Statistical and geometrical tools for the classification of highly textured polarimetric SAR images

Pierre Formont
ONERA / SONDRA

PhD defense
Under the supervision of Frédéric Pascal, Jean-Philippe Ovarlez & Laurent Ferro-Famil (PhD director)

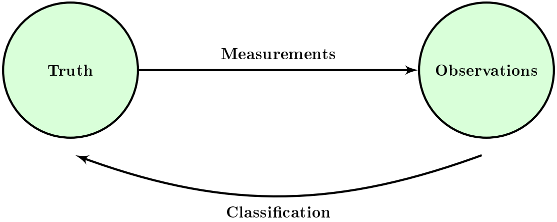
Co-funded by the ONERA and the DGA

December 10, 2013

Classification

Goal

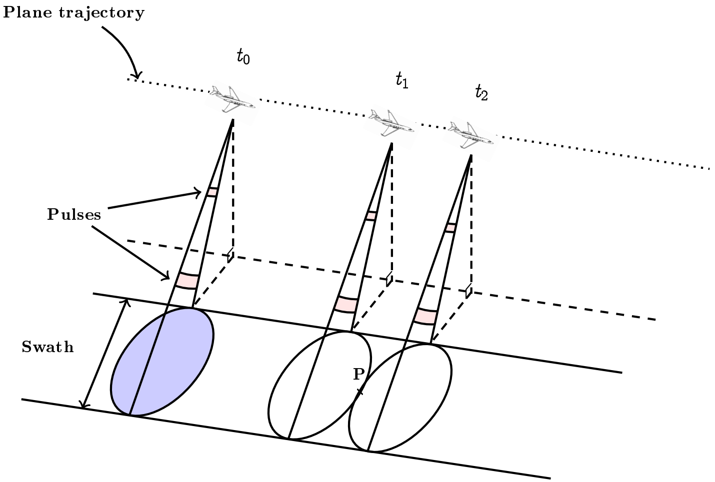
Sort pixels in a polarimetric SAR image in different groups thanks to their polarimetric properties, in an unsupervised way.



Outline

Outline

Principle of SAR



Measured signal : **k** is a complex value.

Polarimetry

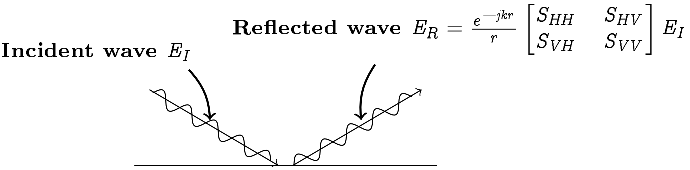


Figure : Polarimetry

- Polarization: orientation of the electric field of the EM wave
- Several possible polarizations $\Rightarrow$ horizontal and vertical
- Monostatic configuration $\rightarrow S_{HV} = S_{VH}$.
- Measured signal : $\mathbf{k} = \begin{bmatrix} S_{HH} \\ \sqrt{2} S_{HV} \\ S_{VV} \end{bmatrix}$ is a complex vector of size $m = 3$.

Random modeling of the signal

- Interferences inside the resolution cells, non-stationarity, ... $\longrightarrow$ model $\mathbf{k}$ as a random variable.
- Common assumption: $\mathbf{k} \sim \mathcal{CN}(\mathbf{0}, \mathbf{T})$
  - Low resolution
  - Large number of scatterers in each resolution cell
  - Central Limit Theorem
- In high resolution images, number of scatterers in each resolution cell smaller $\longrightarrow$ CLT not applicable.
- $\mathbf{k}$ is no longer Gaussian-distributed

Need to model the non-Gaussianity

Introduction of a non-Gaussian model.

Outline

Non-Gaussian models for SAR

Previously proposed distributions:

- K-distribution: Oliver (1984), Jao (1984), Ulaby (1986).
- $\mathcal{G}$ distribution: Frery (1997 & 2003).
- KummerU distribution: Bombrun (2008).
- Fisher distribution: Tison (2004).
- K-Wishart distribution by Doulgeris (2008).

The SIRV (Spherically Invariant Random Vectors) model

$$\mathbf{k} = \sqrt{\tau}\mathbf{x}$$

- $\mathbf{x}$ (speckle): complex circular zero-mean Gaussian m-vector
- τ (texture): positive random variable.

Used extensively in radar detection. Recently, at ONERA, PhD thesis of E. Jay (2002), F. Pascal (2006) and M. Mahot (2012) on detection and estimation with SIRV + postdoc of G. Vasile (2009) on classification.

Non-Gaussian models for SAR

Previously proposed distributions:

- K-distribution: Oliver (1984), Jao (1984), Ulaby (1986).
- $\mathcal{G}$ distribution: Frery (1997 & 2003).
- KummerU distribution: Bombrun (2008).
- Fisher distribution: Tison (2004).
- K-Wishart distribution by Doulgeris (2008).

The SIRV (Spherically Invariant Random Vectors) model

$$\mathbf{k} = \sqrt{\tau}\mathbf{x}$$

- $\mathbf{x}$ (speckle): complex circular zero-mean Gaussian m-vector
- τ (texture): positive random variable.

Used extensively in radar detection. Recently, at ONERA, PhD thesis of E. Jay (2002), F. Pascal (2006) and M. Mahot (2012) on detection and estimation with SIRV + postdoc of G. Vasile (2009) on classification.

Why choose this model ?

- Takes into account the heterogeneity of the signal thanks to the texture τ (local variations of power).
- Contains polarimetric information in $\mathbf{x}$ and $\mathbf{M} = \mathrm{E}\left[\mathbf{x}\mathbf{x}^H\right]$.
- Encompasses many different distributions: Gaussian, K distribution, Weibull, Cauchy, Student-t, Rice, etc, depending on the distribution of τ.
- Provides a strong unified framework, notably for estimation purposes: e.g. covariance matrix estimator.

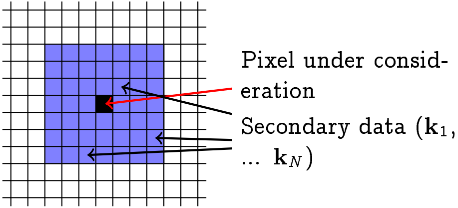
Covariance matrix

Traditionally, $\mathbf{k} \sim \mathcal{CN}(\mathbf{0}, \mathbf{T}) \rightarrow$ need the covariance matrix $\mathbf{T} = \mathrm{E}\left[\mathbf{k}\mathbf{k}^H\right]$.

Problem

$\mathbf{T}$ unknown and only one observation of $\mathbf{k}$

Estimation with neighbouring pixels.



Sample Covariance Matrix

$$\widehat{\mathbf{T}}_{SCM} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{k}_i \mathbf{k}_i^H \sim \mathcal{W}\left(\mathbf{T}, N\right)$$

Covariance matrix

Traditionally, $\mathbf{k} \sim \mathcal{CN}(\mathbf{0}, \mathbf{T}) \to$ need the covariance matrix $\mathbf{T} = \mathrm{E}\left[\mathbf{k}\mathbf{k}^H\right]$.

Problem

$\mathbf{T}$ unknown and only one observation of $\mathbf{k}$

Sample Covariance Matrix

$$\widehat{\mathbf{T}}_{SCM} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{k}_i\mathbf{k}_i^H \sim \mathcal{W}\left(\mathbf{T}, N\right)$$

- In the SIRV case, $\mathbf{k} = \sqrt{\tau}\mathbf{x}$ with $\mathbf{M} = \mathrm{E}\left[\mathbf{x}\mathbf{x}^H\right]$.
- The Sample Covariance Matrix of the SIRV covariance matrix $\mathbf{M}$:

$$\widehat{\mathbf{T}}_{SCM} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{k}_i\mathbf{k}_i^H = \frac{1}{N}\sum_{i=1}^{N} \tau_i\mathbf{x}_i\mathbf{x}_i^H \neq \frac{1}{N}\sum_{i=1}^{N} \mathbf{x}_i\mathbf{x}_i^H$$

The Fixed Point Estimator

Under SIRV assumption, the Approximate Maximum Likelihood Estimator of the covariance matrix $\mathbf{M}$ is the solution of the following equation:

$$\widehat{\mathbf{M}} = \frac{m}{N} \sum_{i=1}^{N} \frac{\mathbf{k}_i \mathbf{k}_i^H}{\mathbf{k}_i^H \widehat{\mathbf{M}}^{-1} \mathbf{k}_i} = \frac{m}{N} \sum_{i=1}^{N} \frac{\mathbf{x}_i \mathbf{x}_i^H}{\mathbf{x}_i^H \widehat{\mathbf{M}}^{-1} \mathbf{x}_i}.$$

Called the Fixed Point Estimator $\widehat{\mathbf{M}}_{FPE}$.

Depends only on the speckle part of the signal

No corruption from the heterogeneous power.

Properties of the FPE

- The solution exists and is unique, up to a scalar factor.
- It is unbiased and consistent.
- When N is large: same asymptotic behavior as $\widehat{\mathbf{M}}_{SCM}$ with a different secondary data number: N for $\widehat{\mathbf{M}}_{SCM}$, $\frac{m+1}{m} N$ for $\widehat{\mathbf{M}}_{FPE}$

Outline

Proposed framework

Many existing techniques

Wishart classifier (K-means clustering)

- Initialization: P classes with class centers $\mathbf{C}_1, ..., \mathbf{C}_P$
- Reassignment:

$$\mathbf{T} \in \Omega_k \Leftrightarrow k = \arg\min_p \left(\ln |\mathbf{C}_p| + \mathrm{Tr}\left(\mathbf{C}_p^{-1} \mathbf{T} \right) \right) \quad \text{(Wishart distance: Lee,1994)}$$

- Class center computation:

$$\mathbf{C}_k = \frac{1}{N} \sum_{\mathbf{T}_i \in \Omega_k} \mathbf{T}_i$$

Proposed framework

Many existing techniques

Wishart classifier (K-means clustering)

- Initialization: P classes with class centers $\mathbf{C}_1, ..., \mathbf{C}_P$
- Reassignment:

$$\mathbf{T} \in \Omega_k \Leftrightarrow k = \arg\min_p \left(\ln|\mathbf{C}_p| + \operatorname{Tr}\left(\mathbf{C}_p^{-1}\mathbf{T}\right)\right) \quad \text{(Wishart distance: Lee,1994)}$$

- Class center computation:

$$\mathbf{C}_k = \frac{1}{N} \sum_{\mathbf{T}_i \in \Omega_k} \mathbf{T}_i$$

Proposed framework

Many existing techniques

Wishart classifier (K-means clustering)

- Initialization: P classes with class centers $\mathbf{C}_1, ..., \mathbf{C}_P$
- Reassignment:

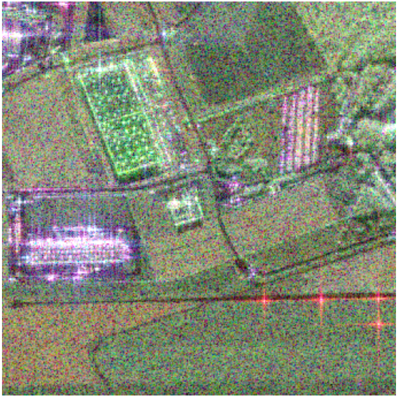
$$\mathbf{T} \in \Omega_k \Leftrightarrow k = \arg\min_p \left(\ln |\mathbf{C}_p| + \mathrm{Tr}\left(\mathbf{C}_p^{-1}\mathbf{T}\right)\right) \quad \text{(Wishart distance: Lee,1994)}$$

- Class center computation:

$$\mathbf{C}_k = \frac{1}{N} \sum_{\mathbf{T}_i \in \Omega_k} \mathbf{T}_i$$

Dataset

(a) Optical view

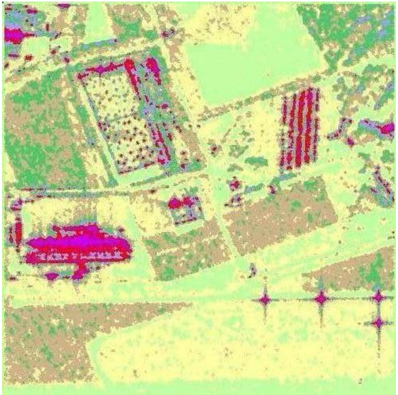


(b) RAMSES data, Pauli basis [1]

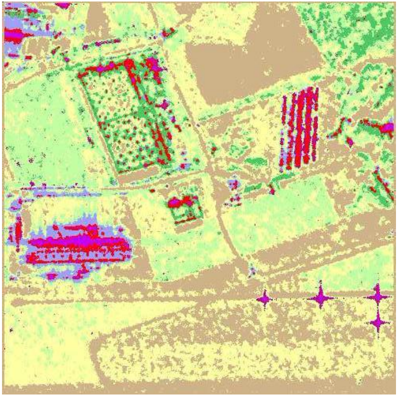
Figure : Dataset, Brétigny

[1] : $\left(\frac{S_{HH}+S_{VV}}{\sqrt{2}}, \frac{S_{HH}-S_{VV}}{\sqrt{2}}, \sqrt{2}S_{HV}\right)$

Limitation of the Gaussian assumption



(a) Using the SCM

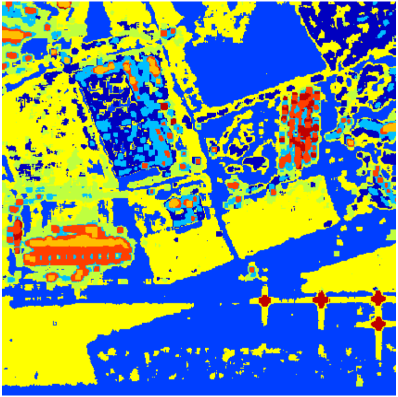


(b) Using only the intensity

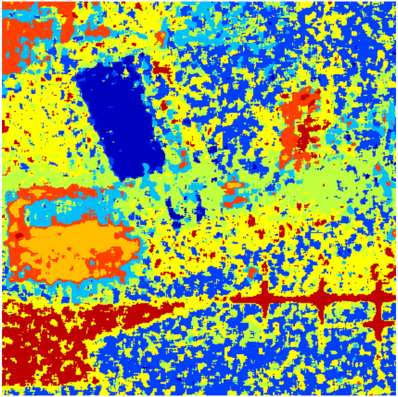
Figure : Wishart classification of the Brétigny area

⇒ same results with $\mathrm{Tr}\,(\mathbf{T})$ and $\mathbf{T}$?

Influence of the SIRV assumption



(a) Using the SCM



(b) Using the FPE

Figure : Wishart classification of the Brétigny area

$\Rightarrow$ better separation of heterogeneous areas

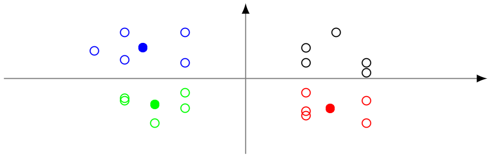
Outline

Motivations

Second step of the Wishart classifier

$$\mathbf{T} \in \Omega_k \Leftrightarrow k = \arg\min_p \left(\ln |\mathbf{C}_p| + \mathrm{Tr}\left(\mathbf{C}_p^{-1}\mathbf{T}\right)\right)$$

- No constraint on the minimum



- Difficulty finding an optimal number of classes.

Motivations

Second step of the Wishart classifier

$$\mathbf{T} \in \Omega_k \Leftrightarrow k = \arg\min_p \left(\ln |\mathbf{C}_p| + \mathrm{Tr}\left(\mathbf{C}_p^{-1} \mathbf{T} \right) \right)$$

- No constraint on the minimum
- Difficulty finding an optimal number of classes.

Proposed approach: hypothesis test

Test if an hypothesis is valid and provides a threshold for the rejection of this hypothesis.

Test construction

Goal

Compare the covariance matrices of two pixels $\mathbf{k}^{(1)}$ and $\mathbf{k}^{(2)}$.

Hypothesis test:

$$\begin{cases} H_0 : & \mathbf{T}_1 = \mathbf{T}_2 = \mathbf{T}, \\ H_1 : & \mathbf{T}_1 \neq \mathbf{T}_2, \end{cases}$$

$\mathbf{T}_1, \mathbf{T}_2, \mathbf{T}$ unknown $\Rightarrow$ estimated from $\left(\mathbf{k}_1^{(1)}, \ldots \mathbf{k}_{N_1}^{(1)}\right)$ and $\left(\mathbf{k}_1^{(2)}, \ldots \mathbf{k}_{N_2}^{(2)}\right)$

Generalized Likelihood Ratio Test

$$\Lambda = \frac{\sup_\theta L(\mathbf{k}; H_1, \theta)}{\sup_\theta L(\mathbf{k}; H_0, \theta)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta, \quad \text{where } L(\mathbf{k}; H, \theta) = \prod_i f\left(\mathbf{k}_i | H, \theta\right).$$

Case $\mathbf{T}_2$ known

GLRT

$$\ln(\Lambda) = N_1 \left(\ln |\mathbf{T}_2| - \ln \left| \widehat{\mathbf{T}}_1 \right| + \operatorname{Tr}\left(\mathbf{T}_2^{-1} \widehat{\mathbf{T}}_1 \right) - m \right)$$

For both SCM and FPE

$$\ln(\Lambda) = d(\widehat{\mathbf{T}}_1, \mathbf{T}_2) = \left(\ln |\mathbf{T}_2| + \operatorname{Tr}\left(\mathbf{T}_2^{-1} \widehat{\mathbf{T}}_1 \right) \right) \Rightarrow \text{Wishart distance}$$

$\Rightarrow$ Generalization of the Wishart distance

Case where both matrices are unknown

GLRT

$$\Lambda = \frac{\left|\widehat{\mathbf{T}}\right|^{N_1+N_2}}{\left|\widehat{\mathbf{T}}_1\right|^{N_1}\left|\widehat{\mathbf{T}}_2\right|^{N_2}} \exp\left(\mathrm{Tr}\left(\widehat{\mathbf{T}}^{-1}\left[N_1\widehat{\mathbf{T}}_1 + N_2\widehat{\mathbf{T}}_2\right]\right) - (N_1+N_2)\,m\right)$$

- SCM case:

$$\widehat{\mathbf{T}} = \frac{N_1\widehat{\mathbf{T}}_1 + N_2\widehat{\mathbf{T}}_2}{N_1+N_2} \Rightarrow \Lambda = \frac{\left|\widehat{\mathbf{T}}\right|^{N_1+N_2}}{\left|\widehat{\mathbf{T}}_1\right|^{N_1}\left|\widehat{\mathbf{T}}_2\right|^{N_2}}$$

- FPE case:

$$\widehat{\mathbf{T}} = f\left(\widehat{\mathbf{T}}_1, \widehat{\mathbf{T}}_2\right) \;???$$

Box's M-test (Gaussian case)

Bartlett's distance (1937)

$$\Lambda_{Bar} = \frac{\left|\widehat{\mathbf{T}}_1\right|^{\frac{\nu_1}{2}} \left|\widehat{\mathbf{T}}_2\right|^{\frac{\nu_2}{2}}}{\left|\widehat{\mathbf{T}}\right|^{\frac{\nu}{2}}}$$

where $\nu_i = N_i$ and $\nu = N_1 + N_2$ are the degrees of freedom of the estimation of $\widehat{\mathbf{T}}_i$ and $\widehat{\mathbf{T}}$, respectively.

Box's χ^2 approximation (1949)

$$\Lambda_{Box} = -2(1 - c_1)\ln(\Lambda_{Bar}) \sim \chi^2\left(\frac{1}{2}m(m+1)\right)$$

where $c_1 = \left(\sum_{i=1}^{2}\frac{1}{\nu_i} - \frac{1}{\sum_{i=1}^{2}\nu_i}\right)\left(\frac{2m^2+3m-1}{6(m+1)}\right)$.

Box's M-test (SIRV case)

Asymptotic property of the FPE: same asymptotic behavior as $\widehat{\mathbf{M}}_{SCM}$ with a different secondary data number: N for $\widehat{\mathbf{M}}_{SCM}$, $\dfrac{m+1}{m}N$ for $\widehat{\mathbf{M}}_{FPE}$

Box's χ^2 approximation for the SIRV case

$$\Lambda_{Box} = -2(1-c_1)\ln(\Lambda'_{Bar}) \sim \chi^2\left(\frac{1}{2}m(m+1)\right)$$

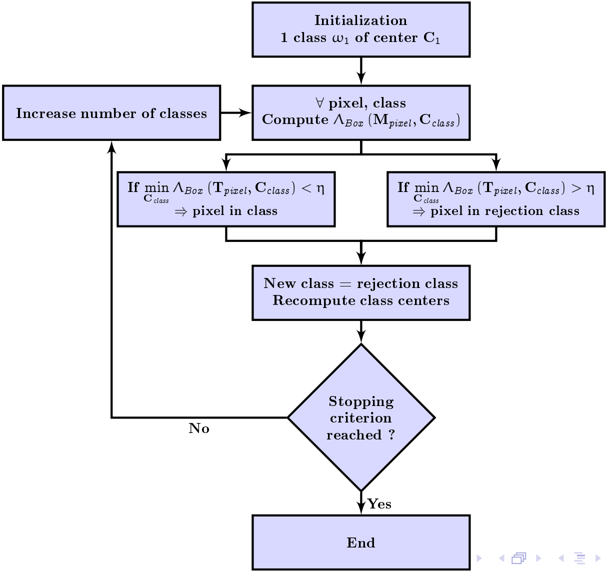
Difference from Gaussian case

$$\nu_i = \frac{m}{m+1}N_i \text{ and } \nu = \frac{m}{m+1}(N_1+N_2).$$

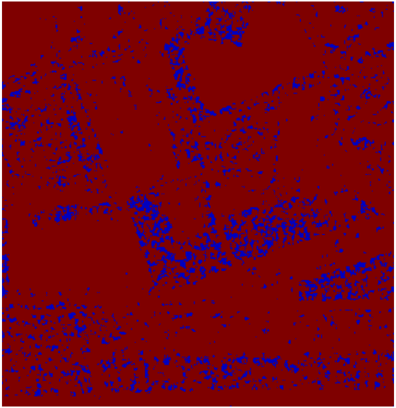
Critical region

$$\Lambda_{Box} \underset{H_0}{\overset{H_1}{\gtrless}} \eta \Rightarrow C_r = \left\{\Lambda_{Box},\ \Lambda_{Box} > \eta = \chi^2_{P_{FA}}\left(\frac{1}{2}m(m+1)\right)\right\}$$

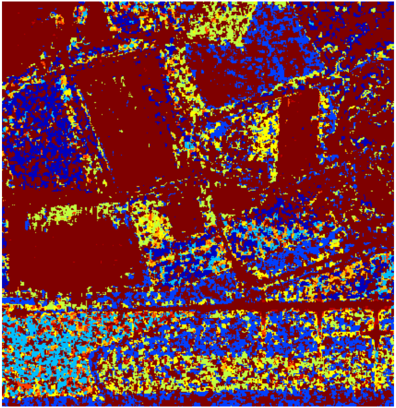
Naive implementation



Naive implementation, classification results using the SCM



(a) 1 iteration

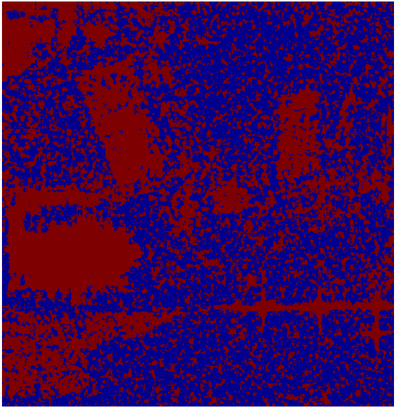


(b) 8 iterations

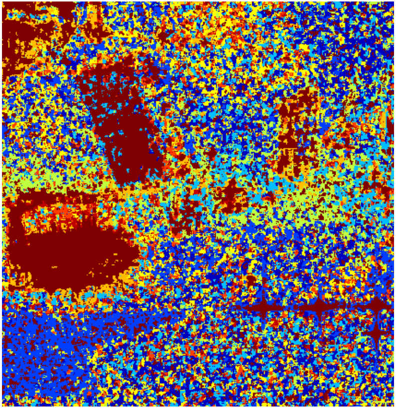
Figure : Classification results with SCM

Naive implementation, classification results using the FPE

P. Formont, F. Pascal, G. Vasile, J.-P. Ovarlez and L. Ferro-Famil, “Statistical Classification for Heterogeneous Polarimetric SAR Images”, *IEEE JSTSP*, 2011.



(a) 1 iteration



(b) 8 iterations

Application to hierarchical clustering

- Hierarchical segmentation: Beaulieu and Touzi(2004).
- Salembier and Alonso-Gonzalez (since 2010).

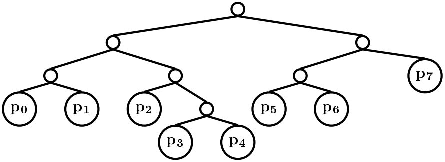


Figure : Hierarchical clustering

- Each pixel initially in its own class (leaf).
- At each iteration, merge closest pixels w.r.t. Λ_{Box}.
- Define a linkage function to merge clusters of pixels:
  - minimum distance
  - maximum distance
  - average distance
- Cut the tree at height given by the threshold η.

Application to hierarchical clustering

- Hierarchical segmentation: Beaulieu and Touzi(2004).
- Salembier and Alonso-Gonzalez (since 2010).

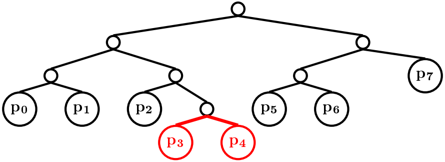


Figure : Hierarchical clustering

- Each pixel initially in its own class (leaf).
- At each iteration, merge closest pixels w.r.t. Λ_{Box}.
- Define a linkage function to merge clusters of pixels:
  - minimum distance
  - maximum distance
  - average distance
- Cut the tree at height given by the threshold η.

Application to hierarchical clustering

- Hierarchical segmentation: Beaulieu and Touzi(2004).
- Salembier and Alonso-Gonzalez (since 2010).

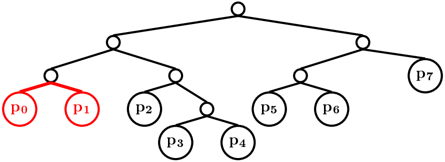


Figure : Hierarchical clustering

- Each pixel initially in its own class (leaf).
- At each iteration, merge closest pixels w.r.t. Λ_{Box}.
- Define a linkage function to merge clusters of pixels:
  - minimum distance
  - maximum distance
  - average distance
- Cut the tree at height given by the threshold η.

Application to hierarchical clustering

- Hierarchical segmentation: Beaulieu and Touzi(2004).
- Salembier and Alonso-Gonzalez (since 2010).

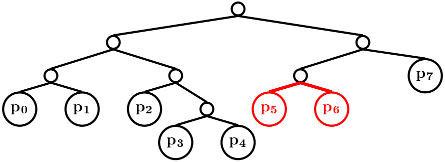


Figure : Hierarchical clustering

- Each pixel initially in its own class (leaf).
- At each iteration, merge closest pixels w.r.t. Λ_{Box}.
- Define a linkage function to merge clusters of pixels:
  - minimum distance
  - maximum distance
  - average distance
- Cut the tree at height given by the threshold η.

Application to hierarchical clustering

- Hierarchical segmentation: Beaulieu and Touzi(2004).
- Salembier and Alonso-Gonzalez (since 2010).

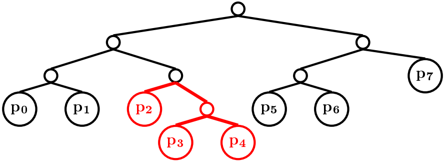


Figure : Hierarchical clustering

- Each pixel initially in its own class (leaf).
- At each iteration, merge closest pixels w.r.t. Λ_{Box}.
- Define a linkage function to merge clusters of pixels:
  - minimum distance
  - maximum distance
  - average distance
- Cut the tree at height given by the threshold η.

Application to hierarchical clustering

- Hierarchical segmentation: Beaulieu and Touzi(2004).
- Salembier and Alonso-Gonzalez (since 2010).

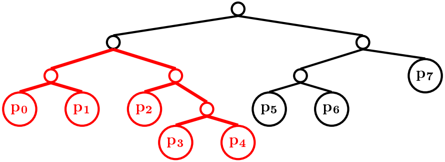


Figure : Hierarchical clustering

- Each pixel initially in its own class (leaf).
- At each iteration, merge closest pixels w.r.t. Λ_{Box}.
- Define a linkage function to merge clusters of pixels:
  - minimum distance
  - maximum distance
  - average distance
- Cut the tree at height given by the threshold η.

Application to hierarchical clustering

- Hierarchical segmentation: Beaulieu and Touzi(2004).
- Salembier and Alonso-Gonzalez (since 2010).

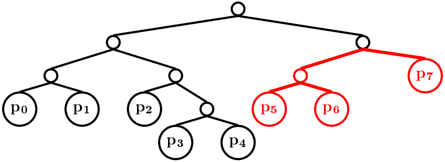


Figure : Hierarchical clustering

- Each pixel initially in its own class (leaf).
- At each iteration, merge closest pixels w.r.t. Λ_{Box}.
- Define a linkage function to merge clusters of pixels:
  - minimum distance
  - maximum distance
  - average distance
- Cut the tree at height given by the threshold η.

Application to hierarchical clustering

- Hierarchical segmentation: Beaulieu and Touzi(2004).
- Salembier and Alonso-Gonzalez (since 2010).

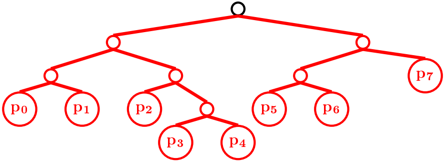


Figure : Hierarchical clustering

- Each pixel initially in its own class (leaf).
- At each iteration, merge closest pixels w.r.t. Λ_{Box}.
- Define a linkage function to merge clusters of pixels:
  - minimum distance
  - maximum distance
  - average distance
- Cut the tree at height given by the threshold η.

Application to hierarchical clustering

- Hierarchical segmentation: Beaulieu and Touzi(2004).
- Salembier and Alonso-Gonzalez (since 2010).

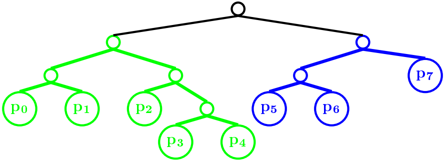


Figure : Hierarchical clustering

- Each pixel initially in its own class (leaf).
- At each iteration, merge closest pixels w.r.t. Λ_{Box}.
- Define a linkage function to merge clusters of pixels:
  - minimum distance
  - maximum distance
  - average distance
- Cut the tree at height given by the threshold η.

Application to hierarchical clustering

- Hierarchical segmentation: Beaulieu and Touzi(2004).
- Salembier and Alonso-Gonzalez (since 2010).

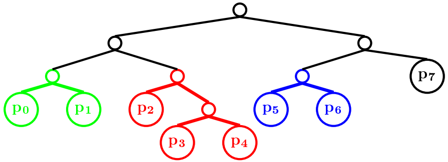
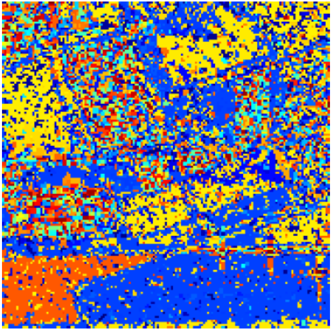


Figure : Hierarchical clustering

- Each pixel initially in its own class (leaf).
- At each iteration, merge closest pixels w.r.t. Λ_{Box}.
- Define a linkage function to merge clusters of pixels:
  - minimum distance
  - maximum distance
  - average distance
- Cut the tree at height given by the threshold η.

Average distance


(a) SCM

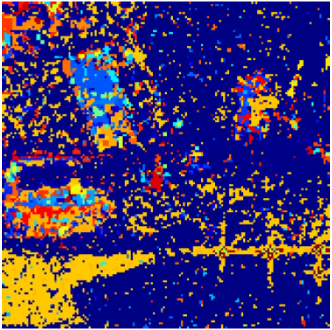

(b) FPE

Figure : Hierachical clustering results with average distance and $P_{FA} = 10^{-4}$

Pierre Formont, Miguel Angel Veganzones, Joana Maria Frontera-Pons, Frédéric Pascal, Jean-Philippe Ovarlez and Jocelyn Chanussot, "CFAR Hierarchical Clustering of Polarimetric SAR Data", *IEEE 2013 International Geoscience and Remote Sensing Symposium (IGARSS)*, Melbourne, Australia, July 21 — 26, 2013.

Outline

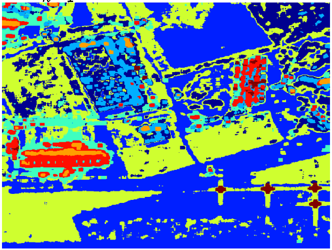
Motivations

Third step of the Wishart classifier

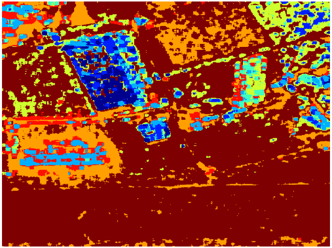
$$\mathbf{C}_k = \frac{1}{N} \sum_{\mathbf{T}_i \in \Omega_k} \mathbf{T}_i$$

Use the pixels of the class directly ?

$$\mathbf{C}_k = \frac{1}{N_k} \sum_{n=1}^{N_k} \mathbf{k}_n \mathbf{k}_n^H$$



(a) Arithmetical mean



(b) Estimation

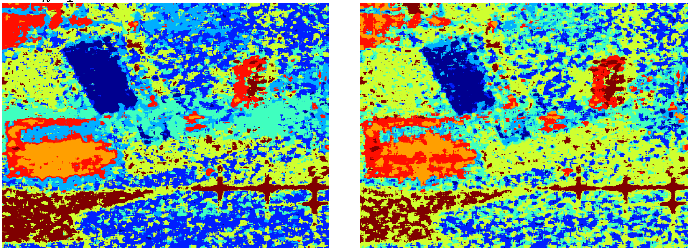
Motivations

Third step of the Wishart classifier

$$\mathbf{C}_k = \frac{1}{N} \sum_{\mathbf{T}_i \in \Omega_k} \mathbf{T}_i$$

Use the pixels of the class directly ?

$$\mathbf{C}_k = \frac{1}{N_k} \sum_{n=1}^{N_k} \mathbf{k}_n \mathbf{k}_n^H$$



(a) Arithmetical mean (b) Estimation

Structure of covariance matrices

- Another way to look at the problem: consider the structure of the manipulated objects (covariance matrices) $\Rightarrow$ Hermitian definite-positive matrices.
- NOT Euclidean space: arithmetical mean not adapted to this space.

Euclidean mean (arithmetic)

$$\underset{\mathbf{M}\in\mathcal{P}(m)}{\arg\min} \sum_{i=1}^{N} d(\mathbf{M},\mathbf{M}_i)^2, \text{ where } d(\mathbf{M},\mathbf{M}_i) = \|\mathbf{M}-\mathbf{M}_i\|_F$$

Riemannian mean (geometric)

$$\underset{\mathbf{M}\in\mathcal{P}(m)}{\arg\min} \sum_{i=1}^{N} d(\mathbf{M},\mathbf{M}_i)^2, \text{ where } d(\mathbf{M},\mathbf{M}_i) = ?$$

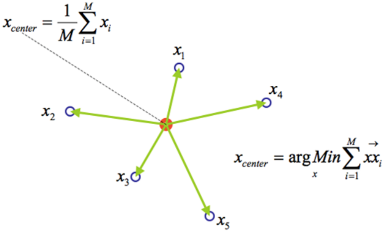
Structure of covariance matrices

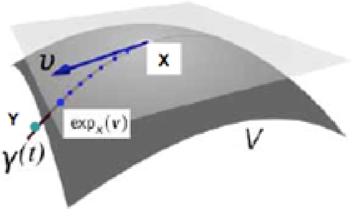
Euclidean mean (arithmetic)

$$d(\mathbf{M}, \mathbf{M}_i) = \|\mathbf{M} - \mathbf{M}_i\|_F$$

Riemannian mean (geometric)

$$d(\mathbf{M}, \mathbf{M}_i) = ?$$





Mean of Hermitian definite positive matrices

Riemannian distance between two matrices

$$d(\mathbf{M}_1, \mathbf{M}_2)^2 = \left\| \log\left(\left(\mathbf{M}_1^{-1/2}\right)^H \mathbf{M}_2 \mathbf{M}_1^{-1/2} \right) \right\|_F^2$$

More convenient expression

$$d\left(\mathbf{M}_1, \mathbf{M}_2\right) = \left[\sum_{k=1}^{n} \left(\log \lambda_k\right)^2 \right]^{1/2}$$

No analytical expression for M !

$$\sum_{i=1}^{N} \log\left(\mathbf{M}_i^{-1}\mathbf{M}\right) = 0.$$

Gradient descent algorithm

$$\mathbf{M}_{n+1} = \left(\mathbf{M}_n^{1/2}\right)^H \exp\left(-\epsilon \sum_{i=1}^{N} \log\left(\left(\mathbf{M}_n^{-1/2}\right)^H \mathbf{M}_i^{-1} \mathbf{M}_n^{-1/2} \right) \right) \mathbf{M}_n^{1/2}$$

Recent work

- Moakher (2005) proposed a differential approach to compute the mean of symmetric positive-definite matrices.
- Devlaminck (2010) demonstrated the added physical interpretation of a Riemannian mean for the covariance matrices in polarized light.
- Wang (2010) used Riemannian geometry for PolSAR classification using the mean-shift algorithm.
- Barbaresco (2010) proposed different approaches for the computation of the mean of Hermitian definite positive matrices and applications to radar signal processing, especially STAP processing.

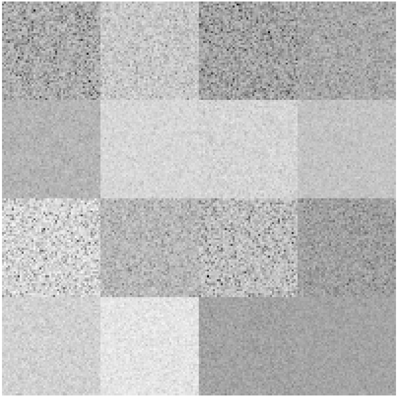
Simulated data

Figure : Extraction of covariance matrices

Simulated data

M_1, λ_1	M_1, λ_2	M_2, λ_1	M_2, λ_2
M_1, λ_3	M_1, λ_4	M_2, λ_3	M_2, λ_4
M_3, λ_1	M_3, λ_2	M_4, λ_1	M_4, λ_2
M_3, λ_3	M_3, λ_4	M_4, λ_3	M_4, λ_4

(a) K-distributed data



(b) Power

Figure : Simulated data

Classification scheme

K-means clustering with 4 classes:

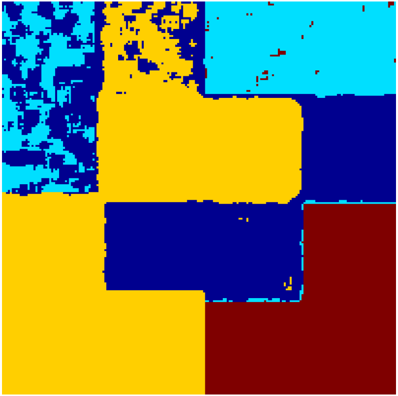
- Choice of Wishart distance or Riemannian distance
- Choice of Euclidean mean or Riemannian mean
- Choice of SCM or FPE
- Choice of supervised case (initial class centers are generating matrices $\mathbf{M}_1$, ..., $\mathbf{M}_4$) or unsupervised case (initial class centers are estimated through random initialization of the data).

Pierre Formont, Jean-Philippe Ovarlez and Frédéric Pascal, "On the use of Matrix Information Geometry for Polarimetric SAR Image Classification", *Matrix Information Geometry*, Springer, pp. 257 — 276, 2013.

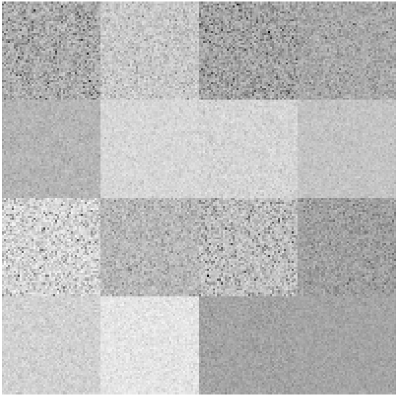
Classification results, simulated data

All cases

SCM polluted by power.



(a) SCM

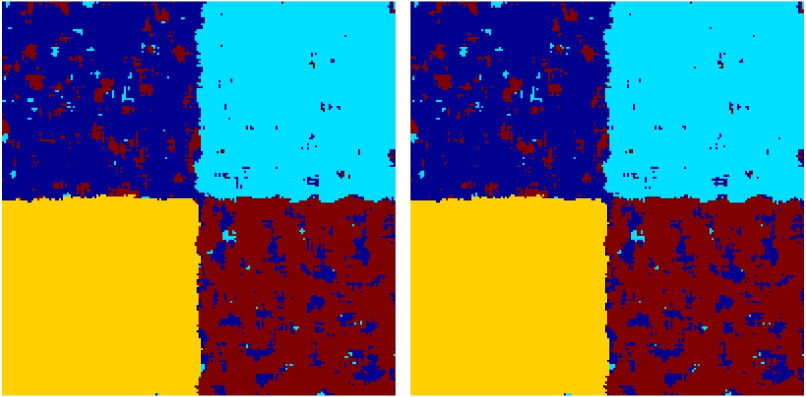


(b) Power

Classification results, simulated data

All cases (FPE, Euclidean mean)

Little difference between Wishart distance and Riemannian distance

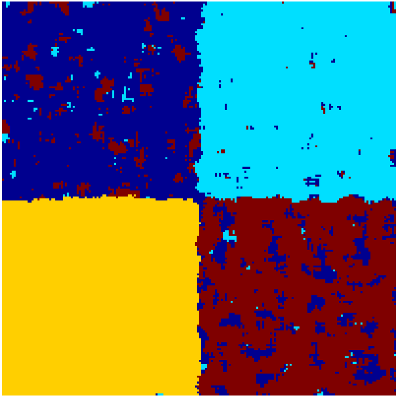


(a) Riemannian distance (b) Wishart distance

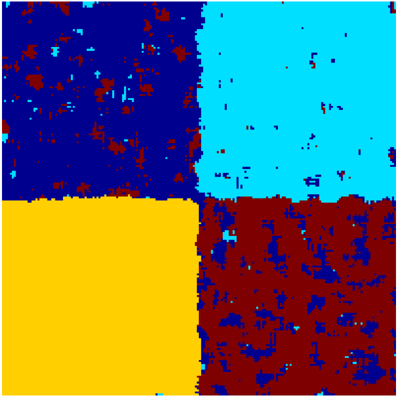
Classification results, simulated data

Supervised case (FPE, Wishart distance)

Little difference between Euclidean mean and Riemannian mean



(a) Euclidean mean

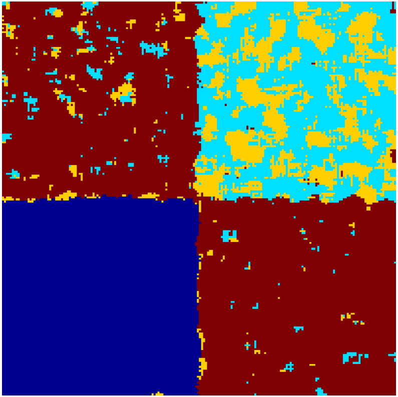


(b) Riemannian mean

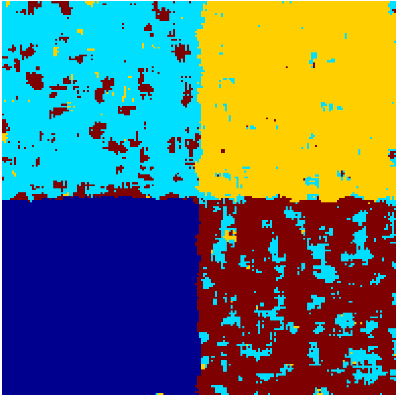
Classification results, simulated data

Unsupervised case (FPE, Wishart distance)

Riemannian mean can perform better when matrices are not known



(a) Euclidean mean



(b) Riemannian mean

Classification scheme for real data

K-means clustering with 8 classes:

- Choice of Wishart distance or Riemannian distance
- Choice of Euclidean mean or Riemannian mean
- Fixed Point Estimator
- Choice of Cloude-Pottier initialization or random initialization.

Cloude-Pottier decomposition

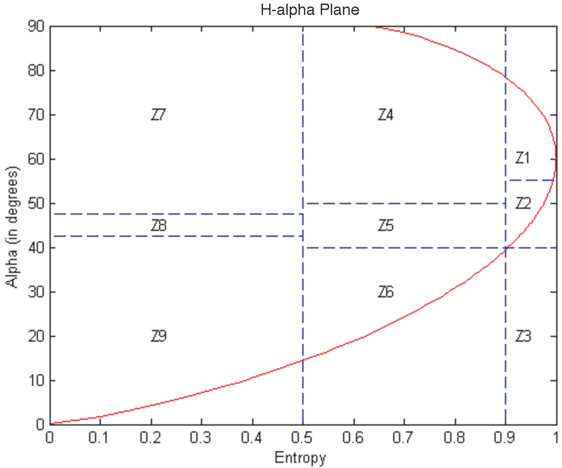
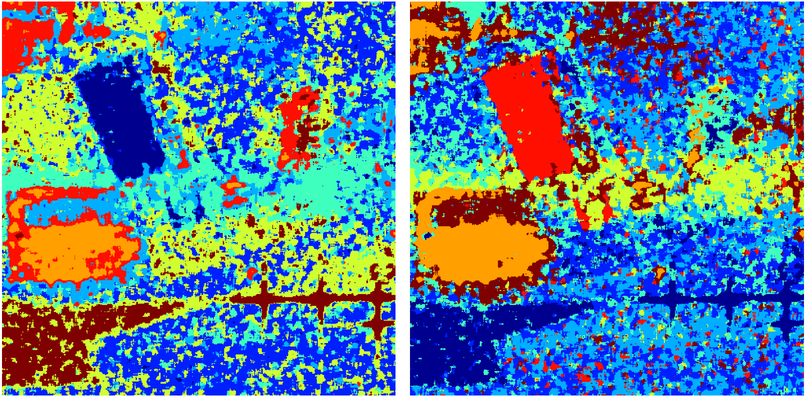


Figure : Entropy - α plane

Classification results, real data

All cases

Little difference between Cloude-Pottier and random initialization.



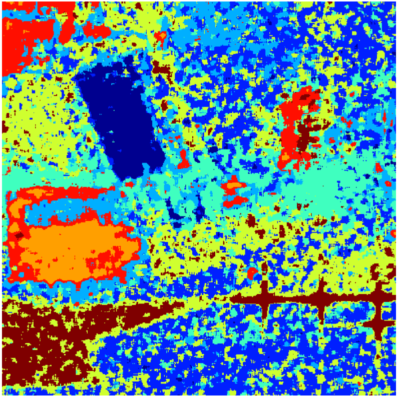
(a) Cloude-Pottier (b) Random

Figure : Euclidean mean, Wishart distance

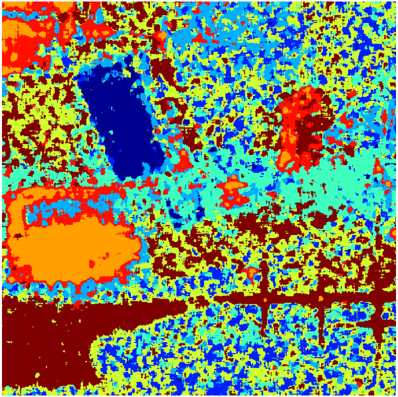
Classification results, real data

All cases

Impact of Riemannian distance difficult to quantify.



(a) Wishart distance



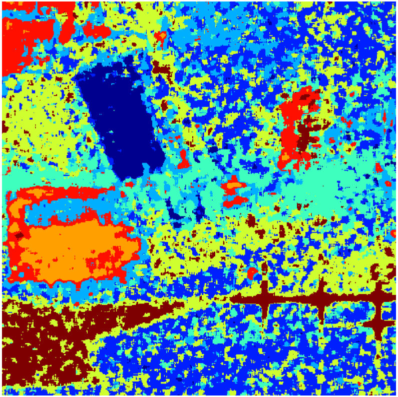
(b) Riemannian distance

Figure : Euclidean mean, Cloude-Pottier initialization

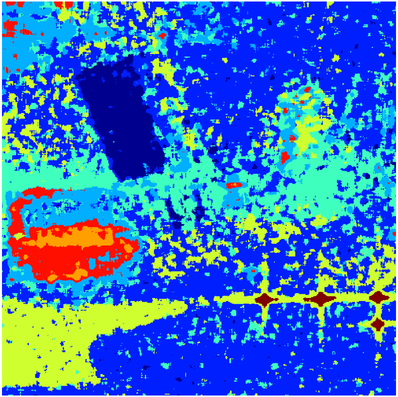
Classification results, real data

Impact of Riemannian mean

Separates some features



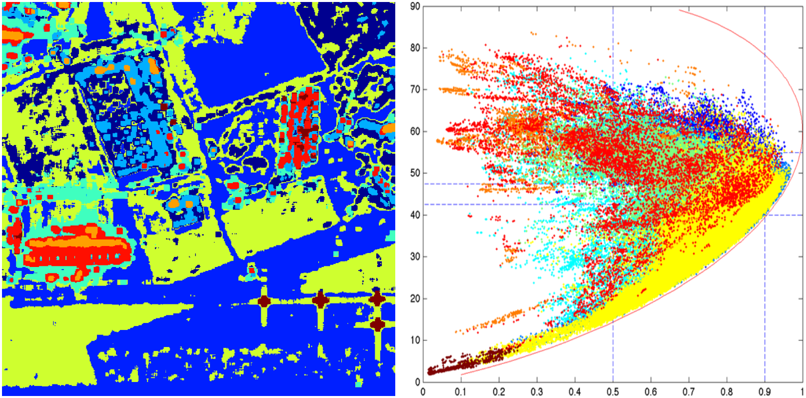
(a) Euclidean mean



(b) Riemannian mean

Figure : Wishart distance, Cloude-Pottier initialization

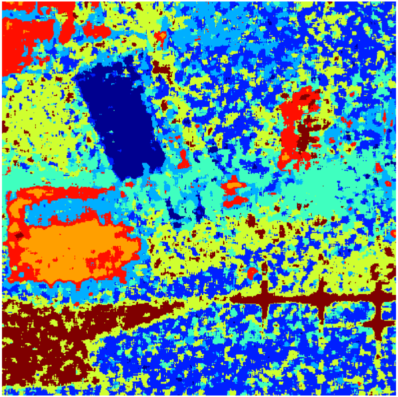
Repartition in the $H - \alpha$ plane



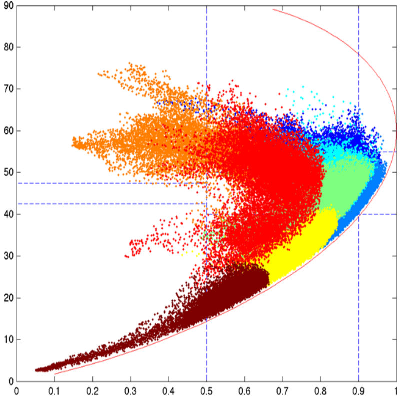
(a) Classification results (b) Repartition

Figure : SCM, Euclidean mean, Wishart distance

Repartition in the $H - \alpha$ plane



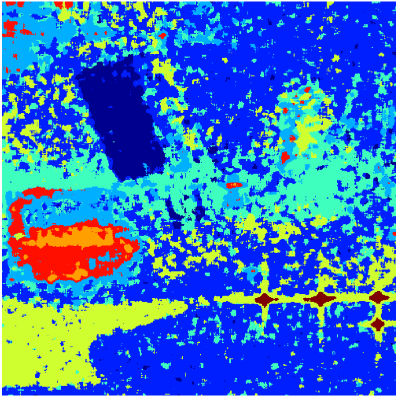
(a) Classification results



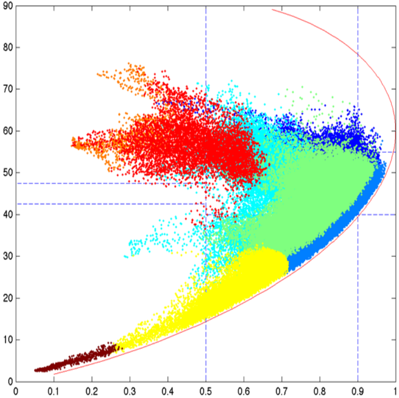
(b) Repartition

Figure : FPE, Euclidean mean, Wishart distance

Repartition in the $H - \alpha$ plane



(a) Classification results



(b) Repartition

Figure : FPE, Riemannian mean, Wishart distance

Outline

Modeling

- Introduction of a non-Gaussian model for polarimetric SAR classification
- Limitations of the traditional Gaussian approach
- Unification of previous work
- Application on real data
- Increase interest of polarimetry

Statistical approach

- Original approach to the classification problem through hypothesis test
- Generalization of the traditional Wishart distance for the SIRV model
- Introduction of a rejection class
- Development of new algorithms and application on real data

Application of information geometry

- Introduction of tools for computation of mean of polarimetric covariance matrices
- Study impact on simulated data
- Application on real data

Perspectives

- Validation of all these techniques: physical interpretation
- Texture can provide polarimetric information jointly with covariance matrix
- Validity of the SIRV model: single texture for all polarisations ?
- Local estimation of the covariance matrix
- Application to other data: hyperspectral, ...

Thanks for your attention