

# THALES

## Apprentissage des environnements fouillis Radar Pathologiques par des techniques de classification supervisées et non-supervisées

Yann Cabanes, doctorant à l'IMB, CIFRE THALES

[www.thalesgroup.com](http://www.thalesgroup.com)



# Plan de l'exposé

---

- 1) Position du problème
- 2) Modélisation et représentation géométrique du problème
- 3) Algorithmes de classification

# THALES



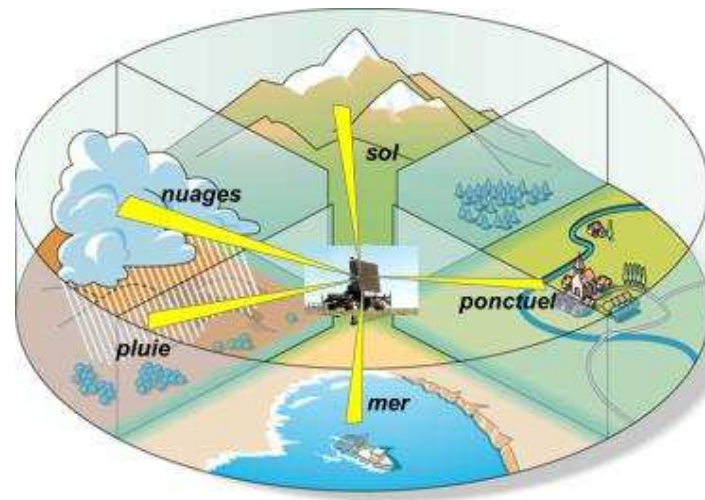
## 1) Position du problème

[www.thalesgroup.com](http://www.thalesgroup.com)



## Position du problème théorique

### Segmentation de l'espace en fonction de la nature du terrain

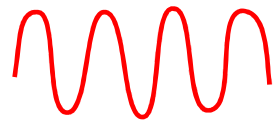


**Objectif** : Rédiger un algorithme qui regroupe les zones de l'espace autour du radar en fonction de la nature du terrain observé (forêt, champs, mer, montagne, désert...)

# Présentation des données

## Données complexes

Signal émis = une impulsion



Signal reçu



Radar



Environnement

Le radar enregistre une donnée complexe :  $x = A * e^{i\varphi}$

A : amplitude du signal reçu

$\varphi$  : phase du signal reçu

# Présentation des données

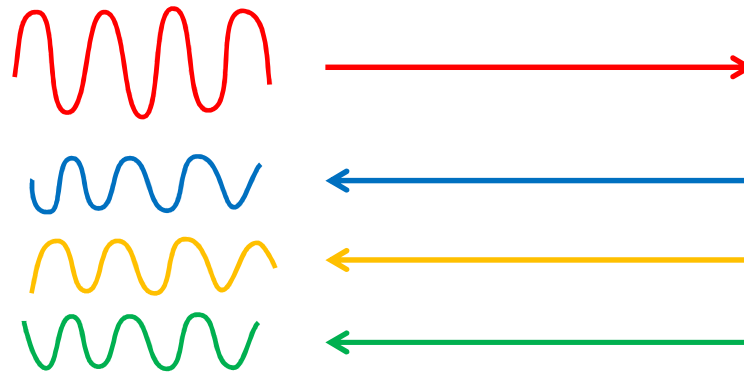
Une récurrence est constituée d'une impulsion et d'un temps d'écoute

Une case distance  
≈ 60 m

Signal émis = une impulsion



Radar



Signaux reçus



Environnement

Le radar enregistre une suite de données complexes :

$$(x_1, x_2, x_3, \dots)$$

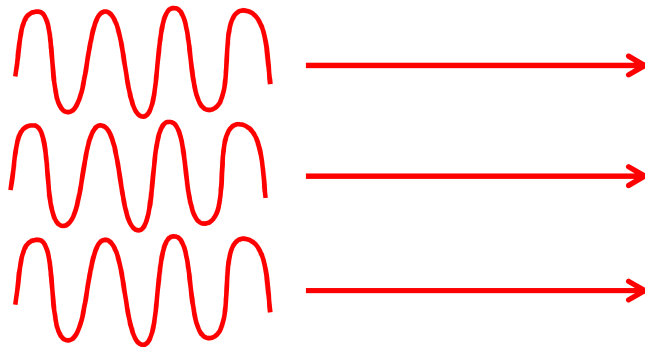
# Présentation des données

This document may not be reproduced, modified, adapted, published, translated, in any way, in whole or in part or disclosed to a third party without the prior written consent of Thales - © Thales 2015. All rights reserved.

Une rafale (ou burst) est une succession de récurrences Une case distance



Impulsions émises



Axe temporel

Le radar enregistre une suite de données complexes :

$$X = \begin{bmatrix} x_{1,1} & x_{1,2} & x_{1,3} & \dots & x_{1,p} \\ x_{2,1} & x_{2,2} & x_{2,3} & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{n,1} & x_{n,2} & x_{n,3} & \dots & x_{n,p} \end{bmatrix}$$

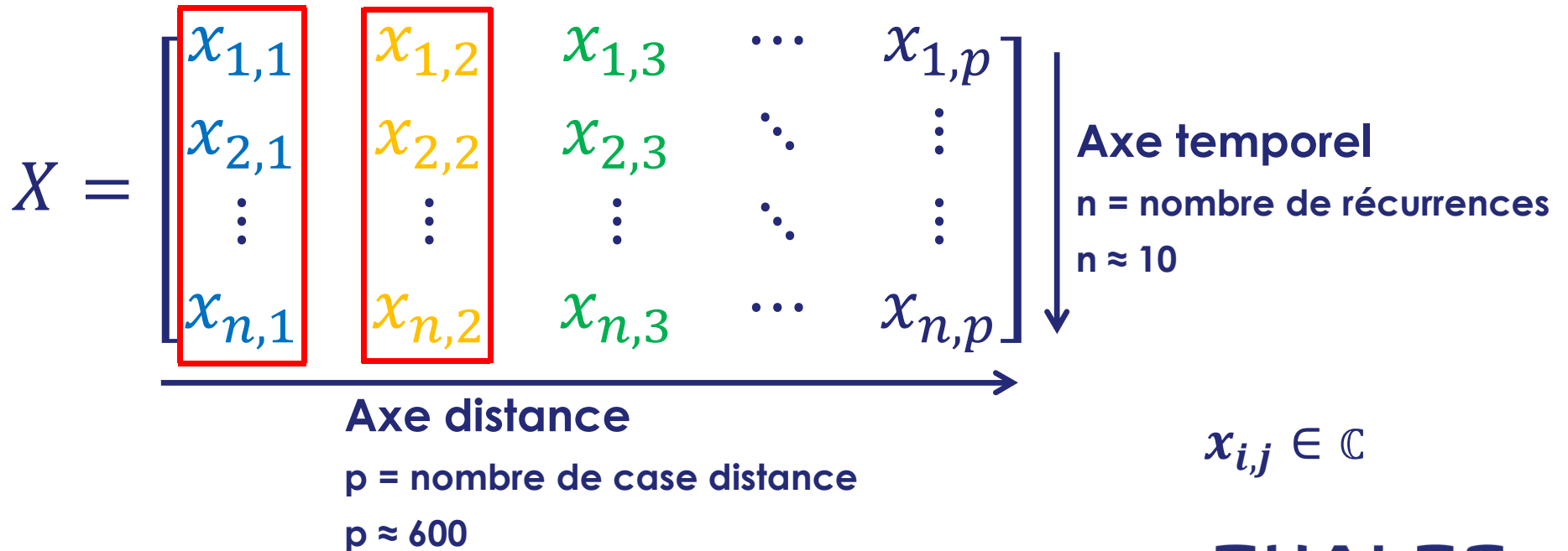




## Classification case par case

### Segmentation case par case

- Données reçues pour une rafale, pour le faisceau d'élévation nulle.

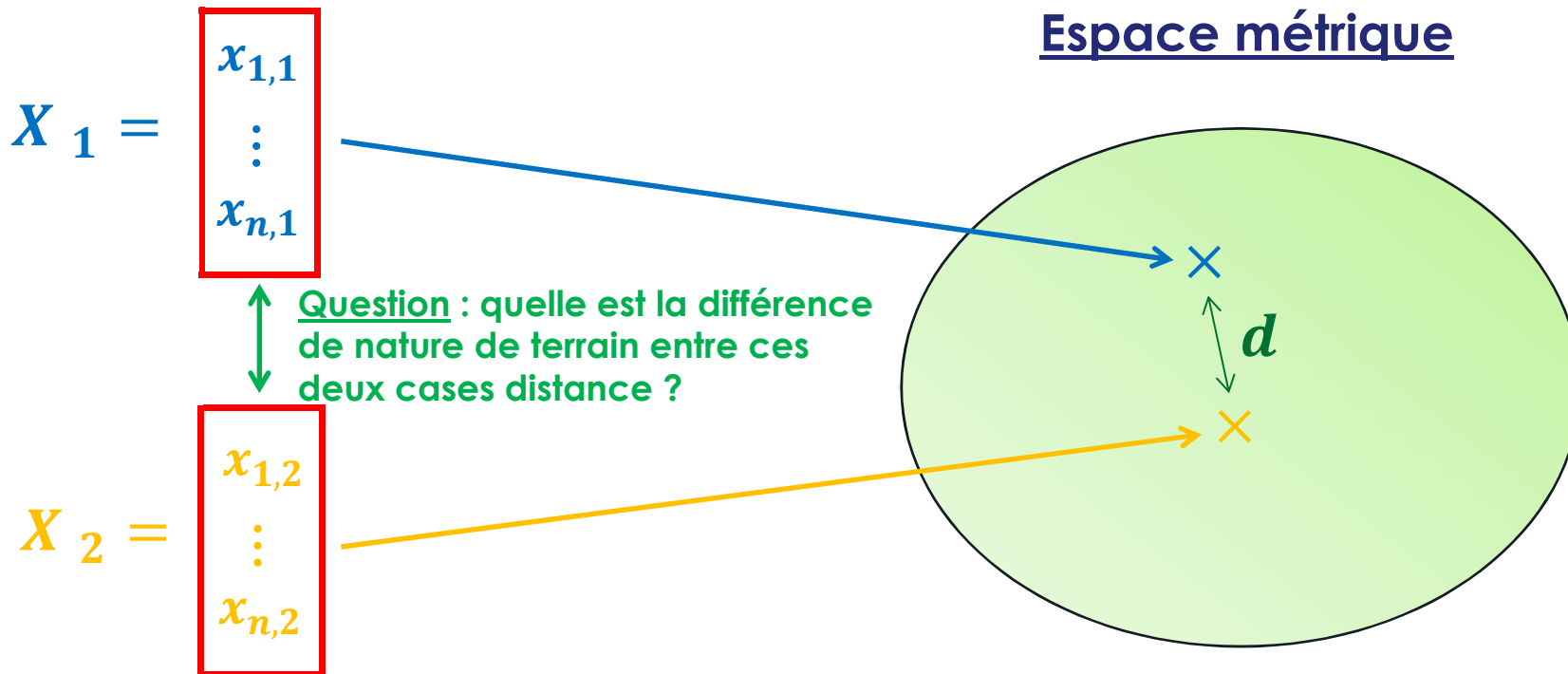




# Problème mathématique

## Segmentation case par case

- Transformation des données vers un espace métrique



This document may not be reproduced, modified, adapted, published, translated, in any way, in whole or in part or disclosed to a third party without the prior written consent of Thales - © Thales 2015. All rights reserved.

# THALES

## 2) Modélisation et représentation géométrique du problème

[www.thalesgroup.com](http://www.thalesgroup.com)



## Hypothèses de modélisation

Hypothèses : Nous supposons le signal stationnaire, de moyenne nulle, modélisable par un processus autorégressif gaussien.

**Formulation équivalente (Equation utilisée pour simuler les données) :**

$$Z = \sqrt{\lambda} R^{1/2} x + b_{radar}$$

Avec :

$\lambda$  : terme de texture (puissance), réel positif

$R$  : matrice Toeplitz Hermitienne Définie Positive

$x, b_{radar}$  : vecteurs gaussiens centrés réduits iid de dimension le nombre d'impulsions de la rafale.

Modèle **SIRV** (Spherically Invariant Random Vectors)

**THALES**

## Lien entre matrice Toeplitz et stationnarité

Mesure de la série temporelle du signal digitalisé (pour une direction donnée)

$$\mathbf{Z} = \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} \text{ avec } z_i \in \mathbb{C}$$

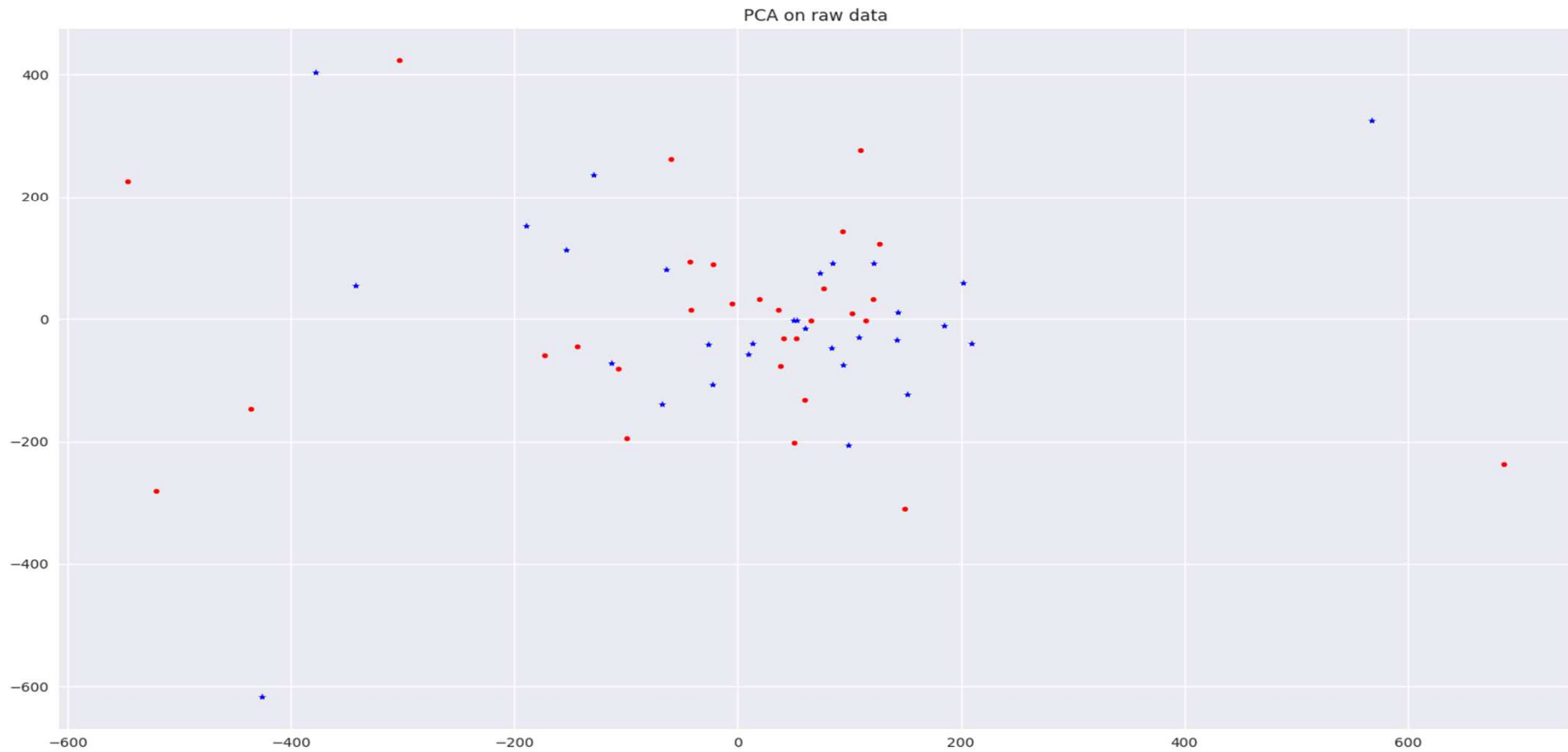
$$\mathbf{R}_n = E[\mathbf{Z}\mathbf{Z}^+] = \begin{bmatrix} r_0 & r_1^* & \cdots & r_{n-1}^* \\ r_1 & r_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & r_1^* \\ r_{n-1} & \cdots & r_1 & r_0 \end{bmatrix} \quad \text{Stationnaire} \implies \mathbf{R} \text{ Toeplitz}$$

avec  $r_k = E[z_m z_{m-k}^*]$

avec  $r_k$  : coefficient de corrélation dont on déduit le spectre Doppler

$$|S_Z(f)|^2 = \left| \sum_{k=-(n-1)}^{n-1} r_k e^{-j2\pi f k} \right|^2$$

## PCA (Principal Component Analysis) sur les vecteurs complexes



# La métrique

## Segmentation case par case

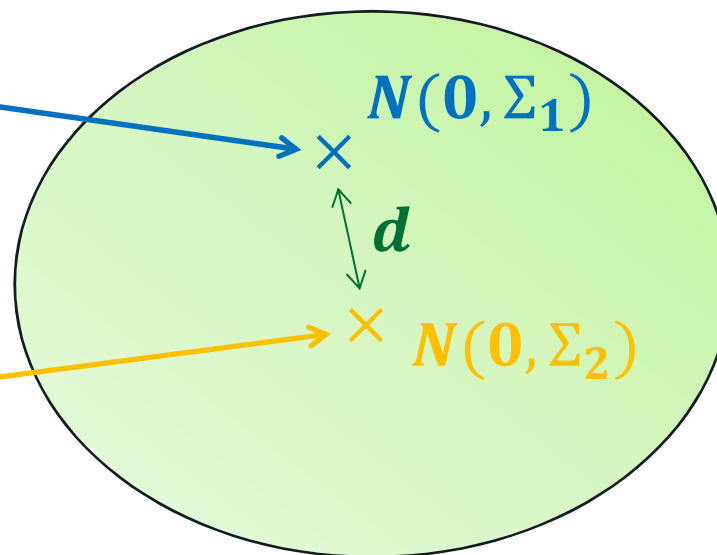
- Transformation des données vers un espace métrique

### Espace métrique des gaussiennes multivariées de moyenne nulle

$$X_1 = \begin{bmatrix} x_{1,1} \\ \vdots \\ x_{n,1} \end{bmatrix}$$

Question : quelle est la différence de nature de terrain entre ces deux cases distance ?

$$X_2 = \begin{bmatrix} x_{1,2} \\ \vdots \\ x_{n,2} \end{bmatrix}$$



$$\Sigma_i = E(X_i X_i^*) \quad \text{THALES}$$

## La métrique de la géométrie de l'information

**Problème** : calculer des distances entre des lois de probabilités.

**Exemple** : Comment calculer la distance entre deux lois gaussiennes  $N(m_1, \sigma^2_1)$  et  $N(m_2, \sigma^2_2)$  ?

→ Nous allons introduire une métrique issue de la géométrie de l'information.

Commençons par définir la matrice de Fischer  $I(\theta)$  d'une densité  $f$  :

$$I(\theta)_{i,j} = -E_{\theta} \left( \frac{\partial^2 \log(f_{\theta}(x))}{\partial \theta_i \partial \theta_j^*} \right)$$



## La métrique de la géométrie de l'information

Pour la loi gaussienne  $N(m, \sigma^2)$ , nous obtenons :

$$I(\theta) = \begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{bmatrix}$$

Remarque : La matrice de Fisher est la matrice qui intervient dans la borne de Cramer-Rao :

$$R_{\hat{\theta}} = E \left[ (\theta - \hat{\theta})(\theta - \hat{\theta})^+ \right] \geq I(\theta)^{-1}$$

Avec  $\hat{\theta}$  estimateur sans biais de  $\theta$ .

## La métrique de la géométrie de l'information

La matrice de Fischer  $I(\theta)$  est une matrice Hermitienne Définie Positive : c'est la matrice d'un produit scalaire.

La longueur d'un chemin  $\gamma$  paramétré par le temps  $t$  entre les instants  $t_1$  et  $t_2$  est :

$$L(\gamma) = \int_{t_1}^{t_2} \sqrt{\|\gamma'(t)\|_{\theta_t}^2} dt \quad \text{avec} \quad \|\gamma'(t)\|_{\theta_t}^2 = \gamma'(t)^* I_{\theta_t} \gamma'(t)$$

On définit la distance entre deux densités de probabilités  $f_1$  et  $f_2$  comme la longueur du plus court chemin qui part de  $f_1$  à l'instant initial et arrive en  $f_2$  à l'instant final.

Remarque : la métrique de Fischer est la seule métrique sur les densités de probabilités qui soit invariante par changement de paramétrisation.

## Théorème de Trench/Verblunsky pour les matrices Toeplitz Hermitiennes Définies Positives

- Toute matrice Toeplitz Hermitienne Définie Positive peut être représentée dans l'espace produit  $(P_0, \mu_1, \dots, \mu_n) \in \mathbb{R}_+^* \times D^{n-1}$  :
- $P_0$  est un paramètre d'échelle (puissance du signal)
  - $\mu_k$  sont des paramètres de forme (forme du spectre) appelés coefficients de réflexion/Verblunsky

$$\begin{aligned} \varphi : THDP(n) &\rightarrow \mathbb{R}_+^* \times D^{n-1} \\ R_n &\mapsto (P_0, \mu_1, \dots, \mu_n) \\ \text{avec } D &= \{z \in \mathbb{C} \mid |z| < 1\} \end{aligned}$$

L'information Doppler peut être représentée dans  $D^{n-1}$

Théorème de Trench : Si l'on considère que la matrice Toeplitz Hermitienne Définie Positive est la matrice de corrélation d'un signal stationnaire, alors les coefficients dans  $D^{n-1}$  sont les coefficients d'un modèle autorégressif.

## La métrique

Grâce à la métrique de la géométrie de l'information restreinte au sous-espace des gaussiennes de moyenne nulle et de matrice de covariance **Toeplitz** Hermitiennes Définies Positives, nous pouvons calculer la distance entre  $N(\mathbf{0}, \Sigma_1)$  et  $N(\mathbf{0}, \Sigma_2)$ .

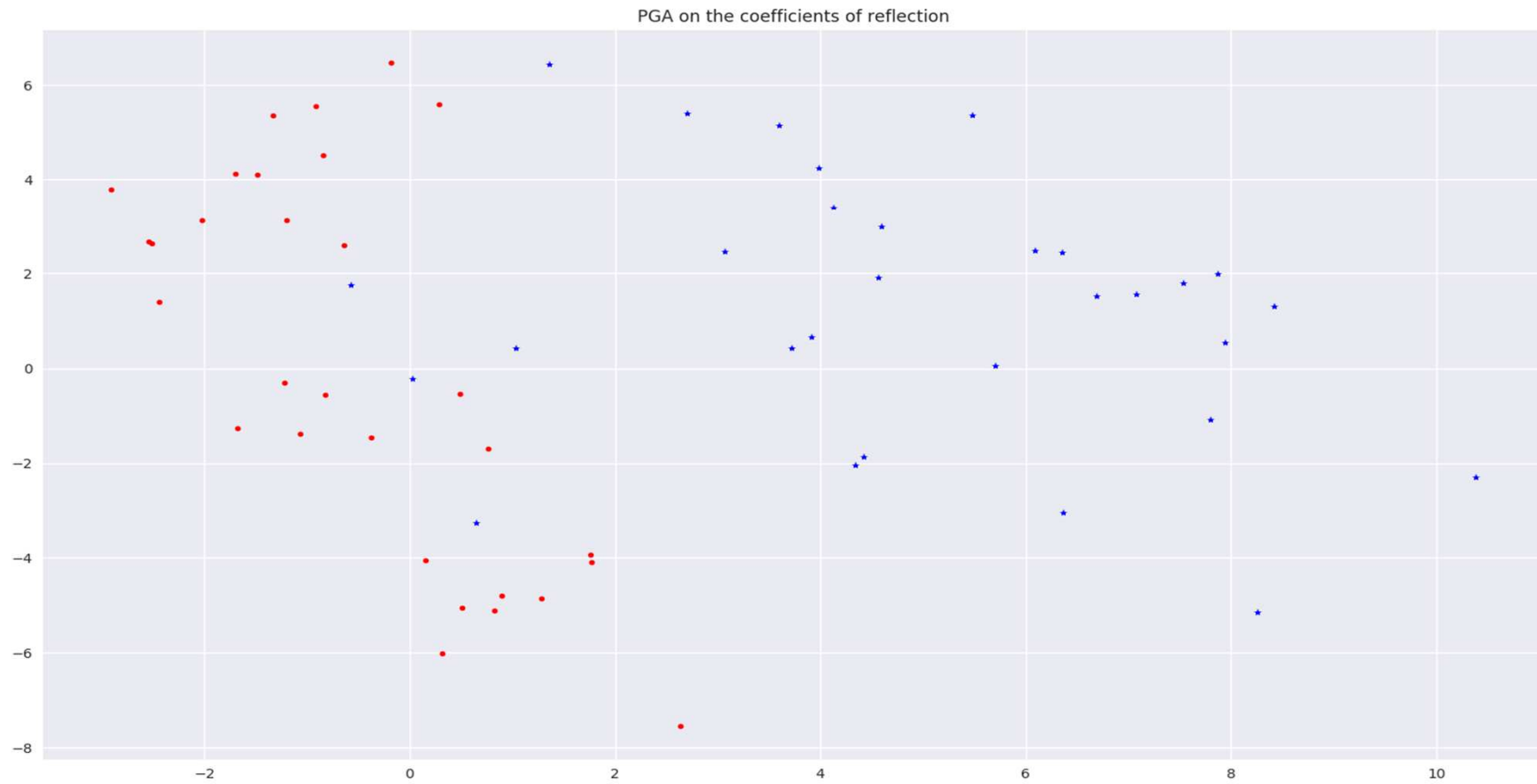
Or les matrices  $\Sigma_1$  et  $\Sigma_2$  sont des matrices Toeplitz Hermitienne Définies Positives, nous pouvons donc les représenter dans l'espace :  $\mathbb{R}_+^* \times D^{n-1}$ .

La métrique se transpose dans cet espace en une métrique produit :

$$d^2 \left[ \left( \bar{P}_{0,1}, \{\bar{\mu}_{i,1}\}_{i=1}^{N-1} \right), \left( \bar{P}_{0,2}, \{\bar{\mu}_{i,2}\}_{i=1}^{N-1} \right) \right] = N \log^2 \left( \frac{\bar{P}_{0,2}}{\bar{P}_{0,1}} \right) + \sum_{i=1}^{N-1} (N-i) \left( \frac{1}{2} \log \left( \frac{1+\delta_i}{1-\delta_i} \right) \right)^2$$

$$\text{avec } \delta_i = \left| \frac{\bar{\mu}_{i,1} - \bar{\mu}_{i,2}}{1 - \bar{\mu}_{i,1} \bar{\mu}_{i,2}^*} \right|$$

## PGA (Principal Geodesic Analysis) on the coefficients of reflection



**THALES**

### 3) Algorithmes de classification

[www.thalesgroup.com](http://www.thalesgroup.com)



## Données simulées

Equation utilisée pour simuler les données :

$$Z = \sqrt{\lambda} R^{1/2} x + b_{radar}$$

Avec :

$\lambda$  : terme de texture (puissance), réel positif

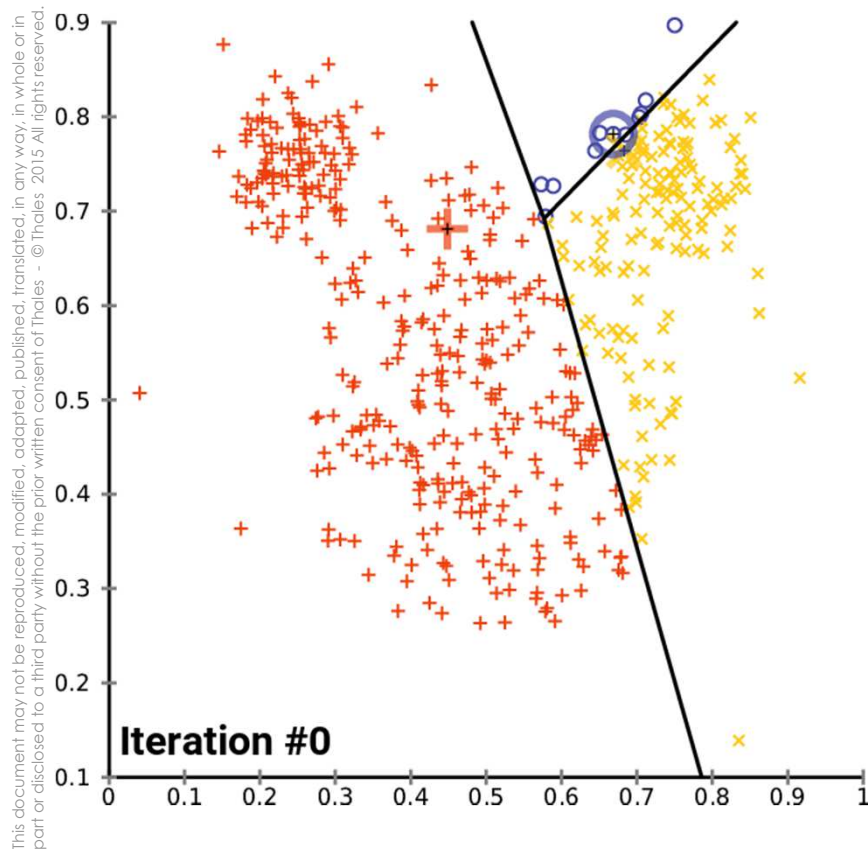
$R$  : matrice Toeplitz Hermitienne Définie Positive

$x, b_{radar}$  : vecteurs gaussiens centrés réduits iid de dimension le nombre d'impulsions de la rafale.

Modèle **SIRV** (Spherically Invariant Random Vectors)



# Un algorithme de classification non-supervisé : les k-means



## Algorithme des k-means pour N clusters :

Initialisation : Choisir au hasard N points parmi les points à classifier. Ils représentent maintenant les barycentres de chaque cluster.

Pour  $i = 1, \dots, n$  :

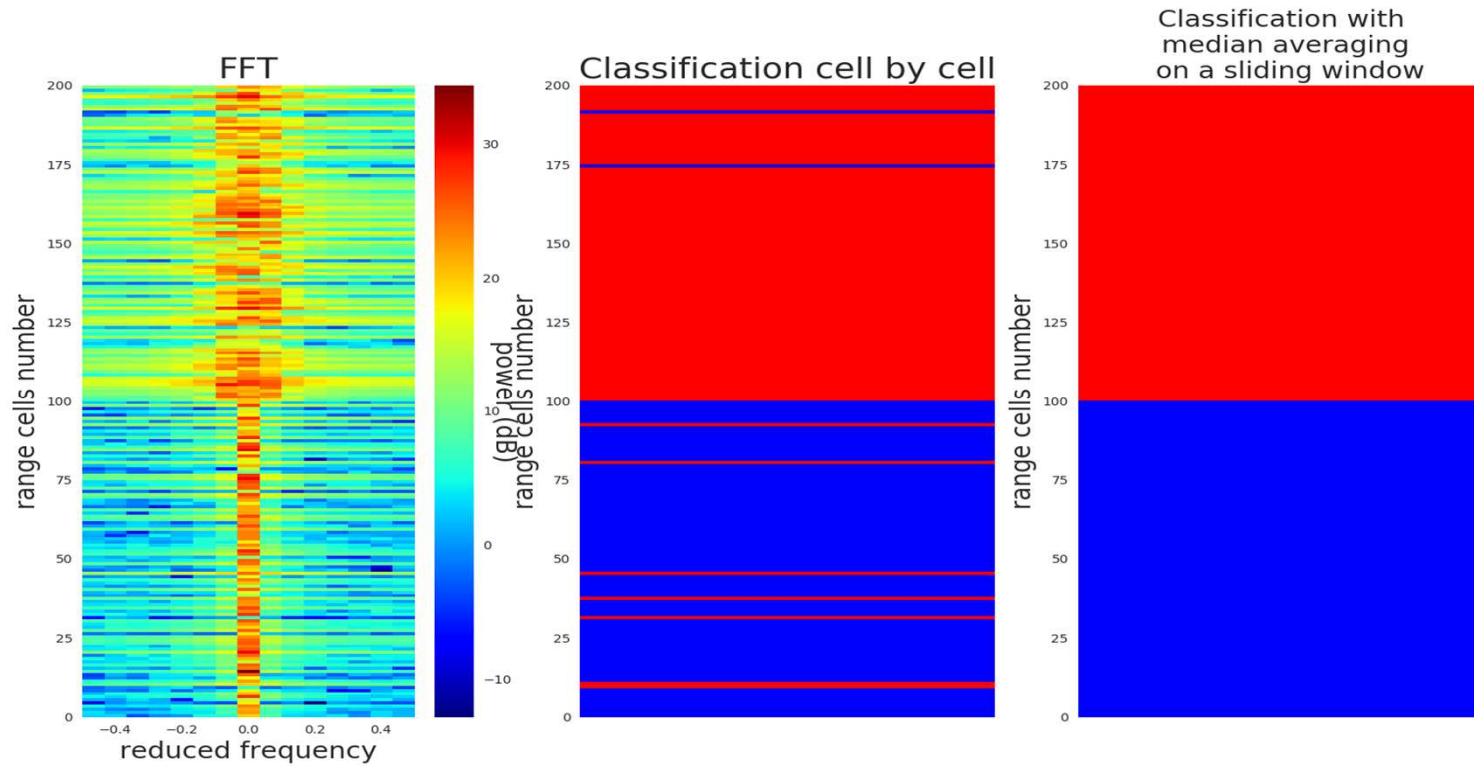
- Affecter chaque point au cluster dont le barycentre est le plus proche.
- Calculer le nouveau barycentre de chaque cluster.

Retourner :

Chaque point est labellisé en fonction de son cluster final.

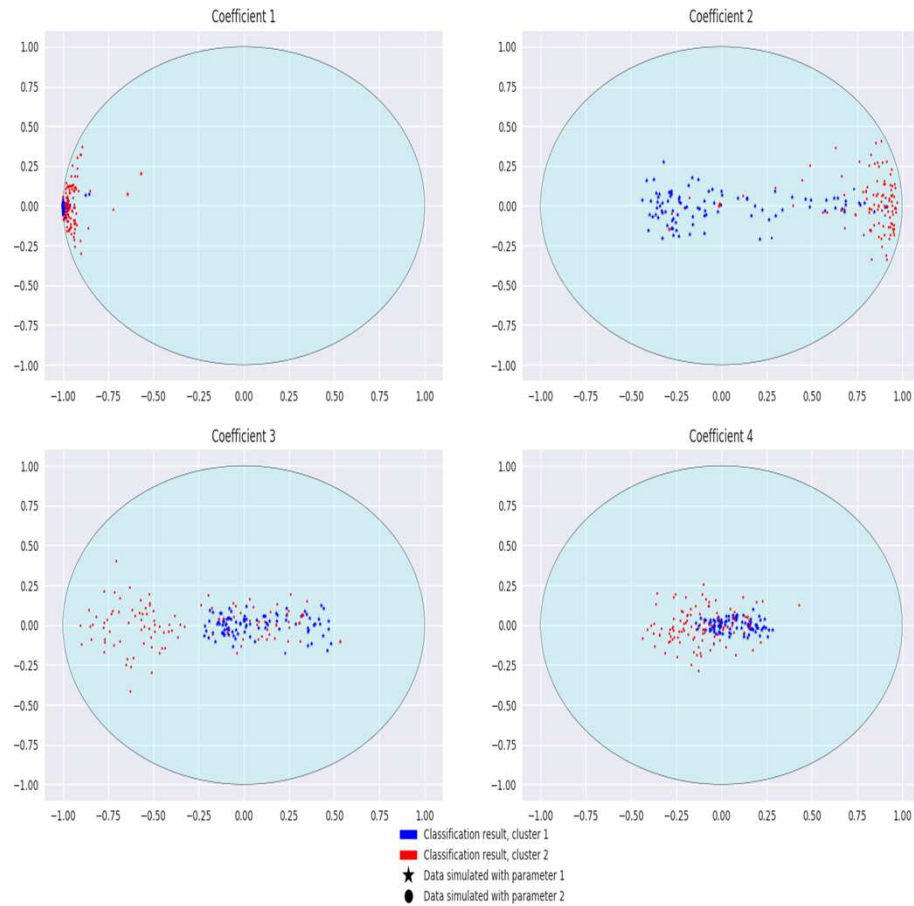
# Visualisation des données simulées et de la classification

This document may not be reproduced, modified, adapted, published, translated, in any way, in whole or in part or disclosed to a third party without the prior written consent of Thales - © Thales 2015. All rights reserved.



# Visualisation des coefficients de réflexion

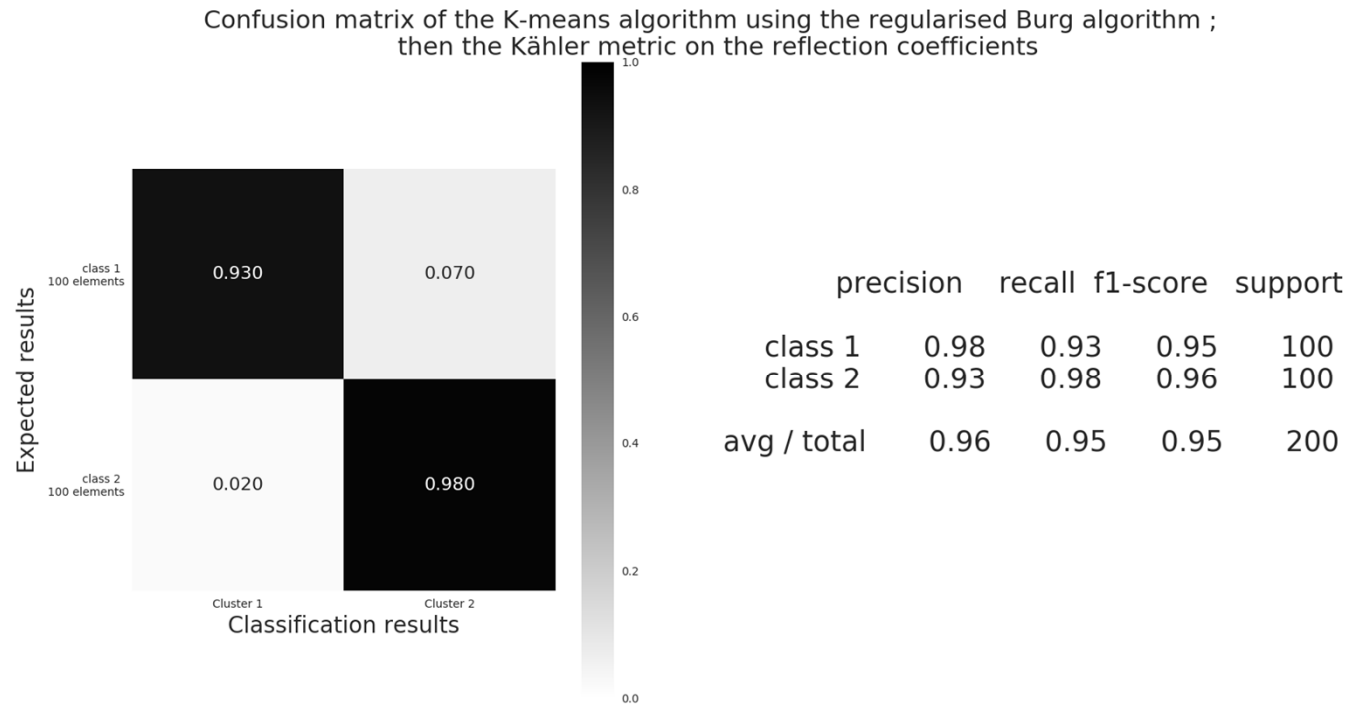
First Coefficients of Reflection



This document may not be reproduced, modified, adapted, published, translated, in any way, in whole or in part or disclosed to a third party without the prior written consent of Thales - © Thales 2015. All rights reserved.

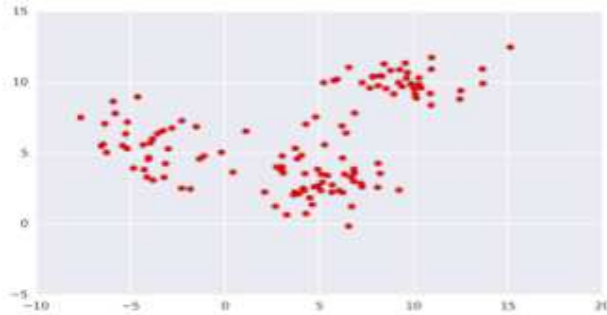
# Visualisation de la classification : matrice de confusion et f1-score

This document may not be reproduced, modified, adapted, published, translated, in any way, in whole or in part or disclosed to a third party without the prior written consent of Thales - © Thales 2015. All rights reserved.

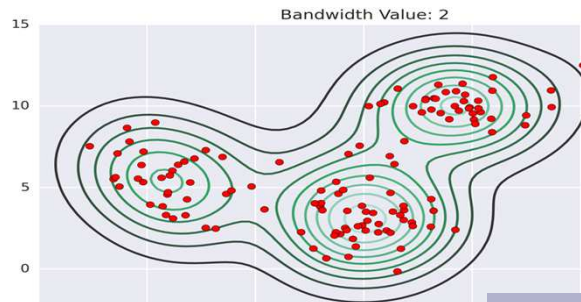
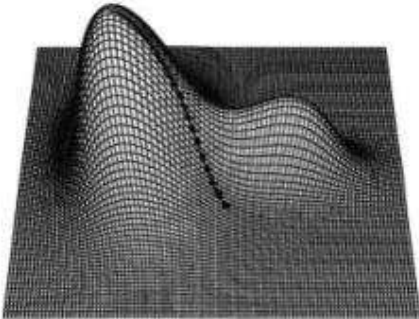


# Un second algorithme de classification non-supervisé : le mean-shift

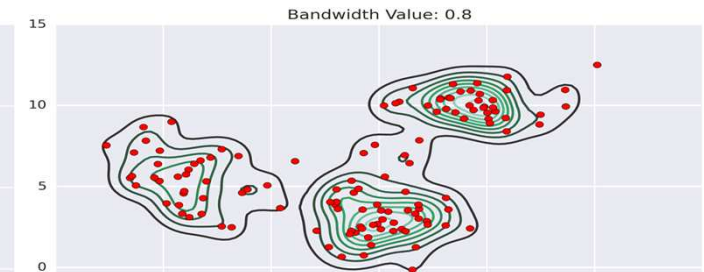
**Etape 1** : choisir une taille de noyau et compter par zone



**Etape 2** : Pour chaque point remonter suivant les pentes de gradients



**Noyau large  
(3 classes)**



**Difficulté: Trouver  
la bonne taille du  
noyau**

**Noyau étroit  
(7 classes)**

**THALES**

## Fin de la présentation

Merci pour votre attention