



Détection d'anomalies sur des images hyperspectrales

Véronique Achard



retour sur innovation

Détection d'anomalies

1. Objectifs

2 . Méthodes de détection d'anomalies

3. Prétraitement / Réduction de dimension

4. Critères d'évaluation de performance

5. Quelles méthodes/ quels prétraitements pour quels types d'images ?

6. Références

Détection d'anomalies

1. Objectifs

2 . Méthodes de détection d'anomalies

3. Prétraitement / Réduction de dimension

4. Critères d'évaluation de performance

5. Quelles méthodes/ quels prétraitements pour quels types d'images ?

6. Références

Détection d'anomalies : Détection de pixels atypiques

- **Détection d'évènements rares (« cibles ») dans une image et qui se distinguent par leur signature spectrale du reste de l'image (« fond »).**
- **Peut être vue comme un problème de classification à deux classes :**
 - **Classe fond (largement représentée, qui peut être hétérogène)**
 - **Classe anomalies (classe cibles)**

Objectifs

Pour le thématicien, qu'est-ce qu'une anomalie ?

- **Végétation/agriculture :**
 - stress hydrique dans un champ
 - arbres infectés dans une forêt
- **Géologie**
 - minéraux rares en géologie
- **Médecine**
 - Recherche de tumeurs
- **Militaire**
 - véhicules, objets manufacturés

Détection d'anomalies

1. Objectifs

2 . Méthodes de détection d'anomalies

- **Principe**
- **RX (Reed-Xiaoli)**
- **GMM : basée sur la modélisation par mélange de gaussiennes**
- **Méthodes non linéaires**

3. Prétraitement / Réduction de dimension

4. Critères d'évaluation de performance

5. Quelles méthodes/ quels prétraitements pour quels types d'images ?

6. Références

Méthodes - Principes

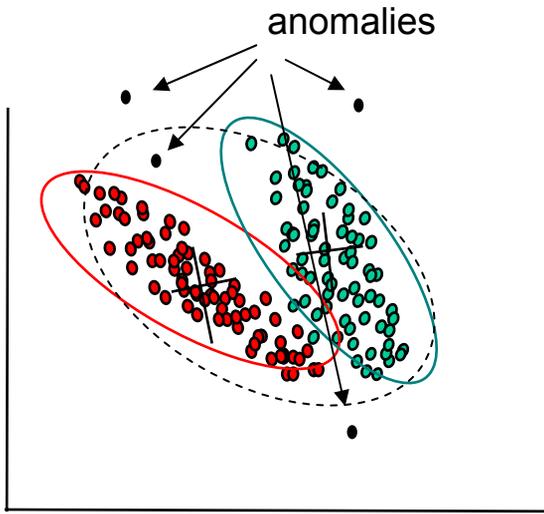
- Approche statistique
 - Image : N pixels, L bandes spectrales
- Test à deux hypothèses :
 - H0 : x est un pixel fond
 - H1 : x est un pixel cible
- Règle de décision de Nieman Pearson basée sur le rapport de vraisemblance des deux densités de probabilité (GLRT) :

$$\Lambda = \frac{f_{\mathbf{x}/H_1}(\mathbf{x})}{f_{\mathbf{x}/H_0}(\mathbf{x})} \underset{<_{H_0}}{\overset{>_{H_1}}{>}} \eta$$

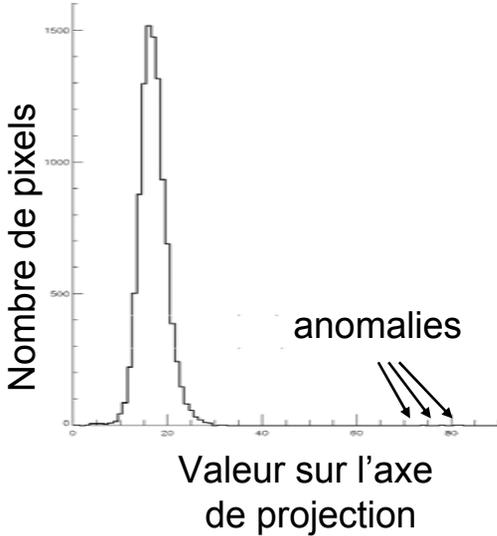
Méthodes - Principes

Modélisation (globale/locale) du fond par des distributions, gaussiennes ou non et recherche des pixels en marge de ces distributions

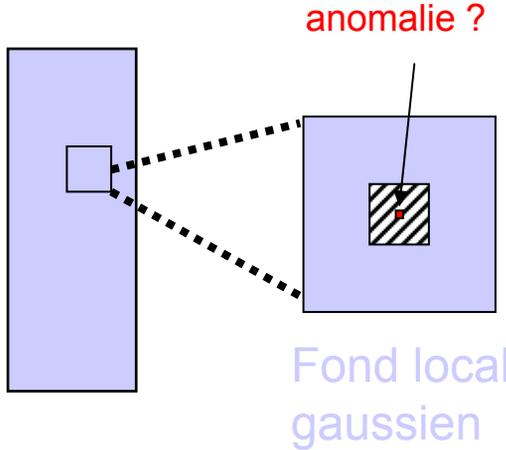
Modélisation du fond par gaussiennes multivariées



Poursuite de projection : Recherche d'axes / la distribution des pixels a des queues lourdes par rapport à une distribution gaussienne



Détecteur RX



L'algorithme de détection RX

Basé sur l'analyse de deux hypothèses :

$H_0 : \mathbf{x} = \mathbf{n}$ (absence de cible),

$H_1 : \mathbf{x} = a\mathbf{s} + \mathbf{n}$ (présence de cible)

Où \mathbf{x} est un pixel de l'image (après centrage des données),

\mathbf{s} , signature spectrale de la cible, $a > 0$: proportion de \mathbf{s} dans le pixel,

et \mathbf{n} un vecteur représentant le « bruit » (bruit instrumental + variabilité naturelle du fond).

Distributions gaussiennes :

Sous H_0 , les données suivent une loi $N(0, \mathbf{C}_b)$

Sous H_1 , les données suivent une loi $N(\mathbf{s}, \mathbf{C}_b)$

Où \mathbf{C}_b est la matrice de covariance des pixels, de taille $L \times L$.

L'algorithme de détection RX

On examine chaque pixel x : Anomalie ?

$$\text{Algorithme RX : } \text{RX}(x) = (x - \hat{\mu}_b)^T \hat{C}_b^{-1} (x - \hat{\mu}_b) \underset{H_0}{\overset{H_1}{>}} \eta$$

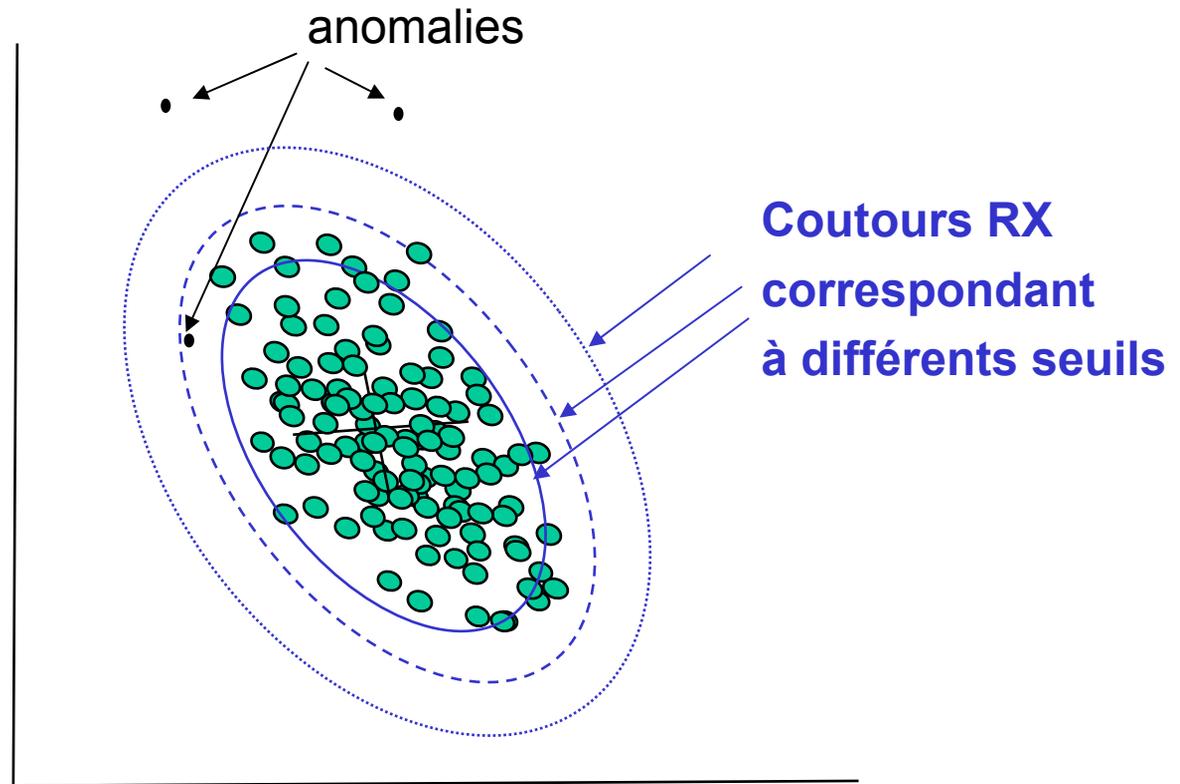
(approximation du GLRT)

\hat{C}_b est la matrice de covariance de la loi normale (de taille $L \times L$)
 $\hat{\mu}_b$ est la matrice des moyennes (pour le centrage des données).

$\text{RX}(x) > \text{seuil} \rightarrow x$ est différent du fond : anomalie

Carte des anomalies : obtenue par seuillage du critère RX

Méthode RX

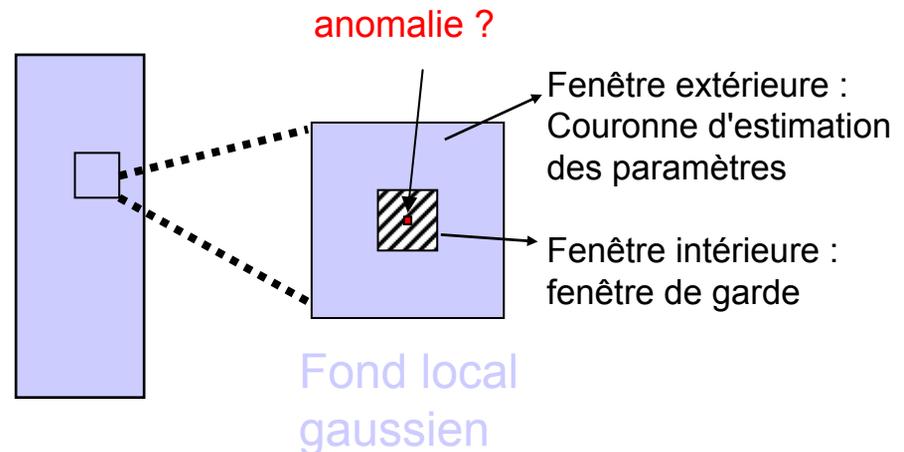


L'algorithme de détection RX

\hat{C}_b et $\hat{\mu}_b$ estimées sur l'image :

GRX : Estimations globales : moyenne et matrice de covariance estimées sur l'ensemble de l'image

LRX : Estimations locales :
Sur une couronne entourant
Le pixel examiné



Méthode RX – GRX et LRX

- **Méthode GRX :**

→ Convient à une image ayant un fond ~ homogène car représentation par une distribution gaussienne.

- **Méthode LRX :**

→ Convient pour des images plus hétérogènes, mais peu adaptée à la détection en milieu urbain

Dans les deux cas, fond modélisé par une distribution normale (estimée globalement ou localement)

Méthode CRX (par classes)

CRX: CB-AD Cluster-based Anomaly detector

- 1ere étape : classification K-means
 - - Choisir k centres de classes distincts (centroïdes) y_1, y_2, \dots, y_k
 - - Construire k classes C_1, C_2, \dots, C_k en affectant chaque pixel de l'image à une des k classes :

$$C_i = \left\{ x \mid \|x - y_i\| < \|x - y_j\| \forall j \neq i \right\}$$

- 2ème étape : détection

$$RX(x) = \min_j \left[(x - \hat{\mu}_j)^T \hat{C}_j^{-1} (x - \hat{\mu}_j) \right] \underset{<_{H_0}}{\overset{>_{H_1}}{}} \eta$$

Méthode CRX (par classes)

Choix du nombre de classes ?

Si le nombre de classes est trop grand : risque que les anomalies forment une classe, donc non détectées \Rightarrow **Nombre minimal de pixels par classes.**

- Contrairement au RX local, tous types de cible peut être détectés (toutes formes et tailles).

Méthodes GMM

GMM : CB-AD Cluster-based Anomaly detector

Modélisation du fond par mélange de gaussiennes multivariées

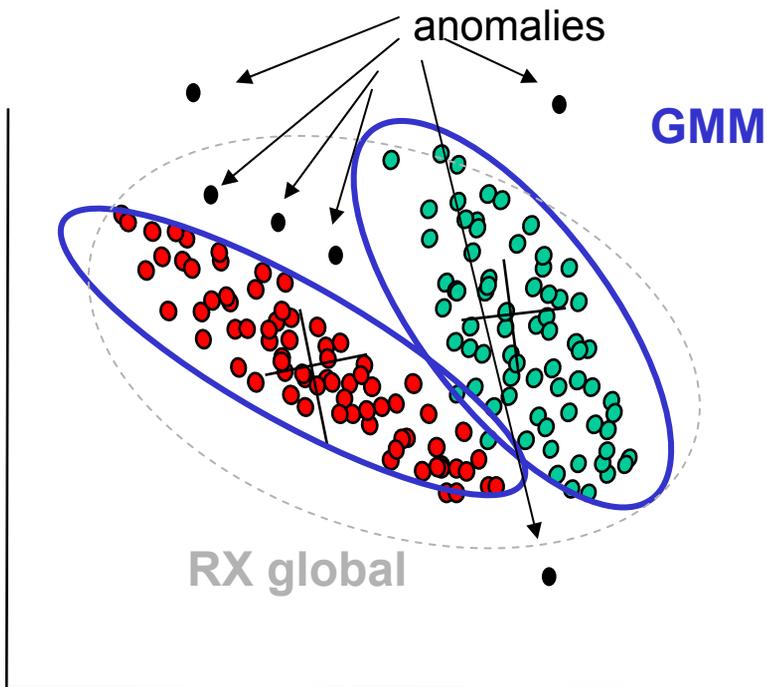
$$f_{X|H_0}(\underline{x}) = \sum_{i=1}^{N_c} \alpha_i \cdot g_N(\underline{x}; \underline{\mu}_i, \underline{C}_i)$$

$$\alpha_i \geq 0, \quad \sum_{i=1}^{N_c} \alpha_i = 1$$

Estimation des paramètres du modèle par EM (Expectation-Maximisation), SEM (Stochastique EM)

Estimation de N_c ?

→ Il existe différentes méthodes pour estimer N_c



1^{ère} approche GMM : CB-AD Cluster-based Anomaly detector

Chaque mode définit une classe et les pixels sont testés par rapport à leur classe j d'appartenance (on ne tient pas compte des α_j) :

$$AD(\underline{x}) = -\log\left(g_N(\underline{x}, \hat{\underline{\mu}}_j, \underline{\underline{C}}_j)\right) \begin{matrix} >_{H1} \\ <_{H0} \end{matrix} \eta$$

- Si N_c est sous-estimé : les classes sont très larges avec de fortes variances \Rightarrow diminution des performances de la détection
- N_c est surestimé : Les pixels cibles peuvent constituer une classe \Rightarrow Ils ne seront pas détectés !

2^{ème} approche GMM : «globale »

On tient compte de la probabilité de chaque mode :

$$AD(\underline{x}) = -\log \left(\sum_{j=1}^{N_c} \hat{\alpha}_j \cdot g_N(\underline{x}, \hat{\underline{\mu}}_j, \underline{C}_j) \right) \begin{matrix} >_{H1} \\ <_{H0} \end{matrix} \eta$$

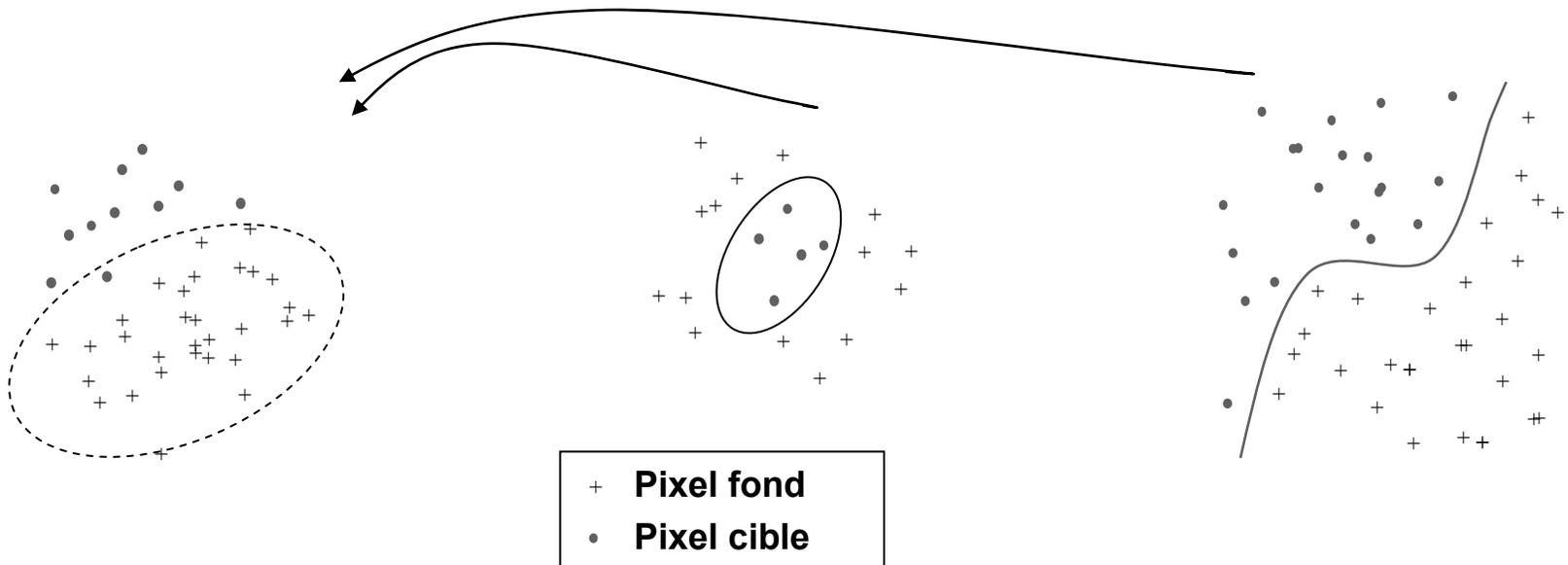
- **Même si les pixels forme une classe à part entière, elle sera peu représentée ($\alpha_j \ll 1$) et donc le mode sera rejeté dans les queues de la distribution \Rightarrow les cibles pourront donc être détectées**
- **Approche moins sensible au choix de N_c**

Détecteurs non linéaires

⇒ Permettre des séparateurs de classes « anomalies » / « fond » plus complexes

Méthodes à noyaux très utilisées pour classification SVM

$\varphi(x)$ (transformation non linéaire)



Méthode Kernel-RX

$\Phi : \mathbb{R}^N \rightarrow \mathcal{F}$, \mathcal{F} peut être de dimension infinie

Deux hypothèses :

- $H_{0\Phi} : \Phi(\mathbf{x}) = \mathbf{n}_\Phi$ (absence de cible)
- $H_{1\Phi} : \Phi(\mathbf{x}) = b \Phi(\mathbf{s}) + \mathbf{n}_\Phi$ (présence de cible)

Où \mathbf{x} est un pixel de l'image, \mathbf{s} la signature spectrale de la cible et \mathbf{n}_Φ un vecteur représentant le bruit.

Algorithme Kernel-RX :

$$RX(\Phi(\mathbf{r})) = (\Phi(\mathbf{r}) - \hat{\boldsymbol{\mu}}_{b\Phi})^T \hat{\mathbf{C}}_{b\Phi}^{-1} (\Phi(\mathbf{r}) - \hat{\boldsymbol{\mu}}_{b\Phi})$$

où $\hat{\boldsymbol{\mu}}_{b\Phi}$ et $\hat{\mathbf{C}}_{b\Phi}$ sont respectivement la moyenne et la matrice de covariance estimées à partir des pixels dans l'espace \mathcal{F}

Méthode Kernel-RX

Astuce du noyau :

Étant donnée une fonction K continue, symétrique, semi-définie positive, elle peut s'exprimer comme un **produit scalaire** dans un espace de grande dimension (théorème de Mercer) :

$K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j)$ produit scalaire dans \mathcal{F} ou noyau (Kernel)

Astuce du noyau : exprimer toutes les équations en fonction de K , et plus de Φ

Méthode Kernel-RX

Expression finale de l'algorithme Kernel-RX ne faisant plus intervenir la fonction Φ mais uniquement les fonctions noyaux, k :

$$KRX(\mathbf{r}) = (\mathbf{K}_r - \mathbf{K}_\mu)^T \hat{\mathbf{K}}_b^{-1} (\mathbf{K}_r - \mathbf{K}_\mu)$$

- $\hat{\mathbf{K}}_b$ est la matrice de Gram centrée, de taille $\mathbf{N} \times \mathbf{N}$,
- estimée à partir des pixels (globalement ou localement)

$$\mathbf{K}_r = \mathbf{k}(\mathbf{X}_b, \mathbf{r})^T - \frac{1}{M} \sum_{i=1}^M k(\mathbf{x}(i), \mathbf{r})$$

$$\mathbf{K}_\mu = \frac{1}{M} \sum_{i=1}^M \mathbf{K}(\mathbf{x}(i), \mathbf{X}_b) - \frac{1}{M^2} \sum_{i=1}^M \sum_{j=1}^M k(\mathbf{x}(i), \mathbf{x}(j))$$

Exemples de fonctions noyaux

- Linéaire : $K(x_i, x_j) = x_i^T \cdot x_j$
- Polynomial : $K(x_i, x_j) = (x_i^T \cdot x_j + 1)^d$
- Gaussien :
$$K(x, y) = \text{Exp}\left(\frac{\|x - y\|^2}{2\sigma^2}\right)$$
- Etc
- Et toute combinaison linéaire de noyaux est un noyau

Quel noyau choisir, avec quels paramètres ?

Détection d'anomalies

1. Objectifs

2. Méthodes de détection d'anomalies

3. Prétraitement

- **Objectifs**
- **Spectral binning**
- **Projections : ACP, KCP**
- **Poursuite de projection**
- **OSP**
- **« Ombre »**

4. Critères d'évaluation de performance

5. Quelles méthodes/ quels prétraitements pour quels types d'images ?

6. Références

Pré- traitements

- Méthode de détection d'anomalies
 1. Pré-traitements : préparer les données pour favoriser la détection
 2. Algorithme de détection
 3. Eventuellement, post-traitements

Le choix du pré-traitement dépend de l'algorithme de détection, du type d'image, d'instrument, etc ...

Pré-traitements - Objectifs

- Réduire la dimension spectrale
- Décorrélérer les bandes spectrales
- Sélectionner des axes spectraux mettant en valeur les anomalies
- Supprimer les composantes « fond »
- Atténuer l'effet des ombres (cibles à l'ombre)
- Réduire le bruit
- Simplifier et améliorer les performances des algorithmes de détection

Prétraitements - Objectifs

Réduction de dimension

→ Pallier le phénomène de Hughes :

Difficultés d'estimer des paramètres dans des espaces pratiquement « vides » :

$f_N = (\text{volume de l'hypersphère} / \text{volume de l'hypercube})$
quand $N \rightarrow \infty$, $f_N \rightarrow 0$

⇒ La qualité de l'estimation de paramètres statistiques, pour un nombre d'échantillons fixés, diminue quand le nombre de bandes spectrales augmente

⇒ Dégradation des résultats d'analyse des images

Spectral binning (élargissement des bandes spectrales)

- Réduction de dimension
- Réduction du bruit

- Nombre de bandes ?
 - Dépend de l'instrument, de la complexité de l'image
- Nombre de bandes conservées est en général $>$ à celui des autres méthodes, car le spectral binning ne cherche pas de représentation optimale

Pourquoi projeter ?

EXPLORATION

Observer sous un angle nouveau.
Faire ressortir des structures
cachées ou inhabituelles.

COMPRESSION

Concentrer les informations
pertinentes sur un nombre
de bandes réduit.

Deux types d'algorithmes :

- **Direct**

- Diagonalisation d'une matrice caractéristique.
- Projection sur la base vecteurs propres.

- **Itératif**

- Optimisation d'un vecteur suivant un indice à chaque itération.
- Projection sur la base obtenue.



Maximisation d'un critère
propre à la projection.

- Variance.
- Kurtosis.
- Etc..

Prétraitements - Projection

→ Renforcer l'hypothèse de normalité

Dans des espaces de grande dimension, les projections linéaires dans des sous-espaces ont tendance à rendre les distributions gaussiennes.

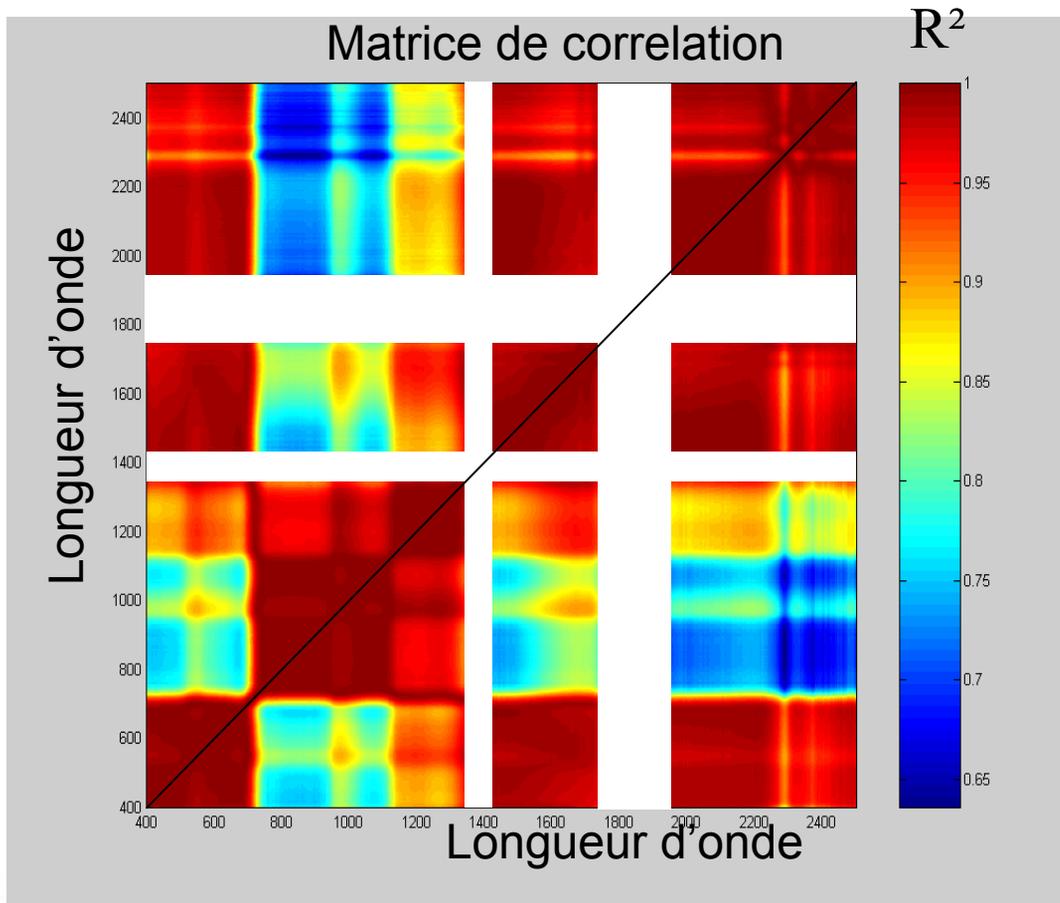
→ remplacer les N variables "bandes spectrales" par $N' < N$ variables décrivant au mieux les données, pour l'application visée

Prétraitements - Projection

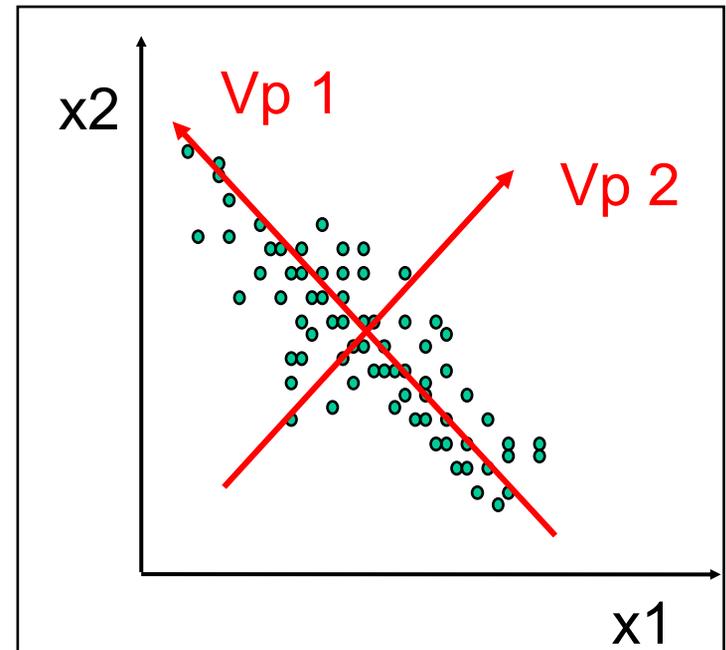
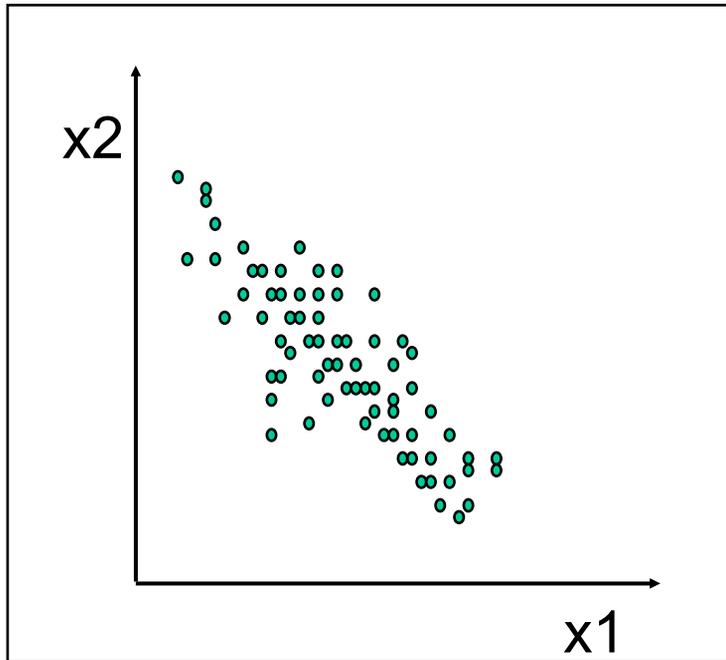
- Filtrer le bruit
- Supprimer la redondance inter-bandes



ACP,
MNF,
etc ...



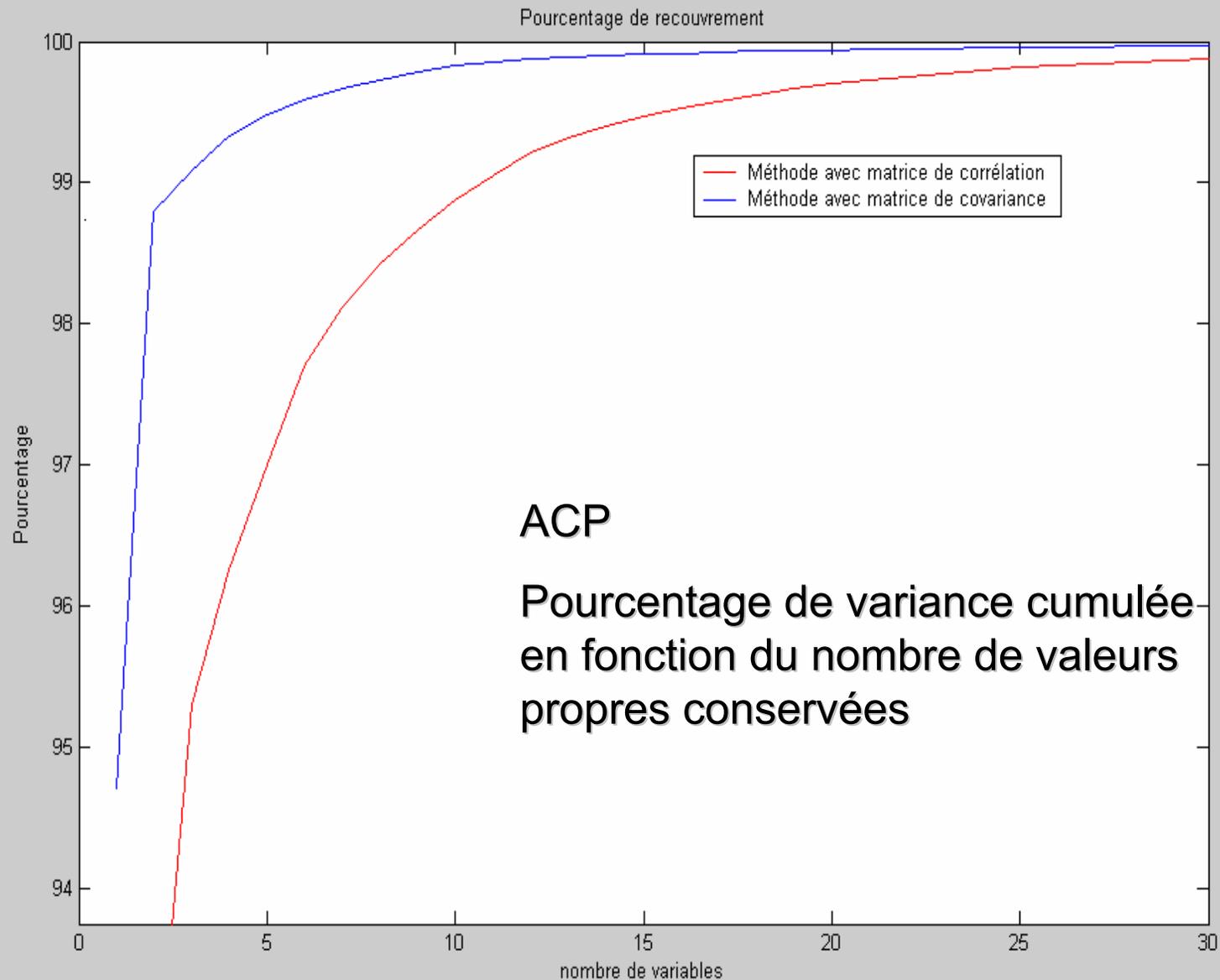
ACP : Diagonalisation de la matrice de variance covariance et projection sur les vecteurs propres de variances significatives



Pré-traitement - ACP

- Nombre de composantes à conservées :
 - Pas trop faible, sinon Probabilité de Détection (PD) diminue
 - Pas trop grand, sinon composantes bruitées \Rightarrow taux de Fausses alarmes (FA) risque d'augmenter
- Critère ?
 - Ex : garder les composantes principales / valeur propre \geq variance du bruit

Prétraitements - Objectifs

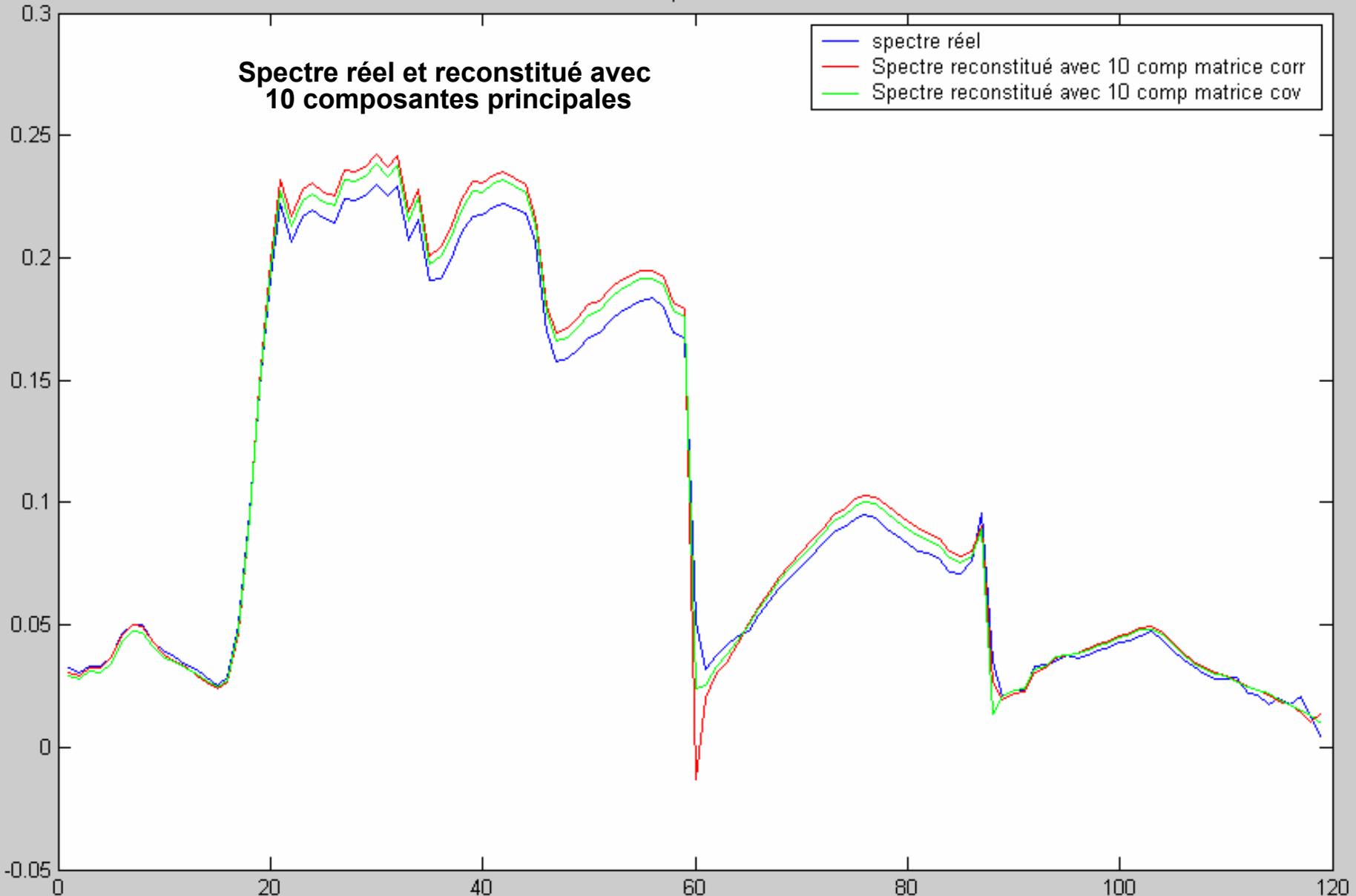


Prétraitements - Objectifs

Spectre

**Spectre réel et reconstitué avec
10 composantes principales**

- spectre réel
- Spectre reconstitué avec 10 comp matrice corr
- Spectre reconstitué avec 10 comp matrice cov



Prétraitements –Projection

- Recherche des axes de projection qui permettent au mieux de mettre en évidence les anomalies – KCP*
- Diagonalisation de :

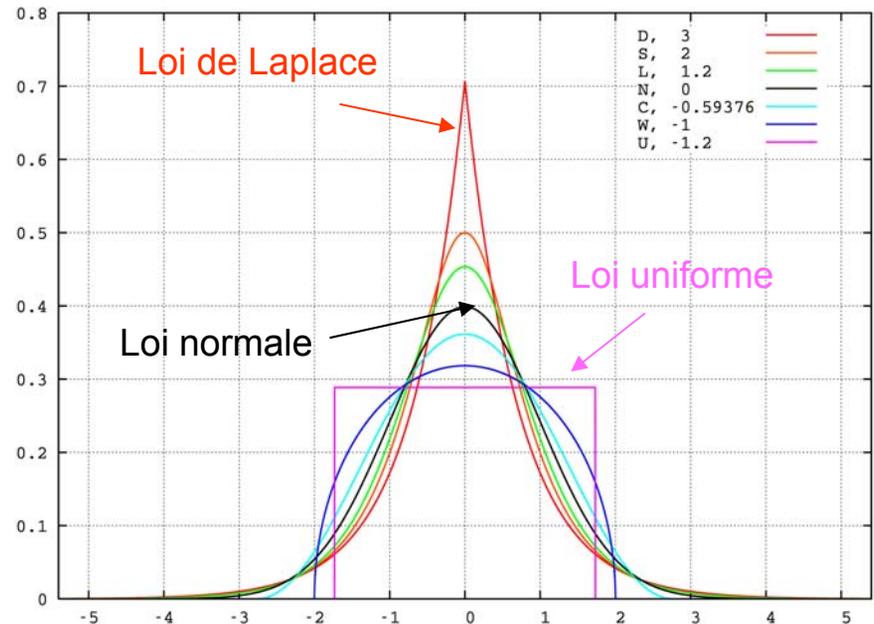
$$\Sigma^{-1} \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^t \Sigma^{-1} (X_i - \mu) (X_i - \mu) (X_i - \mu)^t$$

Matrice de kurtosis multivarié.

- Grandes valeurs propres = axes de fort Kurtosis
→ Met en valeur les anomalies sur des images ayant un fond peu hétérogène.



Des distributions multimodales peuvent avoir de forts Kurtosis !



*Daniel Peña, Francisco J. Prieto, Júlia Viladomat :

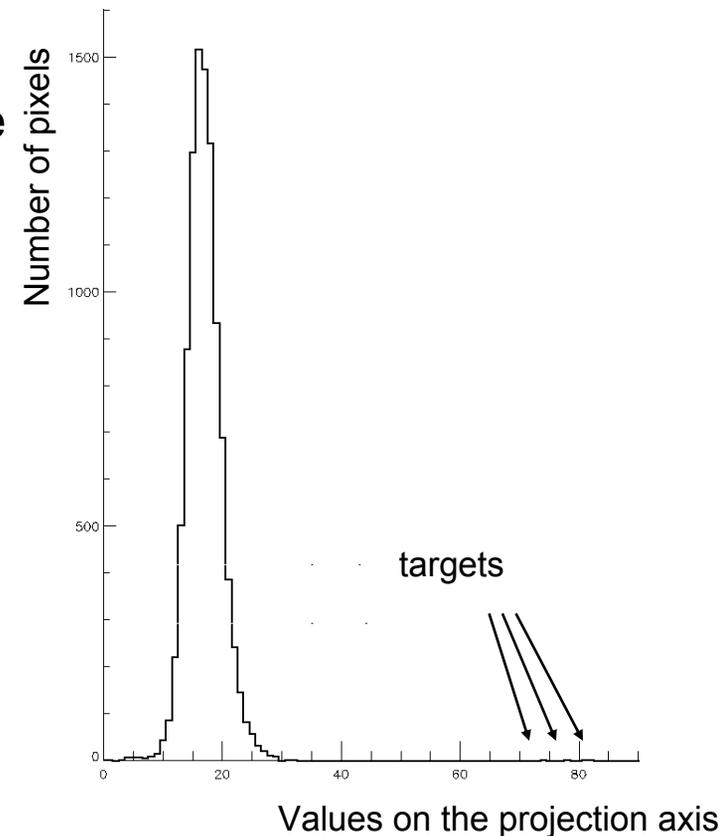
Eigenvectors of a kurtosis matrix as interesting directions to reveal cluster structure

Poursuite de projection

Ex : Détection d'anomalies par poursuite de projection

- **Méthode Itératif**
 - Optimisation d'un vecteur suivant un indice de projection à chaque itération.
 - Projection sur la base obtenue.

→ Recherche d'axes sur lesquels la distribution des pixels a des queues lourdes par rapport à une distribution gaussienne



Indices de projection

Skewness
$$\frac{1}{(N-1)S^3} \sum_{i=1}^n (Y_i - \bar{Y})^3$$

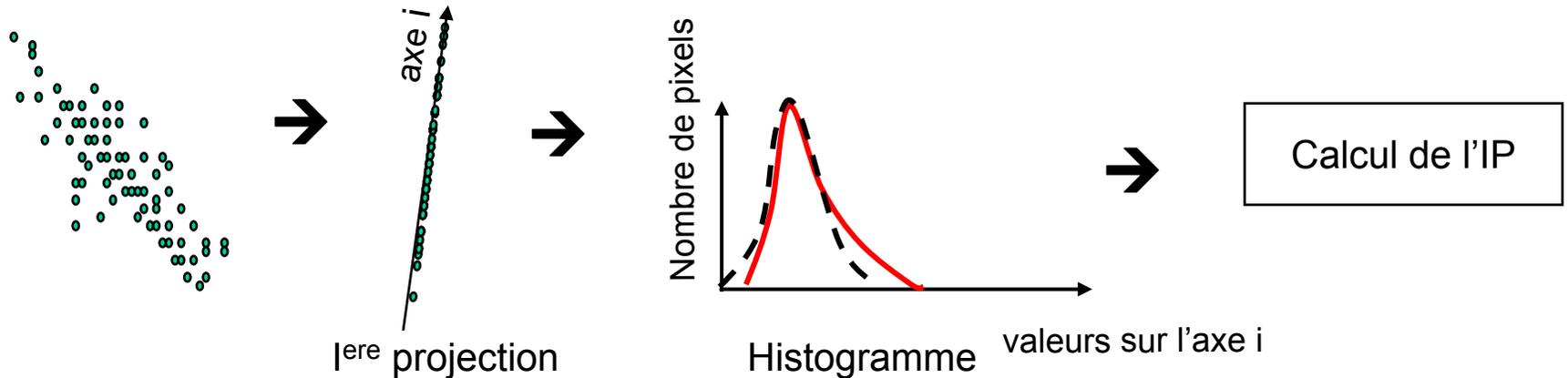
- skewness = 0 : distribution parfaitement symétrique
- skewness < 0 : queue gauche la plus lourde
- skewness > 0 : queue droite la plus lourde

Kurtosis
$$\frac{1}{(N-1)S^4} \sum_{i=1}^n (Y_i - \bar{Y})^4 - 3$$

- Kurtosis > 0 : distribution plus pointue qu'une gaussienne, queues lourdes
- Kurtosis < 0 : distribution plus plate qu'une gaussienne

Fort $|IP| \Rightarrow$ distribution très dissemblable de la gaussienne

Recherche des meilleurs axes de projection



- Plusieurs méthodes pour explorer l'espace spectral :
 - **Tous les vecteurs propres de l'ACP précédemment sélectionnés,**
 - **Projection aléatoire**
 - **Utilisation d'un algorithme génétique.**

Subspace - AD

- **Eliminer les composantes « fond »**

- En partant de l'hypothèse d'un mélange linéaire

$$H_0 : x = \sum_{i=1}^{N_b} \alpha_i b_i + n$$

$$H_1 : x = \alpha_t s + \sum_{i=1}^{N_b} \alpha_i b_i + n$$

- Les composantes « fond » sont largement représentées sur l'image et sont concentrées sur les premières composantes ACP/SVD.
 - supprimer les N_s premières composantes
 - projeter orthogonalement aux N_s premières composantes
- N_s ?
 - Si trop faible, il reste des structures fond
 - Si trop fort : on élimine les cibles également
- Intérêt : Utilisation pour RX global (très rapide), ou local.

Traitements des ombres

- Segmentation de l'image en zones ombre / zones soleil
 - Ex: Méthodes basées sur l'histogramme des luminances :
 - En général : 1er mode = pixels ombre
 - Classe Ombre / classe soleil
- Atténuation des contrastes en zones ombre et soleil :
 - Réduction de la dynamique de l'image en luminances : $\sqrt{\text{image}}$

- Autres méthodes

- MNF
- Méthodes de sélection de bandes
- Décomposition en ondelettes spectrales
- Extraction de caractéristiques spectrales (NDVI, localisation du « red edge », localisation des bandes d'absorption, etc.)
- Extraction de caractéristiques spatiales

Détection d'anomalies

1. Objectifs

2. Méthodes de détection d'anomalies

3. Prétraitement / Réduction de dimension

4. Critères d'évaluation de performance

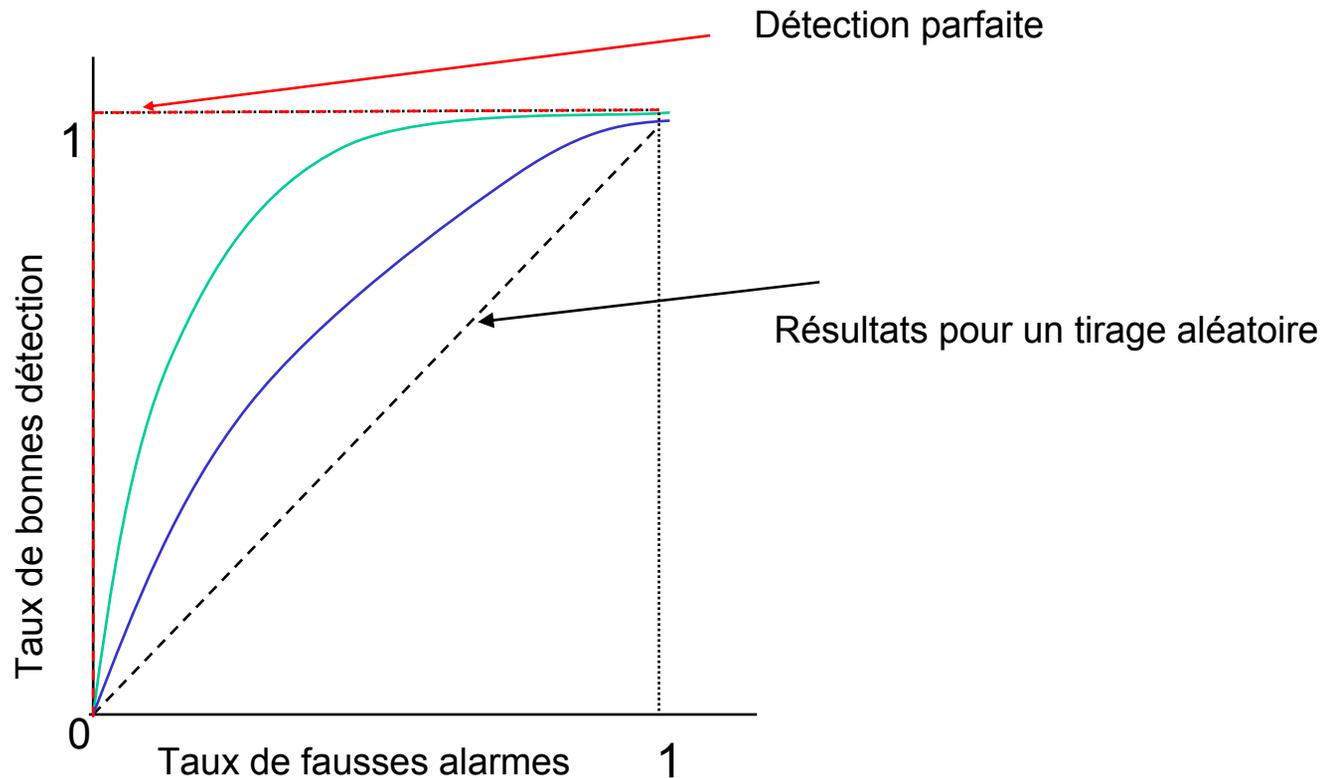
5. Quelles méthodes/ quels prétraitements pour quels types d'images ?

6. Références

Critères d'évaluation de performance

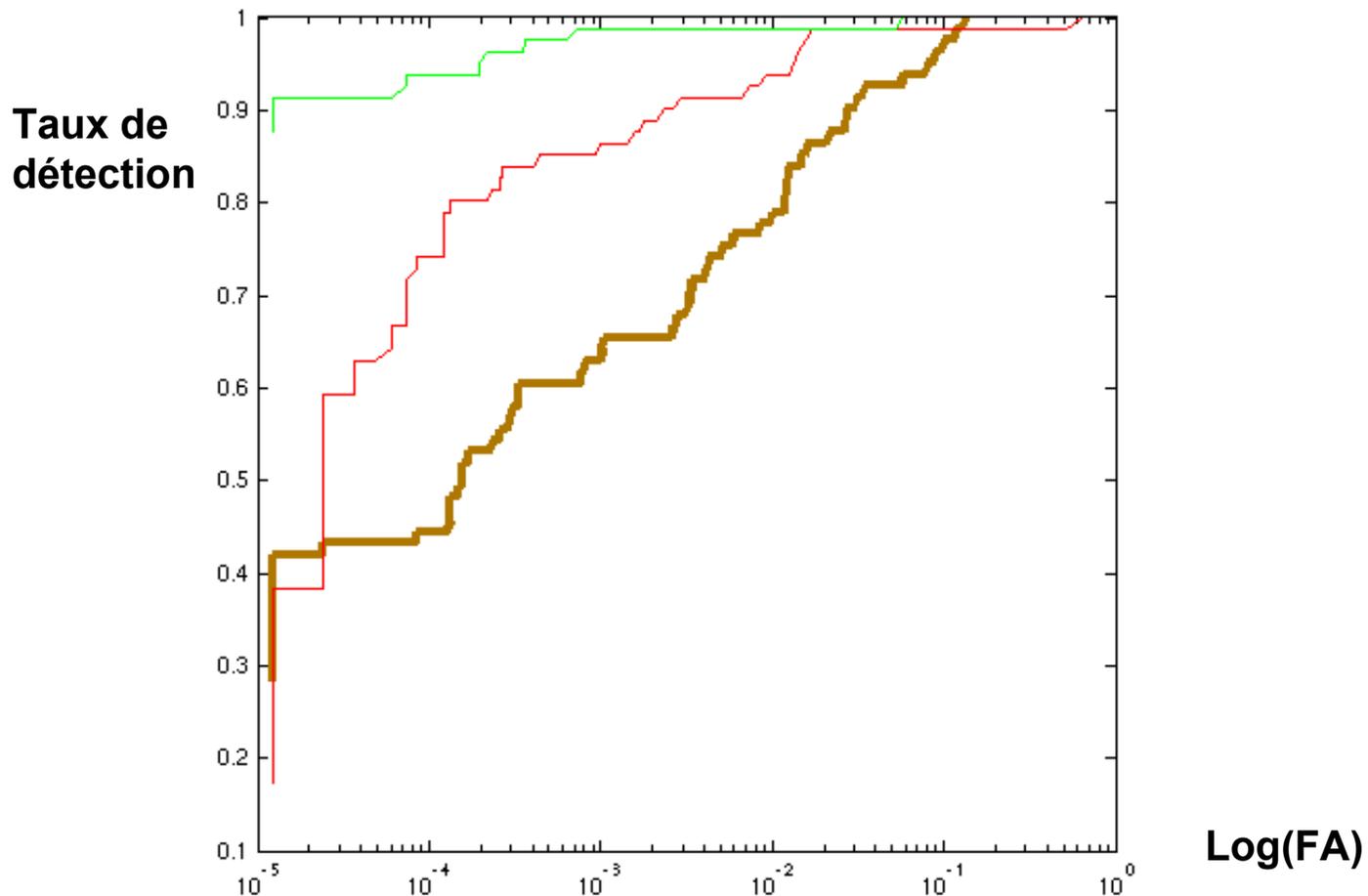
Courbes ROC obtenue en faisant varier la valeur de seuillage du RX :

Si on connaît l'emplacement des anomalies, on va pouvoir compter pour chaque valeur du seuil les bonnes détections et les fausses alarmes et tracer les courbes ROC



Critères d'évaluation de performance

- Courbes ROC généralement tracées en fonction de $\log(\text{FA})$
- LogAUC : Aire sous la courbe ROC avec en abscisse $\log(\text{FA})$



Détection d'anomalies

1. Objectifs

2. Méthodes de détection d'anomalies

3. Prétraitement / Réduction de dimension

4. Critères d'évaluation de performance

5. Exemple de Résultats

Quelles méthodes/ quels prétraitements pour quels types d'images ?

6. Références

Quelle méthode, quel pré-traitement ?

Images peu hétérogènes :

- RX local + KCP

Images très hétérogènes :

- GMM, CB-AD (Spectral binning, ACP)
- SS-AD (ACP, SVD)

Images contenant des cibles à l'ombre :

- traitements spécifiques « ombres » en + des autres pré-traitements

Ce sont de grandes tendances ... mais parfois les résultats sont surprenants !

Exemples de résultat

Ajout d'anomalies dans l'image

Localisation des anomalies

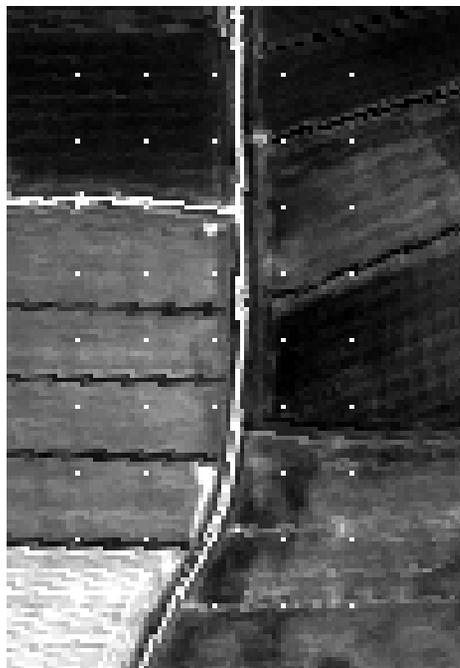
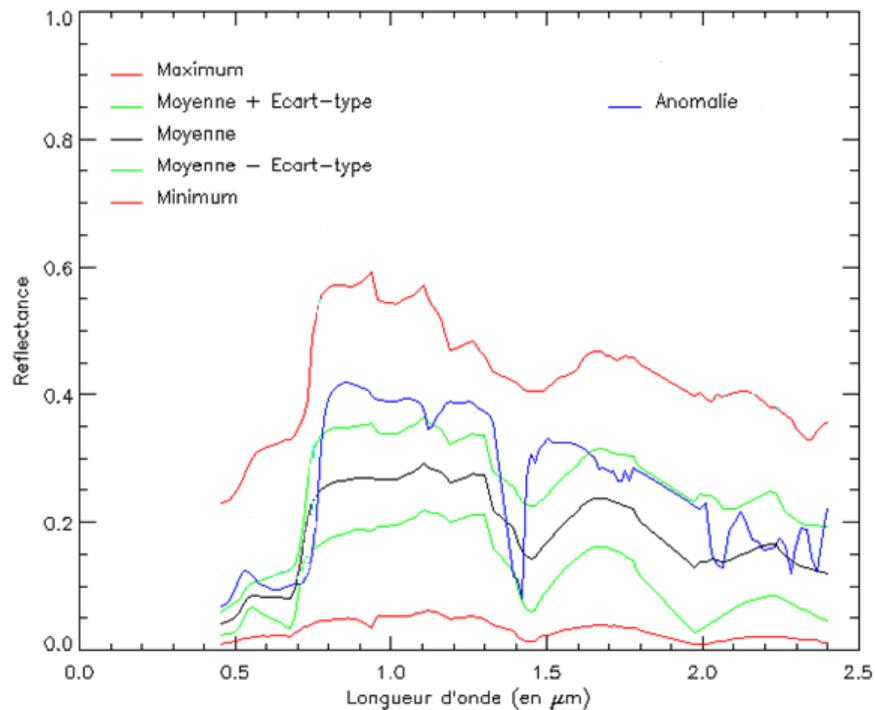


Image HyMap [0.4, 2.5 μm]
(Camargue)

Spectres anomalies/fond

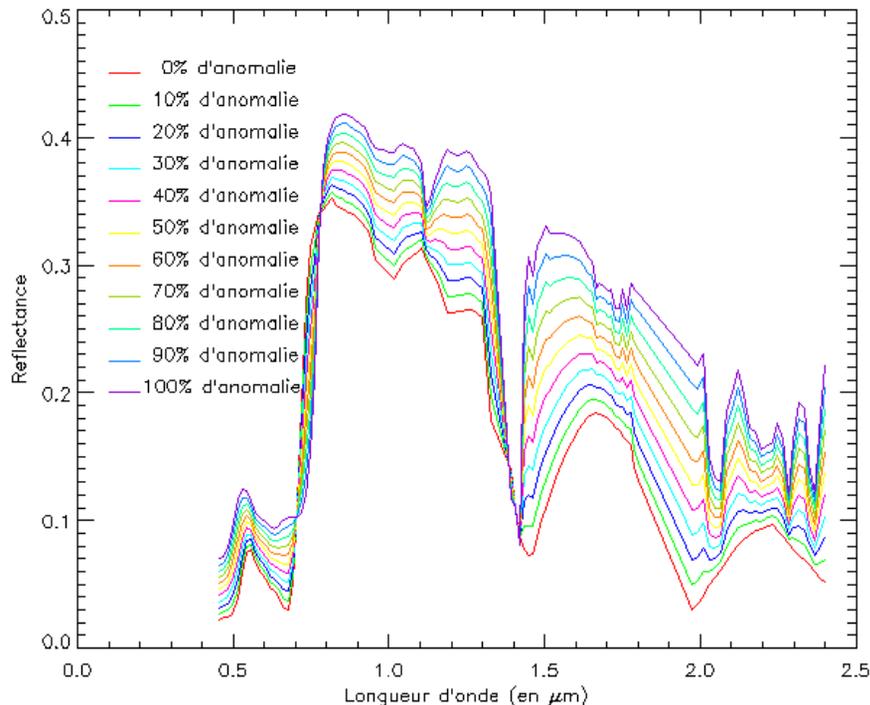


Détection d'anomalies sub-pixelliques

Le spectre à détecter s est une combinaison linéaire du spectre du pixel x lui-même et d'une anomalie a :

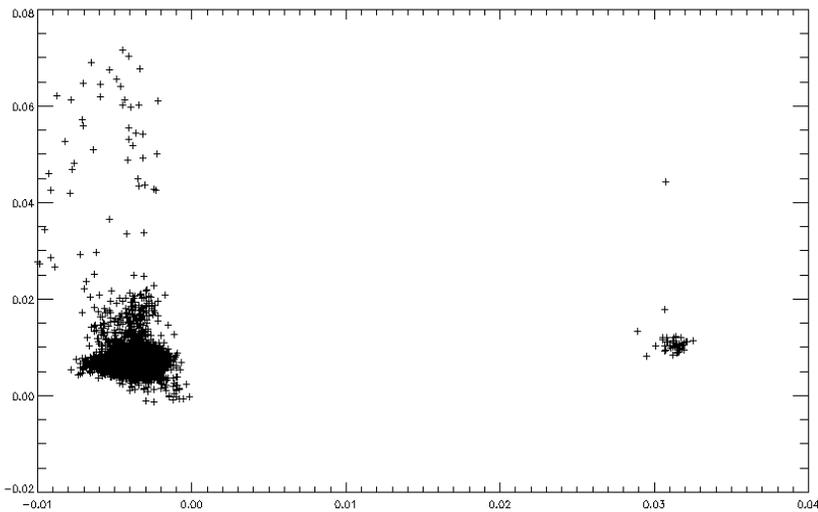
$$s = \alpha a + (1 - \alpha) x$$

où α est le pourcentage d'anomalie à insérer.

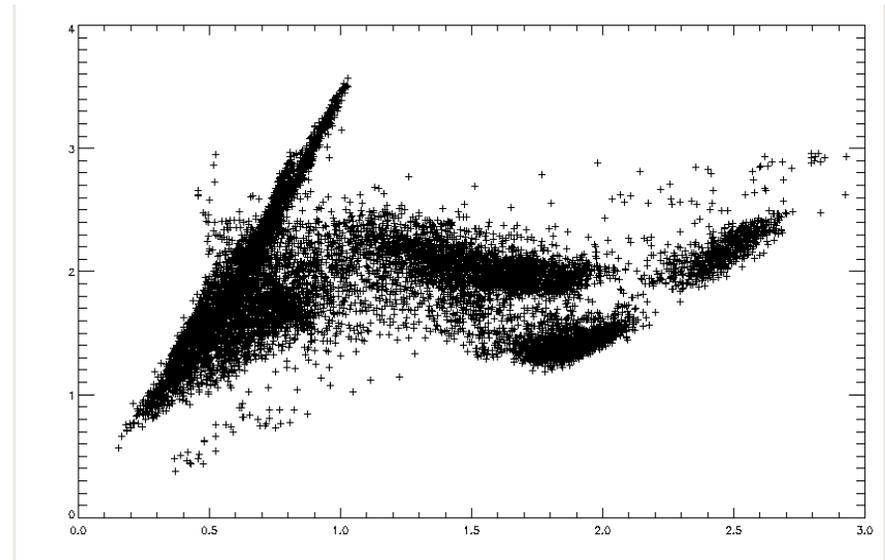


Exemples de résultat

Impact du pré-traitement



**Projection sur les 2 premiers
vecteurs propres KCP**

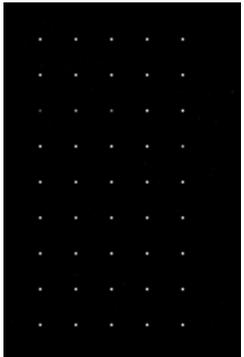
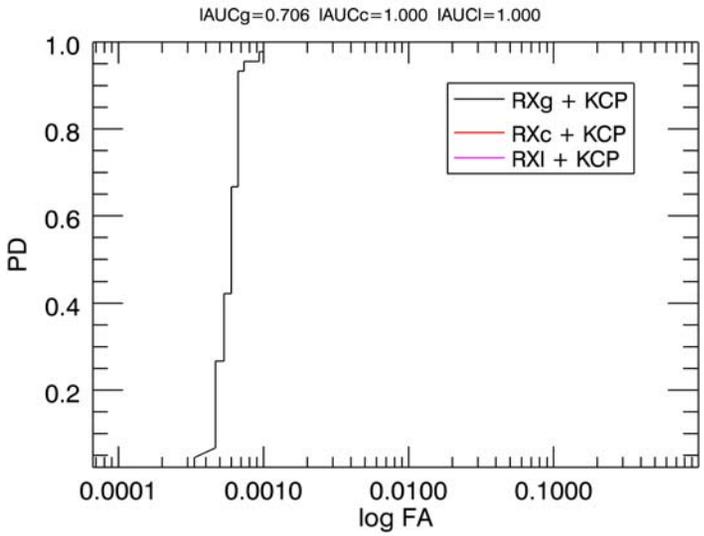


**Projection sur les 2 premiers
vecteurs propres ACP**

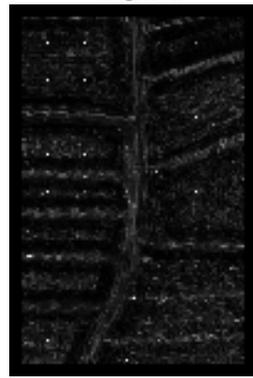
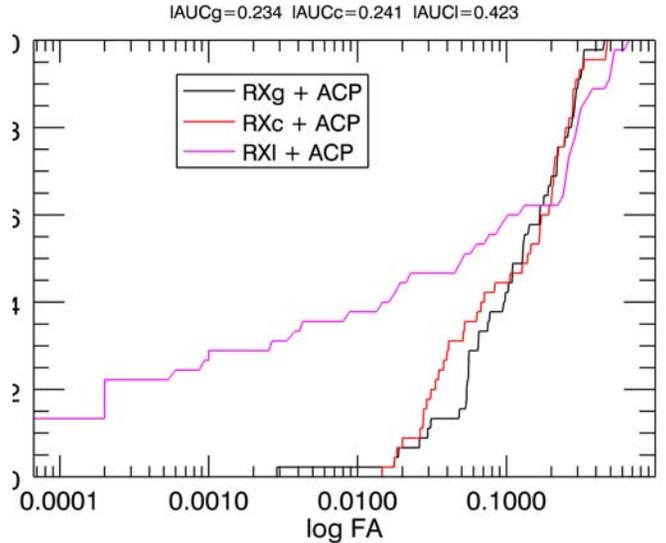
Exemples de résultat

Impact du pré-traitement

Résultat avec 2 composantes KCP



Résultat avec 10 composantes ACP



Références

Overview :

- Stefania MATEOLI, Marco DIANI, Giovanni CORSINI :

A tutorial overview of anomaly detection in hyperspectral image, IEEE A&S Systems Magazine, Part3: Tutorials, vol. 21, no. 3, June 2010.

Méthode RX :

- I. Reed and X. Yu, "Adaptive multiband cfar detection of an optical pattern with unknown spectral distribution," *IEEE ASSP*, vol. 38, no. 10, pp. 1760–1770, Oct 1990.

Méthodes GMM :

C. Willis, "Anomaly detection in hyperspectral imagery using statistical mixture models," in Proc. 2nd EMRS DTC Technical Conference, Edinburgh, 2005. [Online]. Available: <http://www.emrsdtc.com/conferences/2005/downloads/pdf/B16.pdf>

- I. Kasen, P. Goa, and T. Skauli, "Target detection in hyperspectral images based on multicomponent statistical models for representation of background clutter," in Proceedings SPIE, vol. 5612, 2004, pp. 258–264

Méthodes PP :

- Ifarraguerri and C.-I. Chang, "Unsupervised Hyperspectral Images Analysis with Projection Pursuit", *IEEE Trans. Geosci. Remote Sensing*, vol. 38, pp. 2529-2538, Nov. 2000

- S.-S. Chiang, C.-I. Chang and I. W. Ginsberg, "Unsupervised Target Detection in Hyperspectral Images Using Projection Pursuit", *IEEE Trans. Geosci. Remote Sensing*, vol. 39, pp. 1380-1391, Jul. 2001

Comparaison de méthodes

-Hyperspectral anomaly detection: Comparative evaluation in scenes with diverse complexity. D.Borghys, I. Kasen, V. Achard, C. Perneel, submitted to special Issue J. Electrical and Computer Engineering, 2012