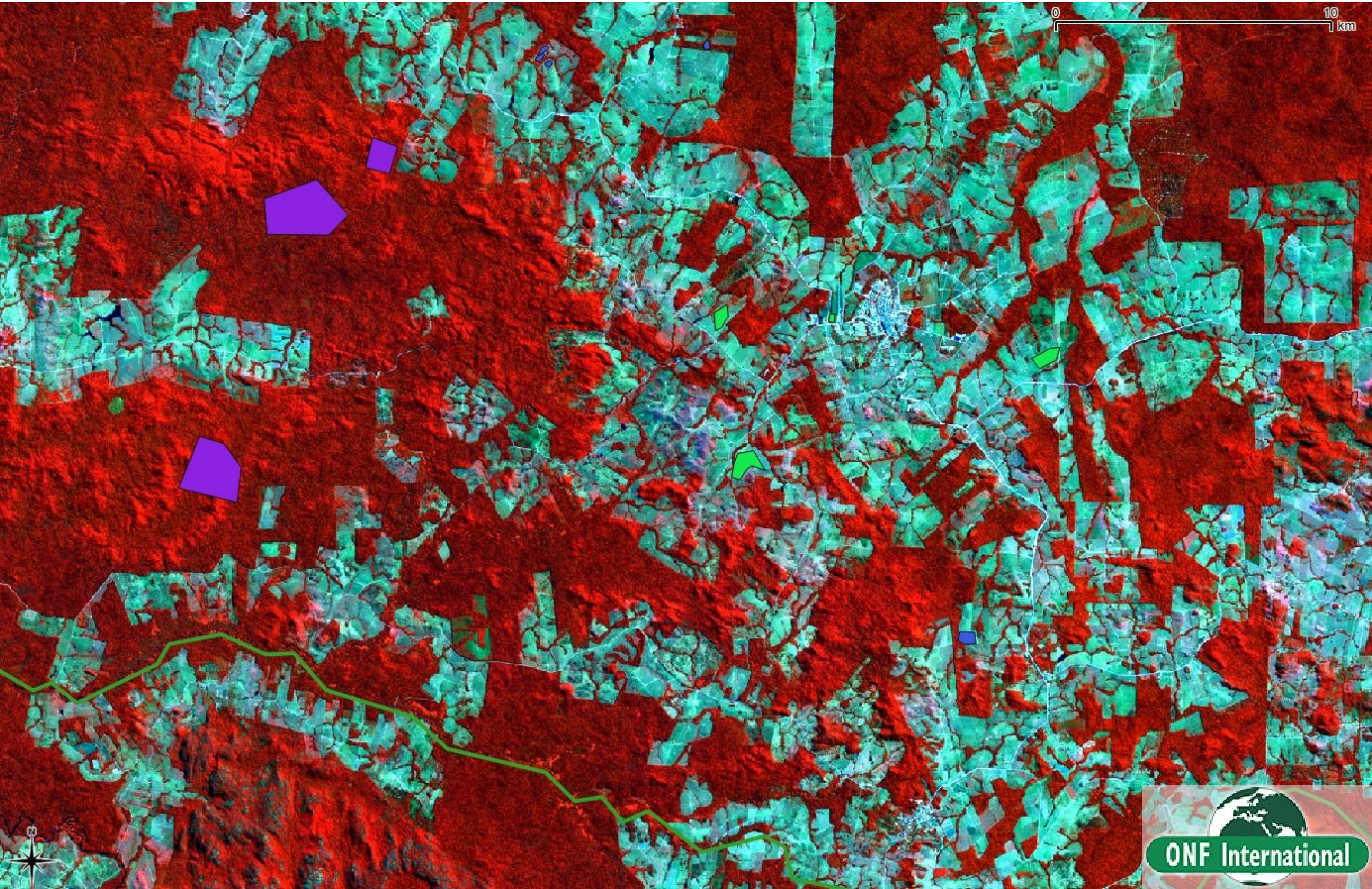
2) The algorithm analyses the spectral properties of each training class

Classfication training** **(training polygons)

LANDSAT TM Image



LANDSAT TM Image



SUPERVISED CLASSIFICATION

*1) The user defines **polygons** on the image*

Several polygons representative of each classes to be determined

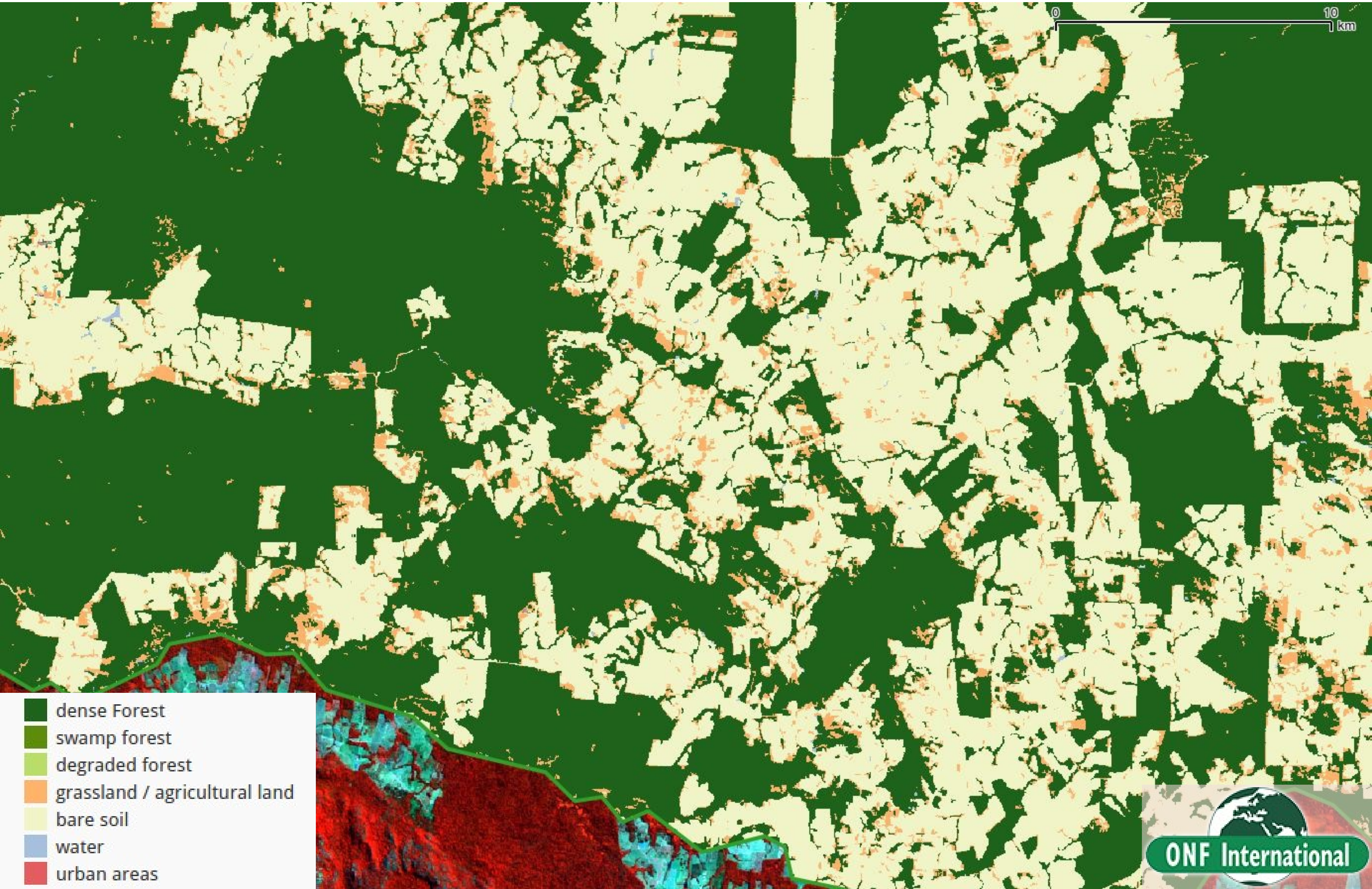
(photo-interprétation, ground survey,....)

2) The algorithm analyses the spectral properties of each training class

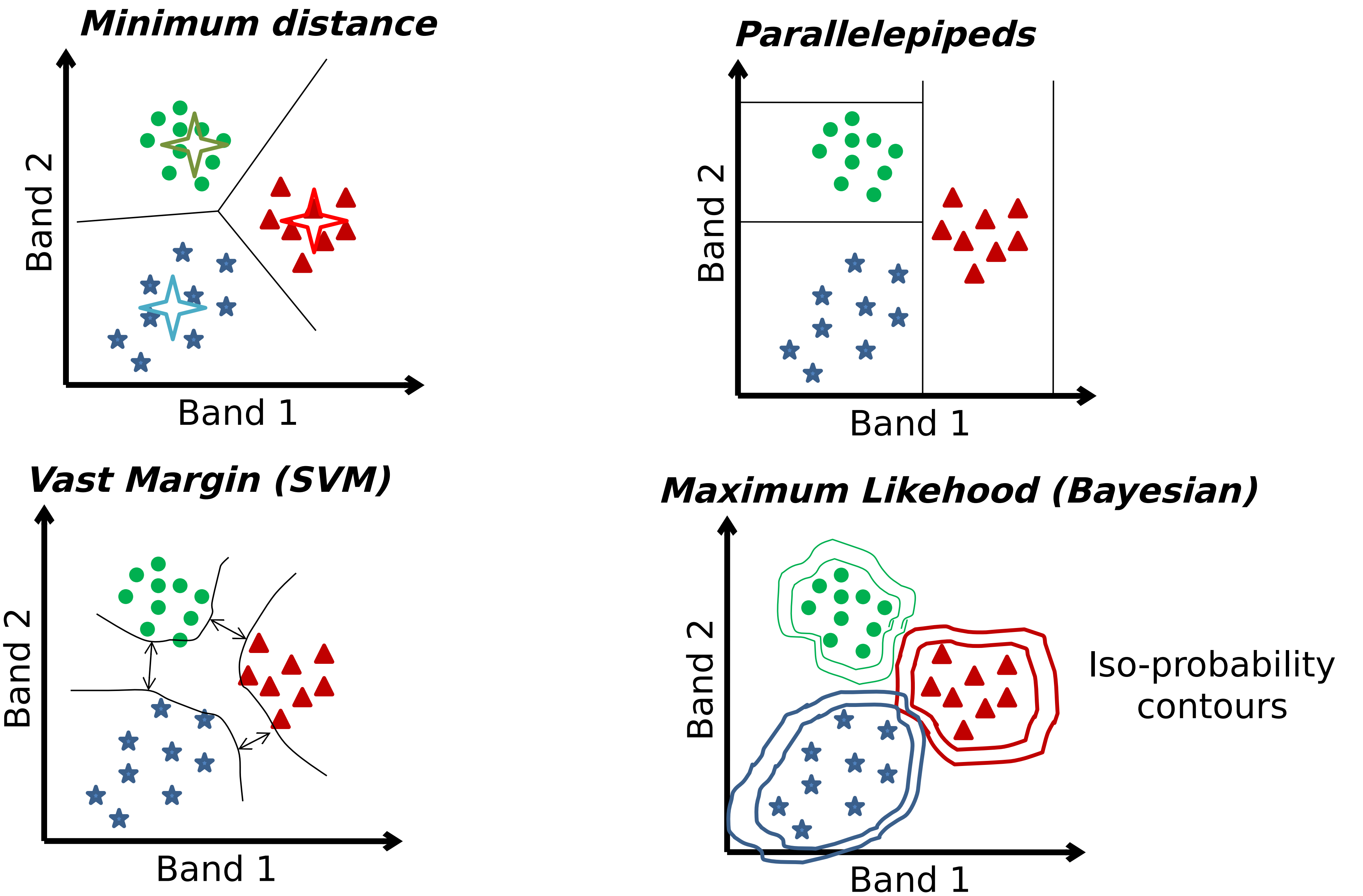
Classfication training (training polygons)

3) The algorithm generalises the training on the whole image

Classification results



Different classification algorithm



Maximum Likehood

Based on Bayes theorem: $$P(C_i/g)=\frac{P(g/C_i)P(C_i)}{P(g)}$$

g: pixel value(N component vector)

N: Bands number of the image

C_1, C_2, …., C_K: class to be determined

$$P(g)=\sum_{i=1}^{k}P(g/C_i)P(C_i)$$ Independent of classes C_i

hence: $g \in C_i$ if $P(g/C_i)P(C_i) \geq P(g/C_j)P(C_j) \quad \forall j=1,2,\ldots,k$

Supposing that each class follows ***a multivariate normal distribution***

$$P(g/C_j)=\frac{1}{(2\pi)^{N/2}\sqrt{|\Sigma_j|}}\exp\left(-\frac{1}{2}(g-\mu)^T\Sigma_j^{-1}(g-\mu)\right)$$ (mean μ, covariance matrix Σ)

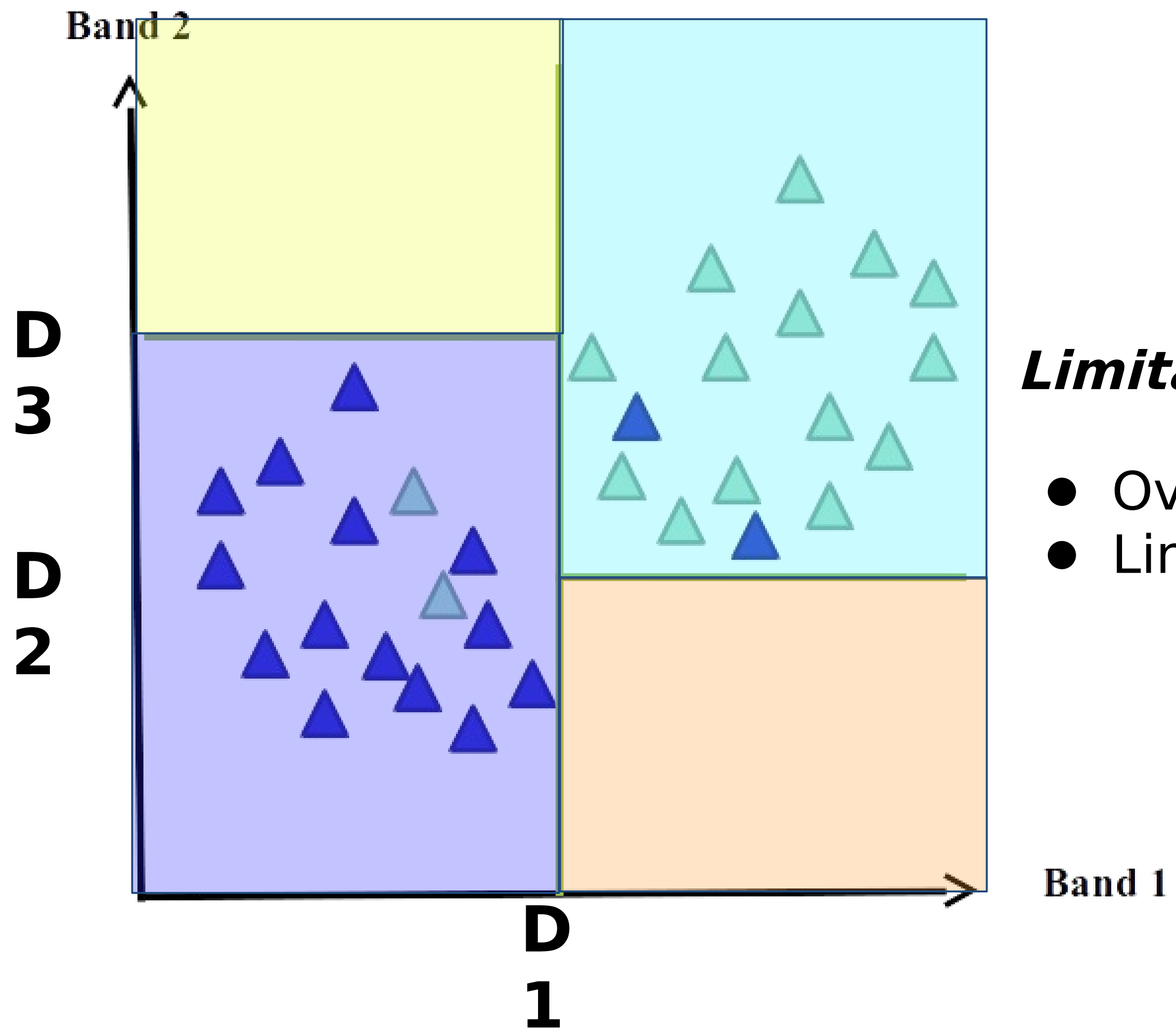
$$d_j(g)=-\log|\Sigma_j|-(g-\mu)^T\Sigma_j^{-1}(g-\mu)$$

$g \in C_i$ si $d_i(g) \geq d_j(g) \quad \forall j=1,2,\ldots,k$

Fast and good results， widely used ***(for N ot too big!)***

Supervised classification

Example of decision Tree



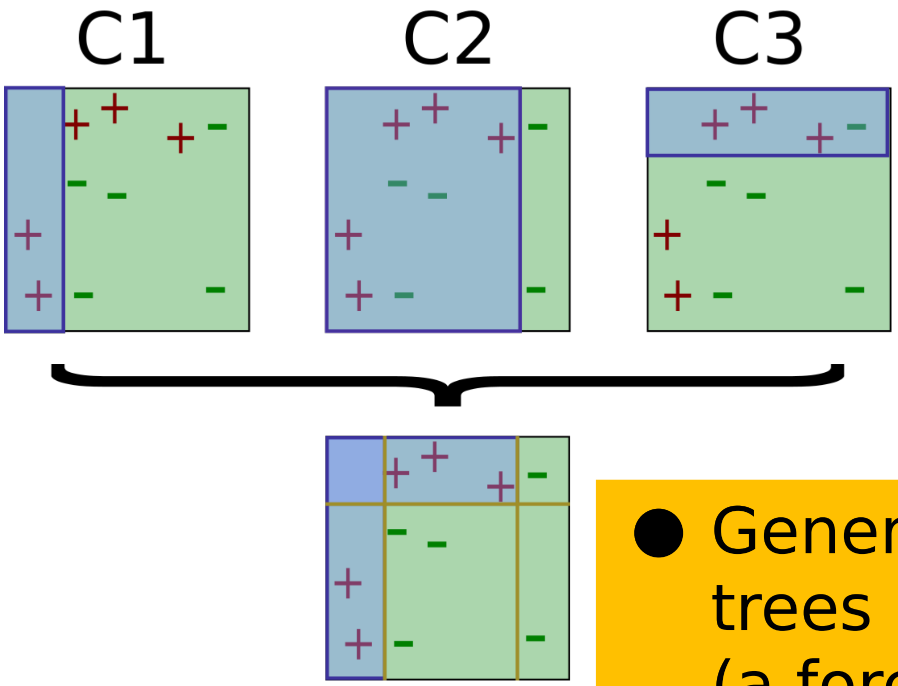
Limitation:

- Overlearing
- Linear border

Supervised classification

Example of Random Forest

- Use classification with decision tree
- Based on boosting principle for prediction improvment



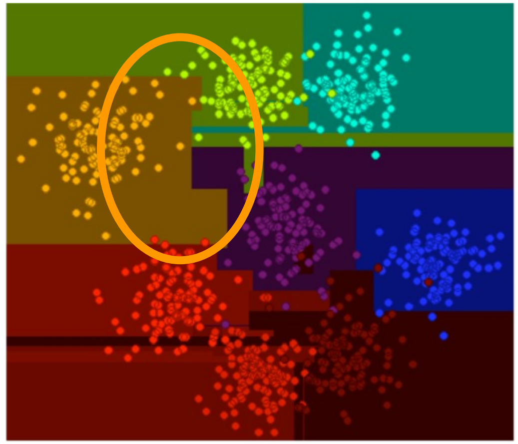
- 3 classifications of moderate performance
- Combination by majority vote
- Results obtained with the best performance

- Generate large number of decision trees
(a forest)
  - Different random training samples
  - Differet random bands
- Combination by majority vote

a quick guide to boosting in ML,
//medium.com/greyatom/a-quick-guide-to-boosting-in-ml-acf7c1585cb5

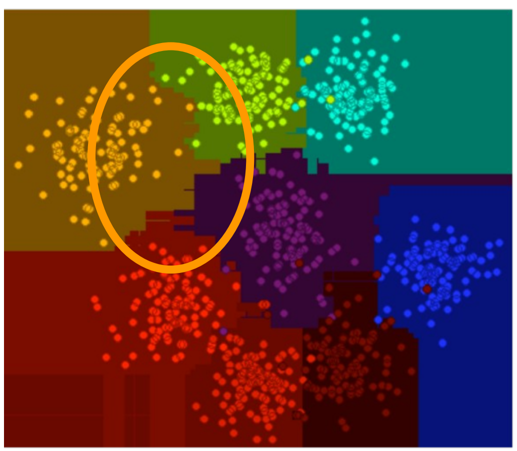
Supervised classification

Example ofRandom Forest

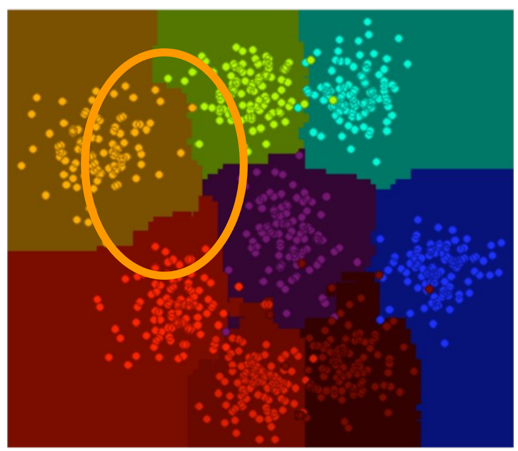


1 Tree

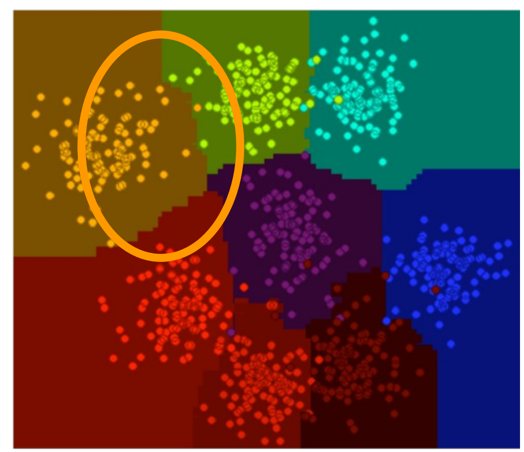
Trees number > 100, improvement the precision of the classes borders



10 Trees



100 Trees



500 Trees

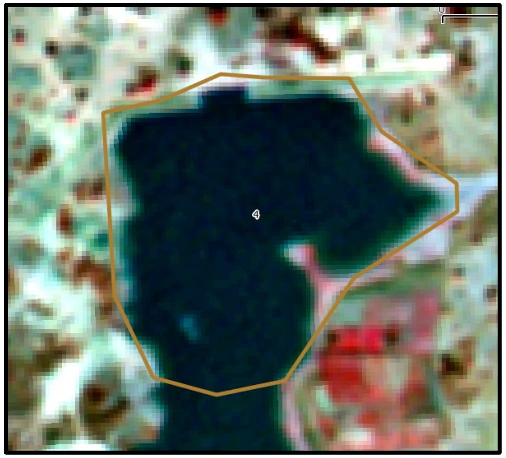
Algorithme RANDOM FOREST

En résumé, pourquoi Random Forest ?

- Nécessite ***un nombre raisonnable de points d'entraînement*** (mais toujours représentatif)
- Possible d’utiliser ***n’importe quels indices*** quel que soit ses statistique (Optical, Radar, Texture…)
- ***Autant d’indices*** que nécessaire : 1-10-100 et plus !!!
- ***Simple d’utilisation !***
- ***Très rapide***

Comment définir ses ROIs ?

1. **Choisir les classes**
   a. *Qu'est ce que je cherche ?*
   b. *Quel taille font mes classes ?*
   c. *A priori que peu discriminer mes données ?*
2. **Créer ROI**
   a. *Doivent être représentatif (dans les données et dans les paysages)*
   b. *Le plus homogène possible (pas trop non plus ;))*



Comment définir ses ROIs ?

1. **Choisir les classes**
   a. *Qu'est ce que je cherche ?*
   b. *Quel taille font mes classes ?*
   c. *A priori que peu discriminer mes données ?*
2. **Créer ROI**
   a. *Doivent être représentatif (dans les données et dans les paysages)*
   b. *Le plus homogène possible (pas trop non plus ;))*
   c. *Faire simple*
      Décomposer en classe radiométrique (si possible)
      Commencer avec les classes les plus simple
      Ensuite ajouter de la complexité avec plus de classes
   d. *Processus itératif: Tester et améliorer les*

Comment améliorer sa classification ?

1. **# de pts d’entrainement > 200pts/classe**
   Ajouter des polygones
2. **Contrôler l’homogénéité des ROI**
   environ 80% du polygone doit correspondre à la classe annoncée
3. **Les stats du classifieur et la qualité globale de la classification doivent avoir les mêmes tendance**
   Si excellentes stats ***≃ ROI non représentatifs→ ajouter des polygones dans des zone d’erreurs***
4. **Si mauvais résultats**
   ajouter des polygones dans des zone d’erreurs
5. **Si toujours mauvais résultats**

Supervised classification

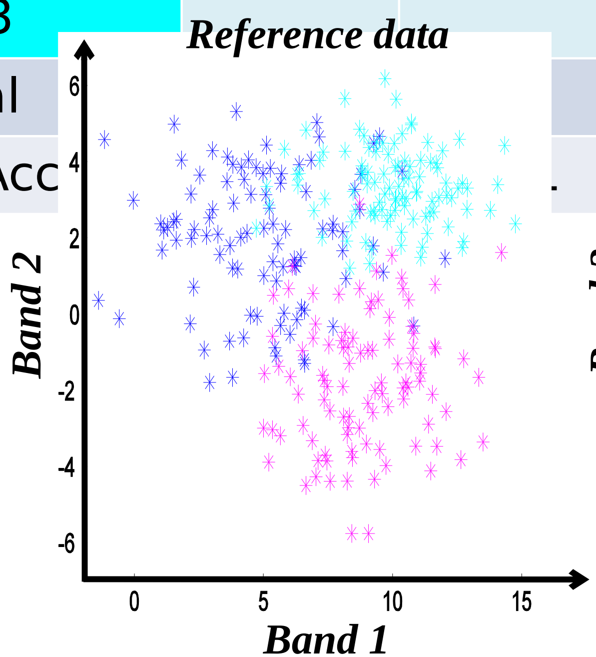
Random Forest

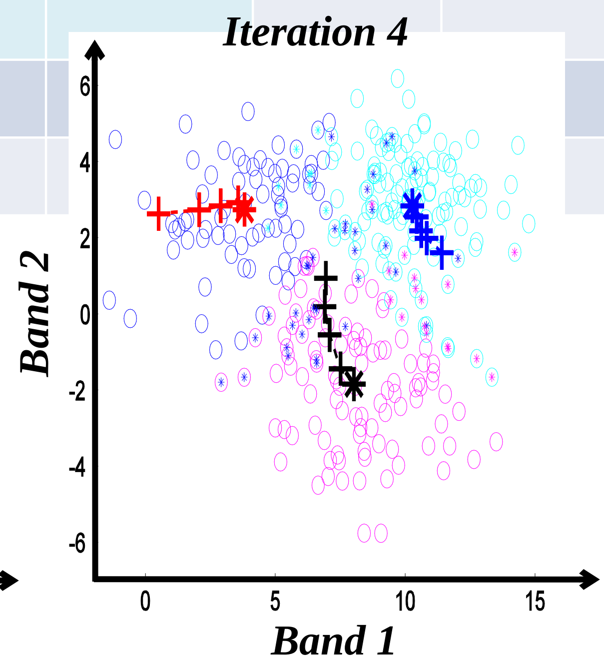
- Needs of reasonable number of training points
- Suited to any features set(optical, radar, texture, ...)
- ***As many indices*** asnécessary : 1-10-100 and more!!!
- ***Simple to use !***
- ***Very fast***

Evaluation of the performance of a classification

CONFUSION MATRICE

		Estimated classes				
		Classe 1	Classe 2	Classe 3	Total	Prod. Acc.
Reference data	Classe 1	65	19	16	100	0.65
	Classe 2	0	84	16	100	0.84
	Classe 3	9	0	91	100	0.91
Total						
User Acc						





Clasification performance

CONFUSION MATRICE

		Estimated classes				
		Classe 1	Classe 2	Classe 3	Total	Prod. Acc.
Reference data	Classe 1	65	19	16	100	0.65
	Classe 2	0	84	16	100	0.84
	Classe 3	9	0	91	100	0.91
Total		74	103	123	300	
User Acc.		0.88	0.81	0.74		

$$\text{Producer Accuracy} = \frac{\text{\# pixels well classified}}{\text{\# pixels reference class}}$$

Omission error = 1 - PA

$$\text{User Accuracy} = \frac{\text{\# pixels well classified}}{\text{\# pixels estimated class}}$$

Comission error = 1 – UA

QUALITE DE LA CLASSIFICATION

MATRICE DE CONFUSION

		Classes estimées				Prod. Acc. Recall
		Classe 1	Classe 2	Classe 3	Total	
Classes réelles	Classe 1	65	19	16	100	0.65
	Classe 2	0	84	16	100	0.84
	Classe 3	9	0	91	100	0.91
Total		74	103	123	300	
User Acc. Precision		0.88	0.81	0.74		

C

Précision totale

$$\text{Producer Accuracy} = \frac{\#\text{ pixels bien classés}}{\#\text{ pixels classe réelle}}$$

Recall

$$\text{Omission error} = \frac{\#\text{ pixels omis}}{\#\text{ pixels classe réelle}}$$

OE = 1 – PA

$$\text{User Accuracy} = \frac{\#\text{ pixels bien classés}}{\#\text{ pixels classe estimée}}$$

Precision

$$\text{Commission error} = \frac{\#\text{ pixels mal classés}}{\#\text{ pixels classe estimée}}$$

CE = 1 – UA

QUALITE DE LA CLASSIFICATION

MATRICE DE CONFUSION

		Classes estimées				Prod. Acc. Recall
		Classe 1	Classe 2	Classe 3	Total	
Classes réelles	Classe 1	65	19	16	100	0.65
	Classe 2	0	84	16	100	0.84
	Classe 3	9	0	91	100	0.91
Total		74	103	123	300	
User Acc. Precision		0.88	0.81	0.74		

C

Précision totale

$$\text{Producer Accuracy} = \frac{\#\ \text{pixels bien classés}}{\#\ \text{pixels classe réelle}}$$

Recall

% qu'un pixel de la classe réelle soit bien classé (PD)

$$\text{User Accuracy} = \frac{\#\ \text{pixels bien classés}}{\#\ \text{pixels classe estimée}}$$

Precision

% qu'un pixel de la classe estimée soit bien classé (1 - FA)

QUALITE DE LA CLASSIFICATION

MATRICE DE CONFUSION

		Classes estimées				Prod. Acc. Recall
		Classe 1	Classe 2	Classe 3	Total	
Classes réelles	Classe 1	65	19	16	100	0.65
	Classe 2	0	84	16	100	0.84
	Classe 3	9	0	91	100	0.91
Total		74	103	123	300	
User Acc. Precision		0.88	0.81	0.74		

$$\textbf{\textit{Precision}} = \frac{\text{Prec}_1 + \text{Prec}_2 + \text{Prec}_3}{3} = \frac{0.88 + 0.81 + 0.74}{3} = 0.81$$

$$\textbf{\textit{Recall}} = \frac{\text{Recall}_1 + \text{Recall}_2 + \text{Recall}_3}{3} = \frac{0.65 + 0.84 + 0.91}{3} = 0.80$$

$$\textbf{\textit{F-Score}} = 2\,\frac{\text{Precision . Recal}}{\text{Precision + Recal}} = 2\,\frac{0.81 \,.\, 0.80}{0.81 + 0.80} = 0.80$$

QUALITE DE LA CLASSIFICATION

MATRICE DE CONFUSION

		Classes estimées				
		Classe 1	Classe 2	Classe 3	Total	Prod. Acc.
Classes réelles	Classe 1	65	19	16	100	0.65
	Classe 2	0	84	16	100	0.84
	Classe 3	9	0	91	100	0.91
Total		74	103	123	300	
User Acc.		0.88	0.81	0.74		

n pixel de l'image classe 1) = 100 / 300 = 33 %

P(un pixel de la classif classe 1) = 74 / 300 = 25 %

xel de la classe 1 soit bien classé par chance) = 0.33 * 0.25 = 8 %

bon accord par chance

$$P_e = \frac{100\,.\,74}{300^2} + \frac{100\,.\,103}{300^2} + \frac{100\,.\,123}{300^2} = 33\%$$

bon accord observé

$$P_0 = \frac{65 + 84 + 91}{300} = 80\%$$

$$Kappa = \frac{P_0 - P_e}{1 - P_e} = 73\%$$

Comparaison entre résultats obtenus et
d'une classification totalement aléatoire

QUALITE DE LA CLASSIFICATION

MATRICE DE CONFUSION

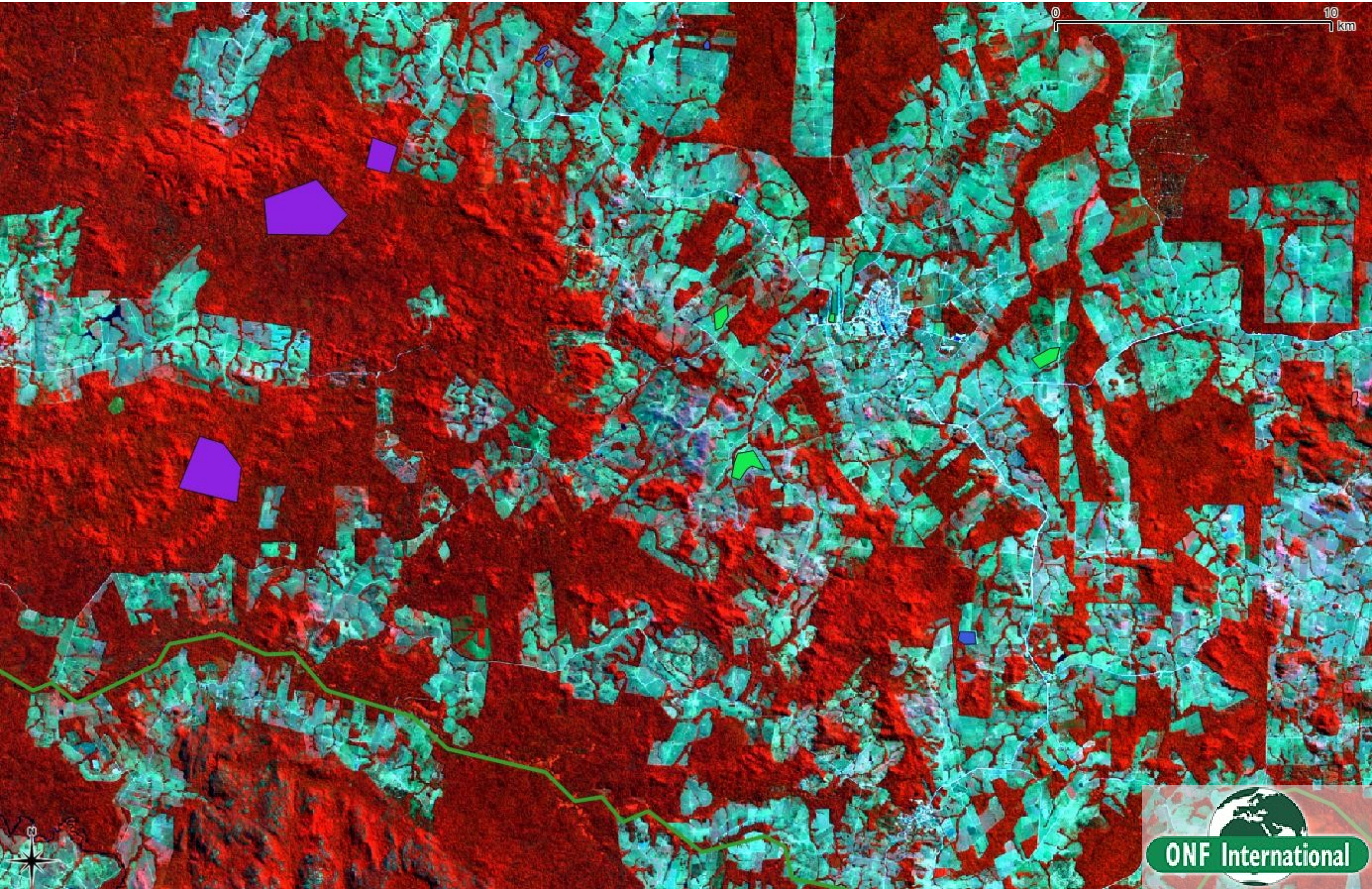
		Classes estimées				
		Classe 1	Classe 2	Classe 3	Total	Prod. Acc.
Classes réelles	Classe 1	65	19	16	100	0.65
	Classe 2	0	84	16	100	0.84
	Classe 3	9	0	91	100	0.91
Total		74	103	123	300	
User Acc.		0.88	0.81	0.74		

Vrais Positifs1 = 65
Faux Négatifs 1 = 19 + 16 = 35
Faux Positifs 1 = 9
Vrais Negatifs 1 = 84 + 16 + 91 = 191

Vrais Positifs 2 = 84
Faux Négatifs 2 = 16
Faux Positifs 2 = 19
Vrais Négatifs 2 = 65 + 9 + 16 + 91 = 181

Vrais Positifs 3 = 91
Faux Négatifs 3 = 16
Faux Positifs 3 = 16 + 16 = 32
Vrais Négatifs 3 = 65 + 19 + 84 = 168

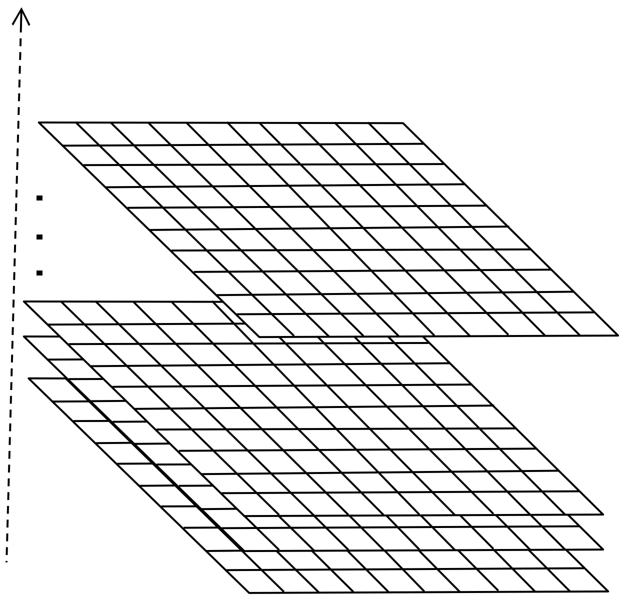
Image LANDSAT TM



REDUCTION DE DIMENSION SPECTRALE

Dimmension Spectrale:

N Canaux



Exemple: image hyperspectrale 50 cana

1.99 – 2.48 μm

Résolution spatiale: 10 nm

Perception œil: 3 parmi N canaux

Bandes: 12 – 22 - 31

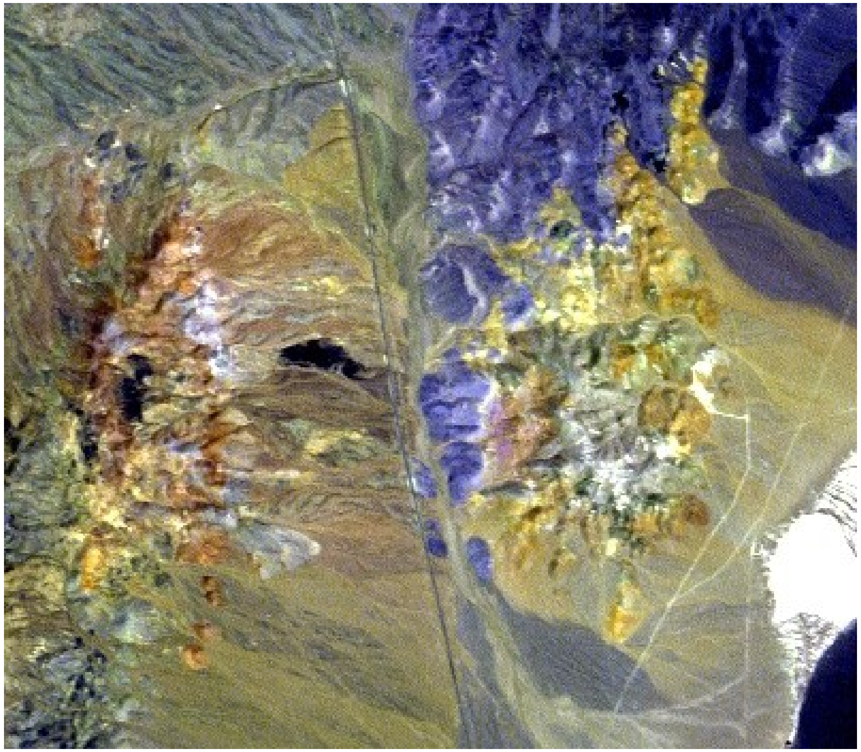
Image aéroportée AVIRIS, Cuprite,

REDUCTION DE DIMENSION SPECTRALE

Exemple: image hyperspectrale 50 canaux
1.99 – 2.48 μm
Résolution spatiale: 10 nm

Perception œil: 3 parmi N canaux

Bandes: 1 – 11 - 50



Bandes: 12 – 22 - 31

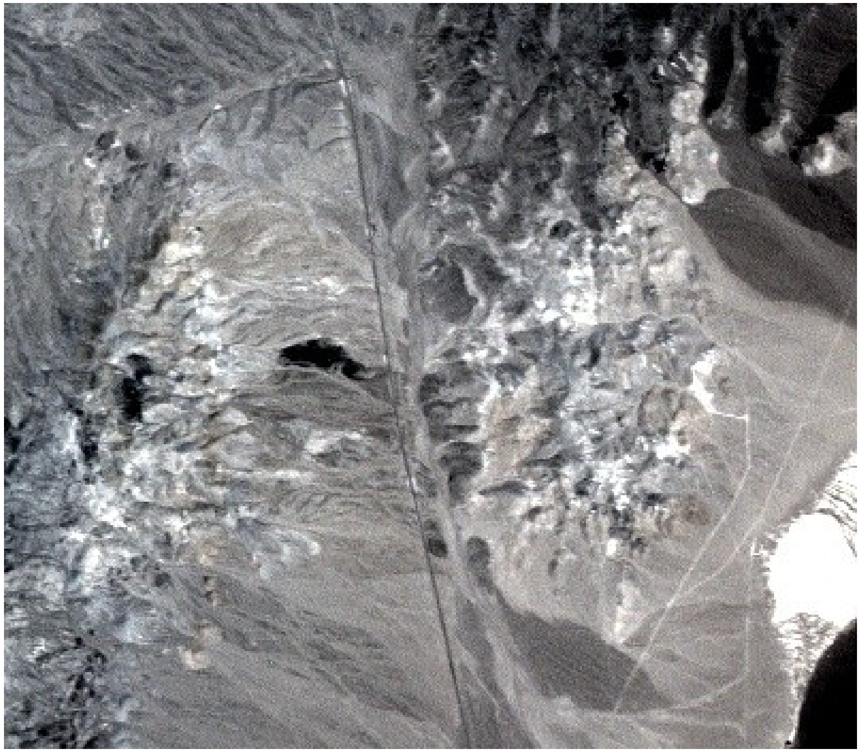
Image aéroportée AVIRIS, Cuprite,

REDUCTION DE DIMENSION SPECTRALE

Exemple: image hyperspectrale 50 canaux
1.99 – 2.48 μm
Résolution spatiale: 10 nm

Perception œil: 3 parmi N canaux

Bandes: 1 –2 - 3

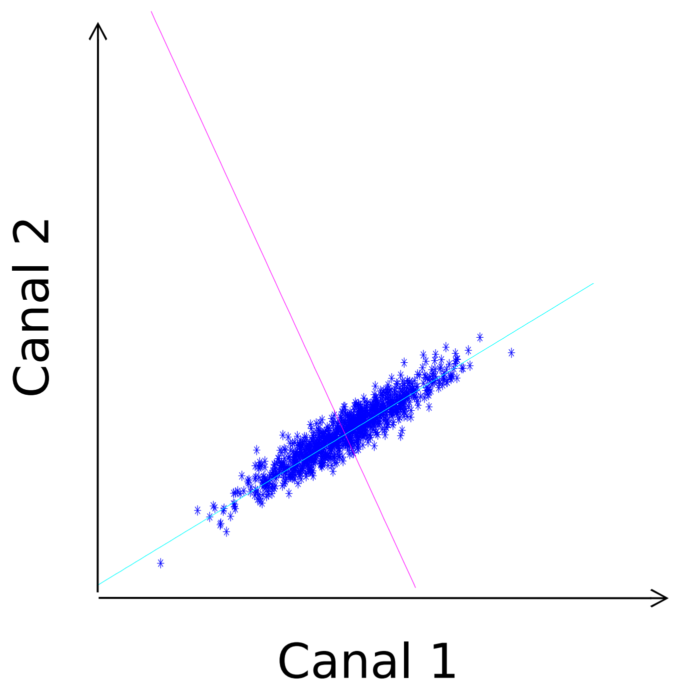


Bandes: 12 – 22 - 31

Image aéroportée AVIRIS, Cuprite,

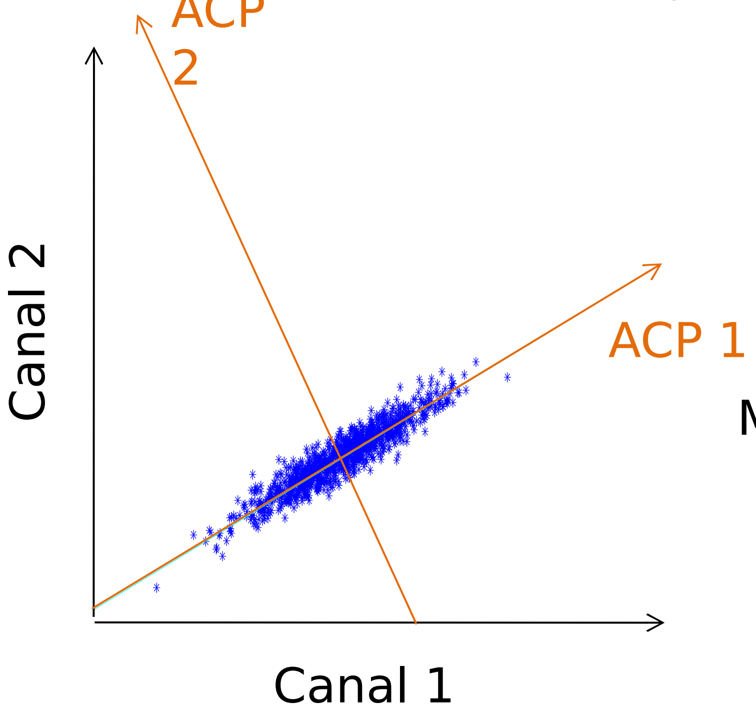
Analyse en Composantes Principales (ACP)

uire la dimension spectrale de l'image pour ne garder que
s composantes contenant le ***plus d'information*** = ***plus grande variance***



Analyse en Composantes Principales (ACP)

uire la dimension spectrale de l’image pour ne garder que
s composantes contenant le ***plus d’information*** = ***plus grande variance***



Matrice de covariance (Canal1, Canal2)

$$C\begin{pmatrix} 99.8 & 49.7 \\ 49.7 & 28.7 \end{pmatrix}$$

Matrice de covariance (ACP1, ACP2)

$$C\begin{pmatrix} 125.4 & 0 \\ 0 & 3.1 \end{pmatrix}$$

REDUCTION DE DIMENSION SPECTRALE

Analyse en Composantes Principales (ACP)

uire la dimension spectrale de l'image pour ne garder que
s composantes contenant le ***plus d'information*** = ***plus grande variance***

Principe: diagonaliser matrice de covariance entre canaux

Exemple: image à ***4 canaux***

Matrice de covariance: $C = \begin{pmatrix} C_{11} & C_{12} & C_{13} & C_{14} \\ C_{21} & C_{22} & C_{23} & C_{24} \\ C_{31} & C_{32} & C_{33} & C_{34} \\ C_{41} & C_{42} & C_{43} & C_{44} \end{pmatrix}$

(matrice symétrique)

$$C_{ij} = \mathrm{cov}(x_i, x_j) = \frac{1}{nb\_pixels} \sum_{k=1}^{nb\_pixels} \left(x_{i,k} - \bar{x}_i\right)\left(x_{j,k} - \bar{x}_j\right)$$

$$C_{ii} = \mathrm{var}(x_i) = \frac{1}{nb\_pixels} \sum_{k=1}^{nb\_pixels} \left(x_{i,k} - \bar{x}_i\right)^2$$

Analyse en Composantes Principales (ACP)

$$P^{-1}\,C\,P = \begin{pmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \lambda_3 & 0 \\ 0 & 0 & 0 & \lambda_4 \end{pmatrix}$$

λ_i : ***Valeurs propres*** de C

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \lambda_4$$

$$\begin{array}{c} \begin{array}{cccc} \vec{V}_1 & \vec{V}_2 & \vec{V}_3 & \vec{V}_4 \end{array} \\ P = \begin{pmatrix} P_{11} & P_{12} & P_{13} & P_{14} \\ P_{21} & P_{22} & P_{23} & P_{24} \\ P_{31} & P_{32} & P_{33} & P_{34} \\ P_{41} & P_{42} & P_{43} & P_{44} \end{pmatrix} \end{array}$$

, ***vecteurs propres*** de C

P : Matrice de changement de base

Matrice de vecteurs propres

$$\left(\vec{e}_1, \vec{e}_2, \vec{e}_3, \vec{e}_4\right) \rightarrow \left(\vec{e}_1^{\,\prime}, \vec{e}_2^{\,\prime}, \vec{e}_3^{\,\prime}, \vec{e}_4^{\,\prime}\right)$$

$$\vec{e}_2^{\,\prime} = P_{12}\vec{e}_1 + P_{22}\vec{e}_2 + P_{32}\vec{e}_3 + P_{42}\vec{e}_4$$

Analyse en Composantes Principales (ACP)

$$P^{-1} C P = \begin{pmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \lambda_3 & 0 \\ 0 & 0 & 0 & \lambda_4 \end{pmatrix}$$

λ_i : ***Valeurs propres*** de C

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \lambda_4$$

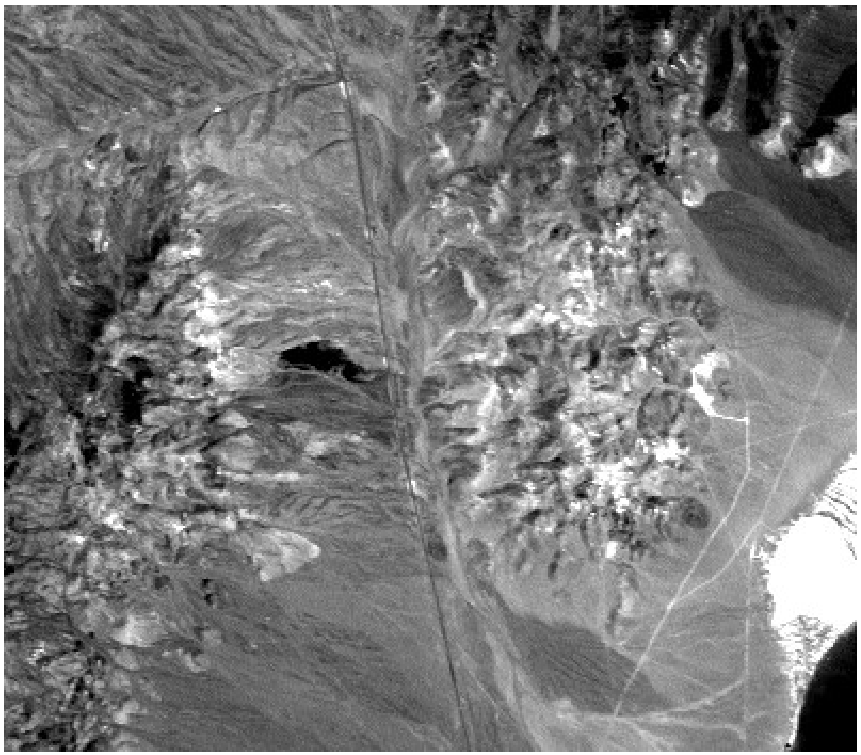
$$\frac{\lambda_i}{\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4}$$ Donne la proportion (%) de la variance portée par l'a

REDUCTION DE DIMENSION SPECTRALE

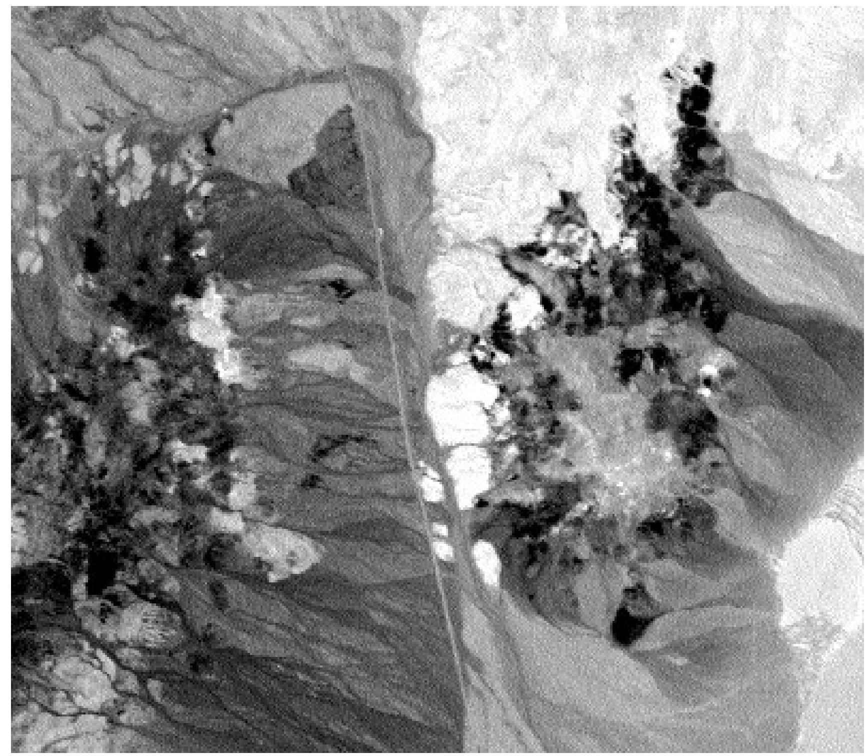
Analyse en Composantes Principales (ACP)

duire la dimension spectrale de l’image pour ne garder que les composantes ontenant le plus d’information

ACP – Axe 1 (90%)



ACP – Axe 2 (6%)



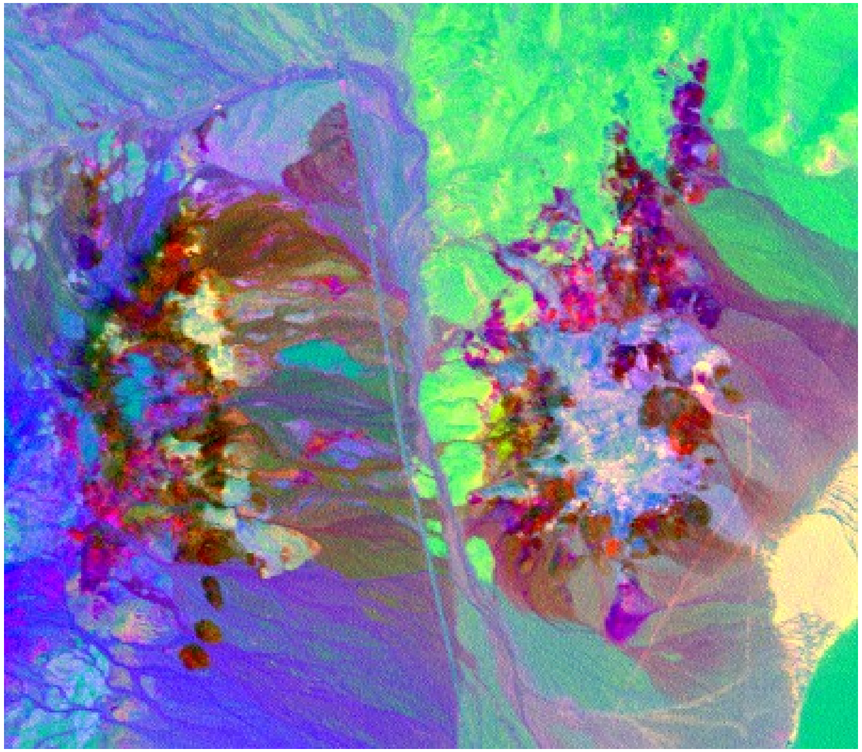
(λ_1=58665, λ_1=4139, $\Sigma\,\lambda_i$ = 6518

REDUCTION DE DIMENSION SPECTRALE

Exemple: image hyperspectrale 50 canaux
1.99 – 2.48 μm
Résolution spatiale: 10 nm
Perception œil: 3 parmi N canaux

Bandes 12 – 22 - 31

ACP: Axe 1 – Axe 2 – Axe 3 (98%)

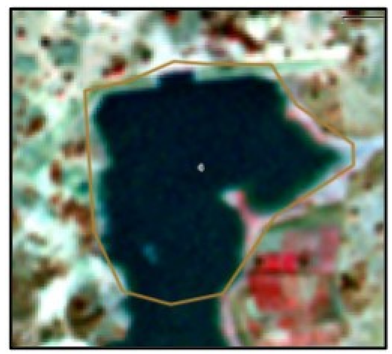


ge aéroportée AVIRIS, Cuprite, USA

Practical recommendations

w to defined polygons (training – validation)?

- **representative**
- **homogeneous**

- **Be simple**

  Most simples classes at the beginninG

  Later add more sophisticated classes

- **Iterative process**

  test and improve the polygons

Practical recommendations

aining samples > 200 / class

several polygons recommended

tistics of classification and visual apect must match

if good stats but poor visual aspect: polygons are ot representative

 add polygons to better match classes

if bad stats

add polygons in error zones

if still bad results

merge confusing classes