



Robust statistics and Riemannian Geometry for Multidimensional SAR Time Series

Ammar Mian

Presented at: Telecom ParisTech

Date: 19 June 2020



Aalto University



Contenu de la présentation

1 Robust Change Detection for SAR Images

- Motivations
- Statistical framework
- Gaussian Assumption
- Extension to robust models

2 Riemannian Geometry and Applications

- Motivations
- Introduction to Riemannian Geometry
- Adapting Euclidean Learning algorithms to Riemannian manifolds
- Applications

Miscellaneous

■ Works presented in this presentation:

- **A. Mian**, G. Ginolhac, J. Ovarlez and A. M. Atto, “New Robust Statistics for Change Detection in Time Series of Multivariate SAR Images”, in *IEEE Transactions on Signal Processing*, vol. 67, no. 2, pp. 520-534, 15 Jan, 2019.
- **A. Mian**, “Contribution to SAR Image Time Series Analysis”, Thesis, Université Paris-Saclay, Sep. 2019.
- **A. Mian**, E. Raninen, E. Ollila, “A Comparative Study of Supervised Learning Algorithms for Symmetric Positive Definite Features,” in *IEEE 28th European Signal Processing Conference (EUSIPCO)*

■ Papers and codes available at:

<https://ammarmian.github.io/>

Contents of the presentation

1 Robust Change Detection for SAR Images

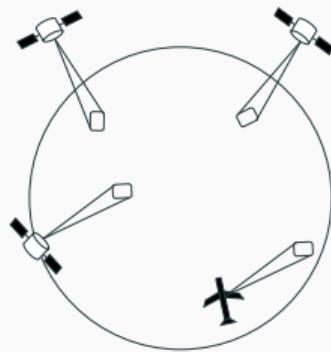
- Motivations
- Statistical framework
- Gaussian Assumption
- Extension to robust models

2 Riemannian Geometry and Applications

- Motivations
- Introduction to Riemannian Geometry
- Adapting Euclidean Learning algorithms to Riemannian manifolds
- Applications

Remote Sensing: big data analysis

Remote sensing allows to obtain image of the Earth's surface for various applications such as **Change Detection** .



Huge increase in the number of available acquisitions:

- Sentinel-1: 12 days repeat cycle, since 2014
 - TerraSAR-X: 11 days repeat cycle, since 2007
 - UAVSAR, ... thousands of flight paths planned

Problem

→ There is a need to process the huge amount of data automatically!

SAR Image Time Series: changes analysis

Change detection is useful for various purposes: Activity monitoring

Figure 1: Terrasar-X images of the Burning-man festival

SAR Image Time Series: changes analysis

Change detection is useful for various purposes: Disaster assessment

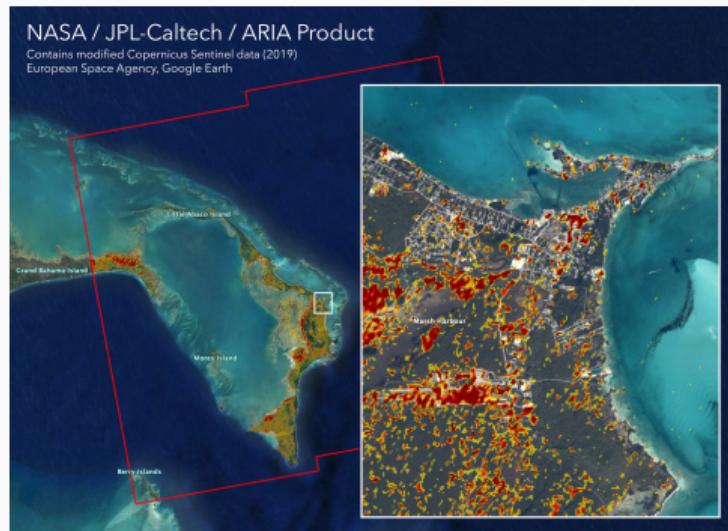
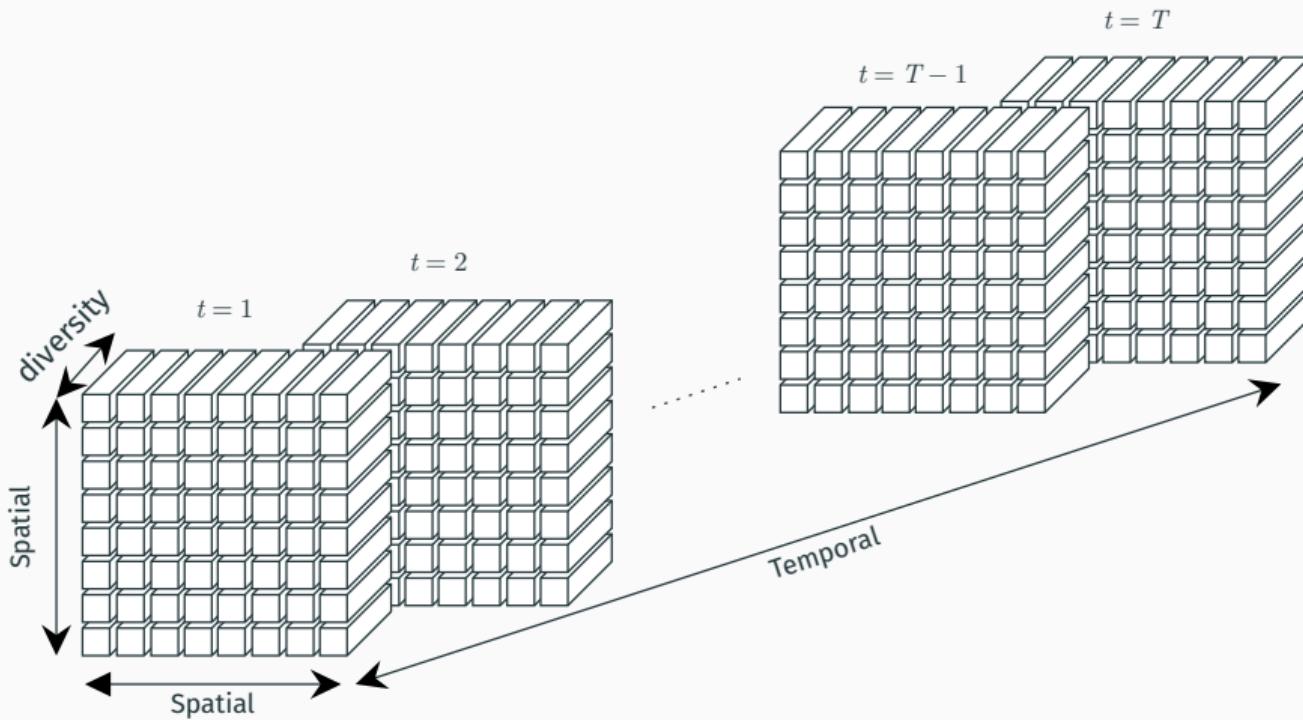
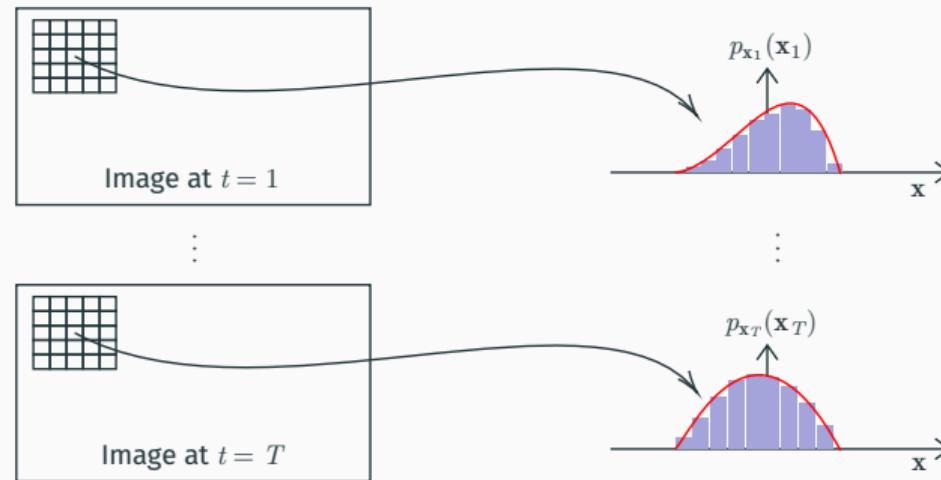


Figure 1: Destruction map of Dorian Hurricane in the Bahamas using Change Detection over Sentinel-1 data

Data dimensionality



Statistical framework: principle



Interest of this approach

- Can account for physical modelling of the data/noise
 - Strong theoretical guarantees from statistical litterature

SAR image time series representation

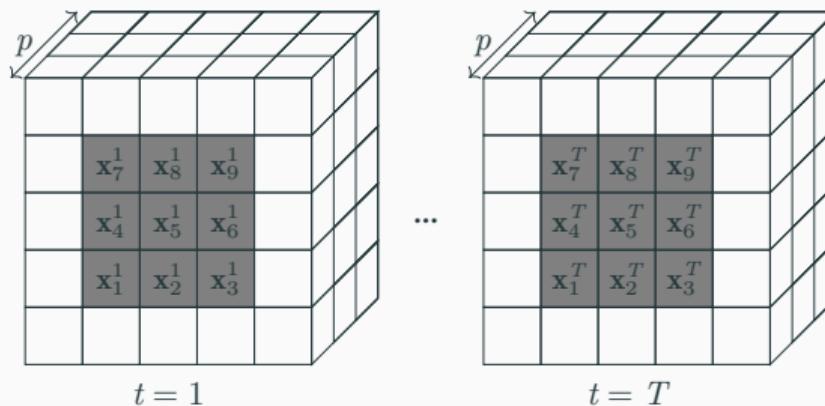


Figure 2: Sliding windows $\mathbb{W}_{1,T}$

p	N	T
dimension of vectors	number of pixels on local window	number of dates in the time series

Parametric change detection

A probability model is assigned to the observations on the windows over time:

$$\mathbf{x}_k^t \sim p_{\mathbf{x}_k^t}(\mathbf{x}_k^t; \boldsymbol{\theta}_t; \boldsymbol{\Phi}_t).$$

The detection is done on some *parameters of interest* $\boldsymbol{\theta}_t$ while the remaining ones are the *nuisance parameters* $\boldsymbol{\Phi}_t$:

$$\begin{cases} H_0 : \boldsymbol{\theta}_1 = \dots = \boldsymbol{\theta}_T = \boldsymbol{\theta}_0, \\ H_1 : \exists (t, t') \in \llbracket 1, T \rrbracket^2, \boldsymbol{\theta}_t \neq \boldsymbol{\theta}_{t'} \end{cases} . \quad (1)$$

Problems

- Specify a model and parameters of interest which are
 - A good fit to the empirical distribution
 - Robust to a large class of distributions and outliers
- Find a test statistic to obtain a rule of decision between the two alternatives.

Test statistic

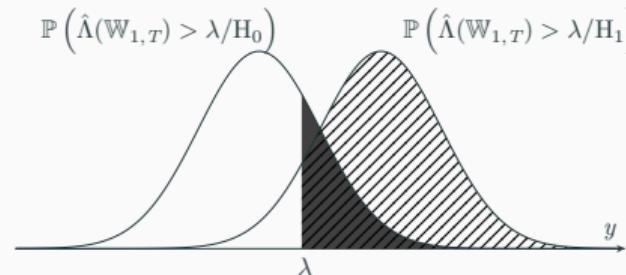
We want to obtain:

- a statistic of decision $\hat{\Lambda}$:
- a threshold λ

$$\begin{aligned}\mathbb{C}^{p \times N} \times \cdots \times \mathbb{C}^{p \times N} &\rightarrow \mathbb{R} \\ \mathbb{W}_{1,T} &\rightarrow \hat{\Lambda}(\mathbb{W}_{1,T})\end{aligned}$$

So that

$\mathbb{P}(\hat{\Lambda}(\mathbb{W}_{1,T}) > \lambda/H_1)$ is high while $\mathbb{P}(\hat{\Lambda}(\mathbb{W}_{1,T}) > \lambda/H_0)$ is low.



Constant False Alarm Rate (CFAR) property

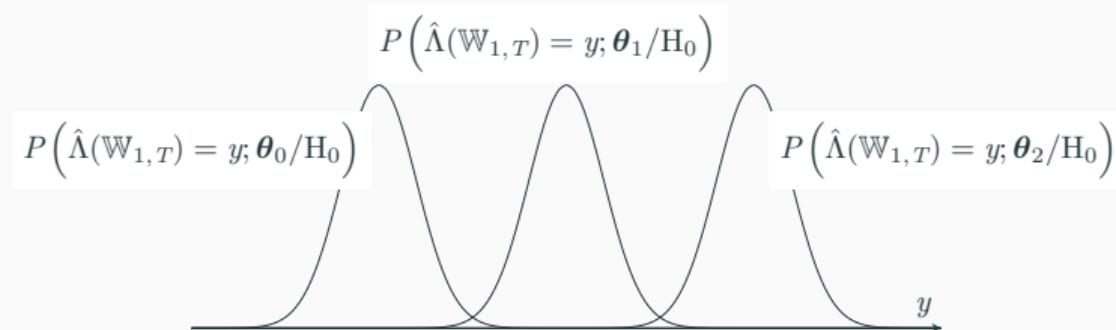
Assume a parametric model: $\forall k, \forall t, \mathbf{x}_k^t \sim p_{\mathbf{x};\boldsymbol{\theta}}(\mathbf{x}; \boldsymbol{\theta})$.

Definition

A statistic $\hat{\Lambda}$ is said to be CFAR if for any set $(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1)$, we have:

$$\mathbb{P}(\hat{\Lambda}(\mathbb{W}_{1,T}; \boldsymbol{\theta}_0/H_0) = x) = \mathbb{P}(\hat{\Lambda}(\mathbb{W}_{1,T}; \boldsymbol{\theta}_1/H_0) = x)$$

Example of a non CFAR statistic:



Generalized Likelihood Ratio Test

Given the change detection decision problem, the GLRT is formulated as follows:

$$\hat{\Lambda} = \frac{\max_{\theta_1, \dots, \theta_T, \Phi_1, \dots, \Phi_T} p_{\mathbb{W}_{1,T}}(\mathbb{W}_{1,T}/H_1; \theta_1, \dots, \theta_T, \Phi_1, \dots, \Phi_T)}{\max_{\theta_0, \Phi_1, \dots, \Phi_T} p_{\mathbb{W}_{1,T}}(\mathbb{W}_{1,T}/H_0; \theta_0, \Phi_1, \dots, \Phi_T)} \stackrel{H_1}{\underset{H_0}{\gtrless}} \lambda. \quad (2)$$

→ Good invariance properties [Kay and Gabriel, 2003].

Gaussian modelling

Definition

A vector $\mathbf{x} \in \mathbb{C}^p$ is said to Gaussian distributed with mean parameter $\boldsymbol{\mu} \in \mathbb{C}^p$ and covariance parameter $\boldsymbol{\Sigma} \in \mathbb{S}_{\mathbb{H}}^p$, denoted $\mathbf{x} \sim \mathbb{C}\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ if the Probability Distribution Function (PDF) of its distribution is the following:

$$p_{\mathbf{x}}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\pi^p |\boldsymbol{\Sigma}|^{-1}} \exp \left\{ -(\mathbf{x} - \boldsymbol{\mu})^H \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\}. \quad (3)$$

Introduced by [Conradsen et al., 2003] which is a reference in the domain. The data is modelled by a Gaussian model as follows:

$$\mathbf{x}_k^t \sim \mathbb{C}\mathcal{N}(\mathbf{0}_p, \boldsymbol{\Sigma}_t).$$

Detection test: $\boldsymbol{\theta}_t = \boldsymbol{\Sigma}_t \quad \& \quad \Phi_t = \emptyset$

Prior works in Gaussian context

We have the following test:¹:

GLRT for covariance homogeneity test in Gaussian context

$$\hat{\Lambda}_G = \frac{\left| \frac{1}{T} \sum_{t=1}^T \hat{\Sigma}_t \right|^{TN}}{\prod_{t=1}^T \left| \hat{\Sigma}_t \right|^N} \stackrel{H_1}{\gtrless} \lambda \quad (4)$$

where $\forall t$, $\hat{\Sigma}_t = \frac{1}{N} \sum_{k=1}^N \mathbf{x}_k^t (\mathbf{x}_k^t)^H$.

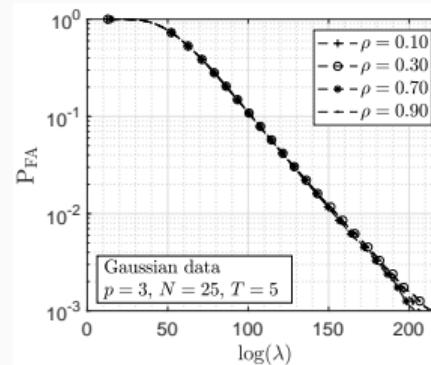
¹Many other statistics based on other principles exist as described in [Ciuonzo et al., 2017].

Some properties of the statistic i

CFARness property

The GLRT statistic is CFAR towards the covariance parameter.

In simulation: $\mathbf{x}_k^t \sim \mathbb{C}\mathcal{N}(\mathbf{0}_p, (\rho^{|i-j|})_{ij})$ with 10^5 Monte-Carlo trials.



Some properties of the statistic ii

False alarm/threshold relationship [Anderson, 2003]

Under H_0 hypothesis:

$$P\left\{2\rho \log(\hat{\Lambda}_G) \leq z\right\} \approx P\left\{\chi^2(f^2) \leq z\right\} + \omega_2 \left[P\left\{\chi^2(f^2 + 4) \leq z\right\} - P\left\{\chi^2(f^2) \leq z\right\} \right]$$

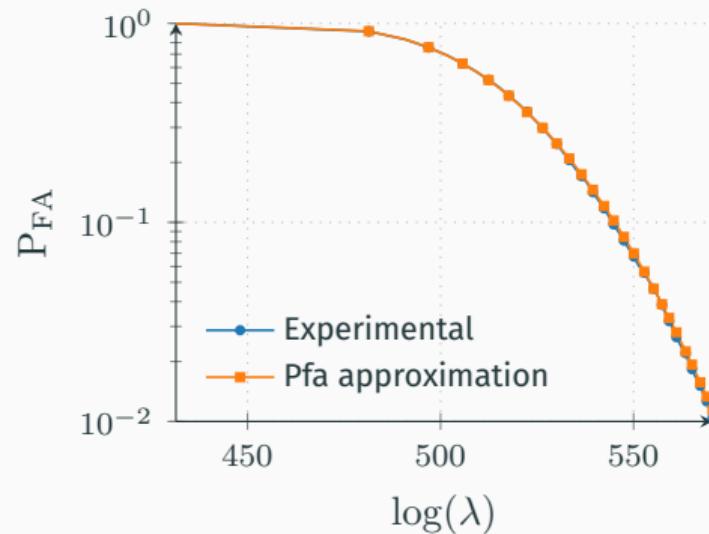
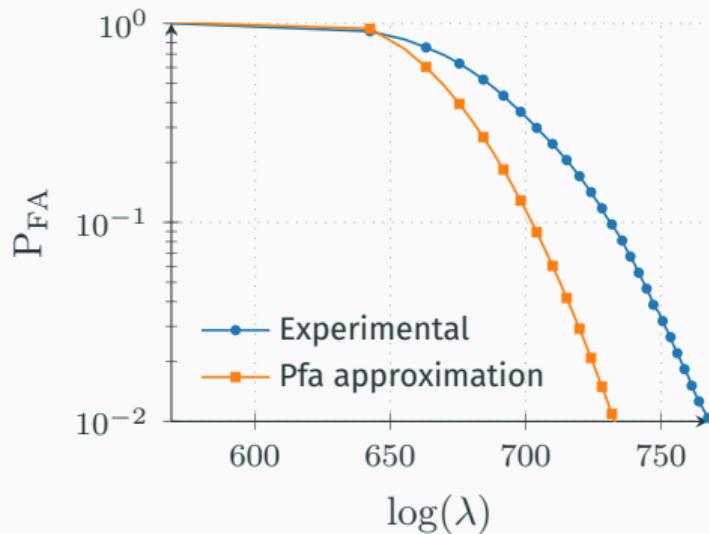
$$f = (T-1)p^2, \rho = 1 - \frac{(2p^2 - 1)}{6(T-1)p} \left(\frac{T}{N} - \frac{1}{NT} \right),$$

$$\omega_2 = \frac{p^2(p^2 - 1)}{24\rho^2} \left(\frac{T}{N^2} - \frac{1}{(NT)^2} \right) - \frac{p^2(T-1)}{4} \left(1 - \frac{1}{\rho} \right)^2$$

Note: A less accurate approximation has also been derived by Box in [BOX, 1949]

Verification of False alarm/threshold relationship

Parameters of the simulation: $p = 15$, $T = 5, 10^4$ trials. Left: $N = 20$. Right: $N = 50$.



Non-Gaussianity of High-resolution images

Issue: Data is *non-Gaussian* !

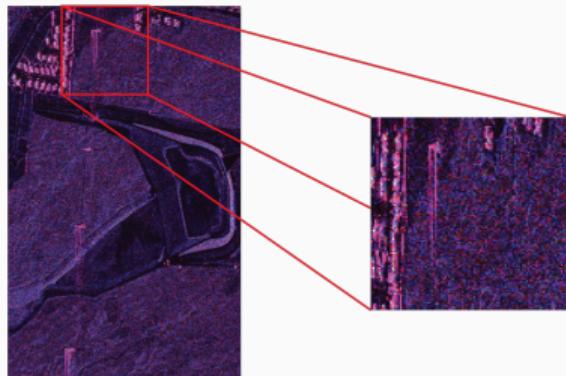
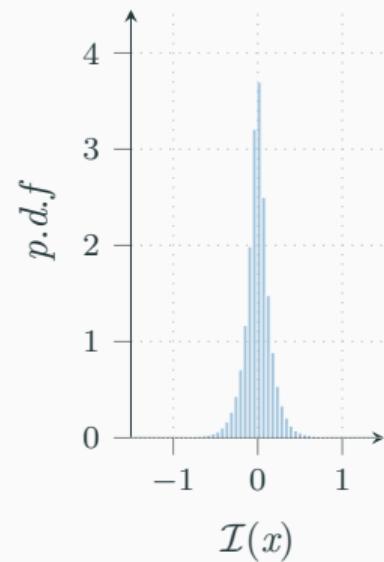
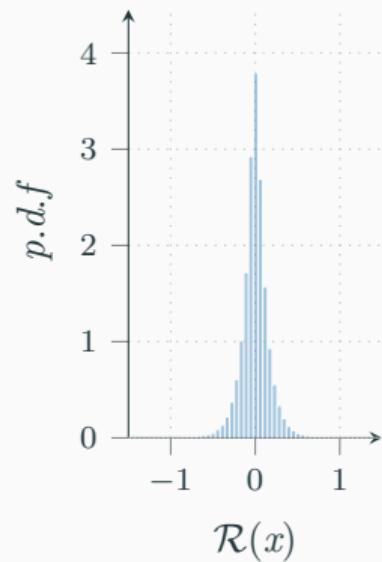
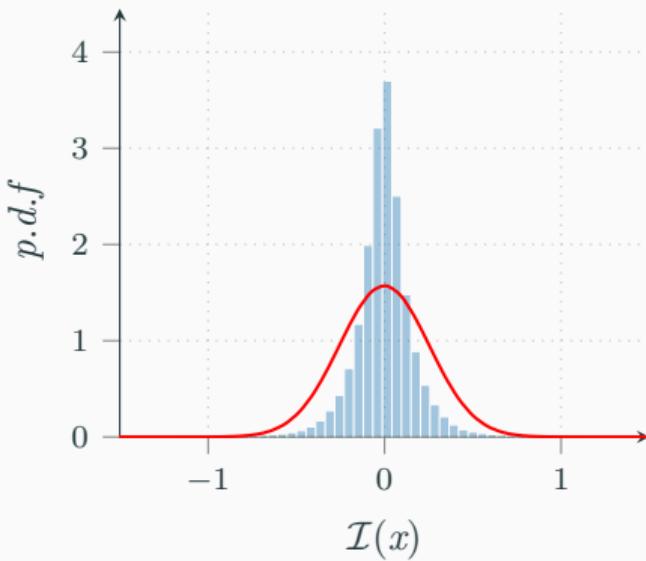
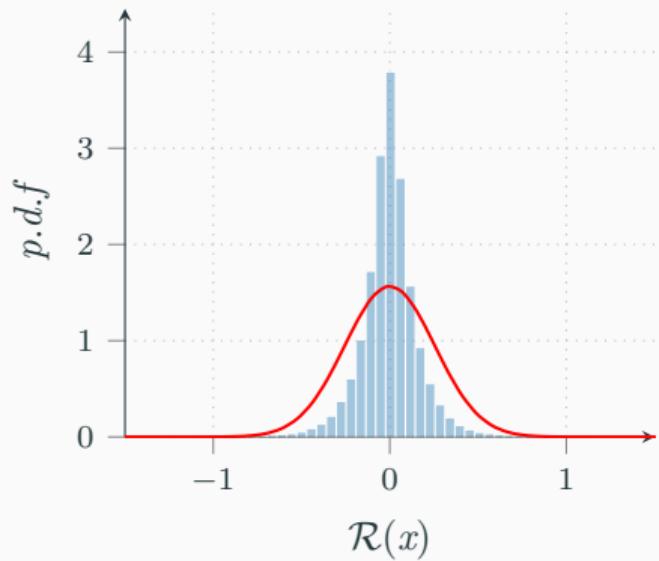


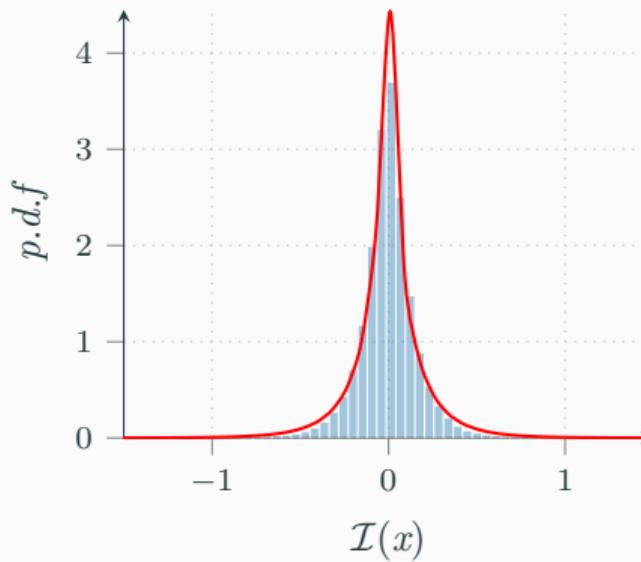
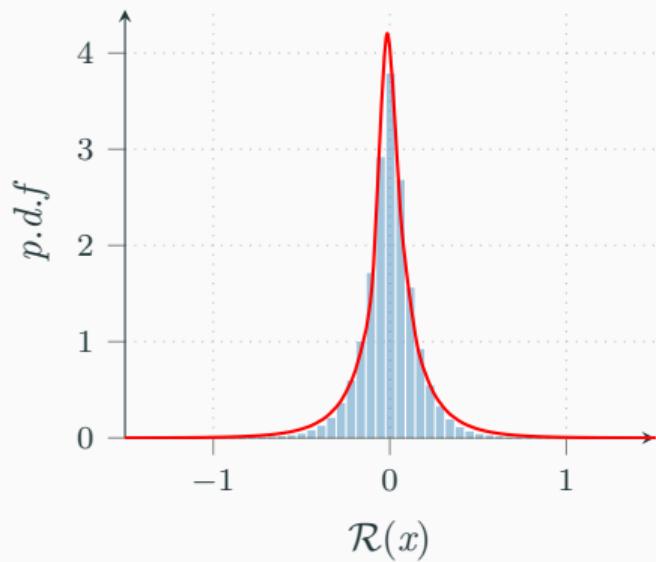
Figure 3: UAVSAR data (Courtesy NASA/JPL-Caltech)



Gaussian distribution fitting

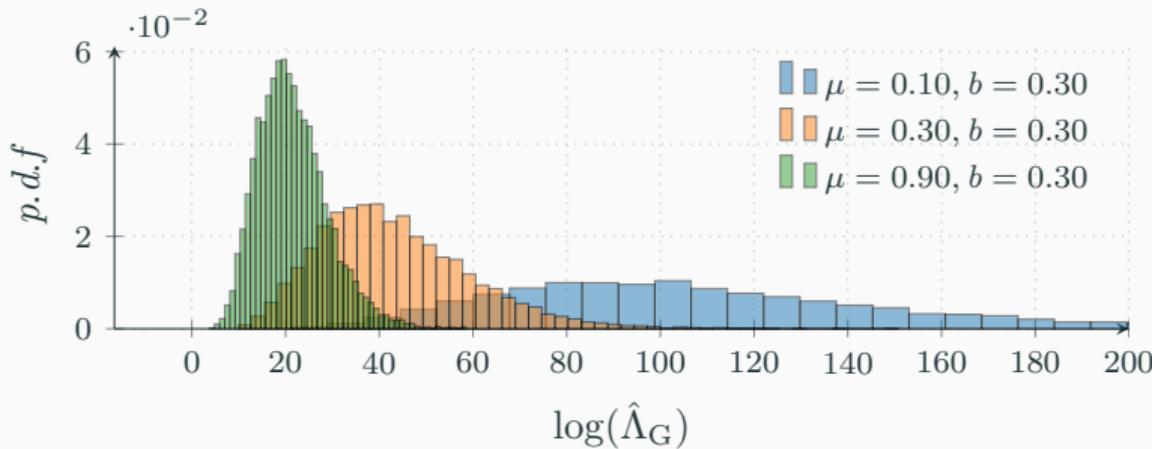


Generalized Gaussian distribution fitting



Non CFAR behaviour in non-Gaussian context: Experimental results

In simulation: $\mathbf{x}_k^t = \sqrt{\tau_k^t} \mathbf{z}_k^t$ where $\mathbf{z}_k^t \sim \mathbb{C}\mathcal{N}(\mathbf{0}_p, (0.5^{|i-j|})_{ij})$ and $\tau_k^t \sim \Gamma(\mu, b)$ with $p = 3, N = 10, T = 3 \cdot 10^4$ Monte-Carlo trials.



→ The Gaussian GLRT is **not CFAR** in the context of compound-Gaussian distributions !

Complex Elliptical Symmetric modelling

Definition

A vector $\mathbf{x} \in \mathbb{C}^p$ is said to be Complex Elliptical Symmetric (CES) distributed with density generator function $g : \mathbb{R}^+ \rightarrow \mathbb{R}^+$, mean parameter $\mu \in \mathbb{C}^p$ and scatter matrix parameter $\Sigma \in \mathbb{S}_{\mathbb{H}}^p$, denoted $\mathbf{x} \sim \mathbb{C}\mathcal{E}(g, \mu, \Sigma)$ if its PDF is of the following form:

$$p_{\mathbf{x}}(\mathbf{x}; g, \mu, \Sigma) = \mathfrak{C}_{p,g} |\Sigma|^{-1} g \left\{ (\mathbf{x} - \mu)^H \Sigma^{-1} (\mathbf{x} - \mu) \right\}. \quad (5)$$

	Gaussian	Generalized Gaussian	Student-t	W-distribution	K-distribution
$g(t)$	$\exp(-t)$	$\exp(-t^s/b)$ $s, b > 0$	$(1 + t/d)^{-(d+p)}$ $d > 0$	$t^{s-1} \exp(-t^s/b)$ $s, b > 0$	$\sqrt{t}^{\nu-p} K_{\nu-p}(2\sqrt{\nu t})$ $\nu > 0$
$\mathfrak{C}_{p,g}$	π^{-p}	$\frac{s\Gamma(p)b^{-p/s}}{\pi^p\Gamma(p/s)}$	$\frac{\Gamma(p+d)}{\pi^m d^p \Gamma(d)}$	$\frac{s\Gamma(p)b^{-(p+s-1)/s}}{\pi^p \Gamma((p+s-1)/s)}$	$2 \frac{\nu^{(\nu+p)/2}}{\pi^p \Gamma(\nu)}$

[Ollila et al., 2012] proposed to use elliptical distributions for modelling the clutter of HR SAR images:

$$\mathbf{x}_k^t \sim \mathbb{C}\mathcal{E}(\mathbf{0}_p, g, \Sigma_t).$$

1st approach: plug-in of robust estimates

Idea: Use the Gaussian GLRT (T=2)

$$\hat{\Lambda} = \frac{\left| \hat{\Sigma}_0 \right|^N}{\left| \hat{\Sigma}_1 \right|^{\frac{N}{2}} \left| \hat{\Sigma}_2 \right|^{\frac{N}{2}}} \stackrel{H_1}{\gtrless} \lambda \stackrel{H_0}{\lessgtr}$$

where $\hat{\Sigma}_t$ is a **robust estimator**.

Tyler's estimator [Formont et al., 2011]

$$\hat{\Sigma}_t = \frac{p}{N} \sum_{k=1}^N \frac{\mathbf{x}_k^t \mathbf{x}_k^{t^H}}{\mathbf{x}_k^{t^H} \{\hat{\Sigma}_t\}^{-1} \mathbf{x}_k^t},$$

→ Loss of power information

Student M-estimator [Drašković et al., 2019]

$$\hat{\Sigma}_t = \frac{p + \nu/2}{N} \sum_{k=1}^N \frac{\mathbf{x}_k^t \mathbf{x}_k^{t^H}}{\nu/2 + \mathbf{x}_k^{t^H} \{\hat{\Sigma}_t\}^{-1} \mathbf{x}_k^t},$$

→ Choice of ν

2nd approach: GLRT in CES context (1/3)

- **Model:** $\mathbf{x}_k^t \sim \mathbb{CE}(\mathbf{0}_p, g, \Sigma_t)$.
- **Detection test:** $\theta_t = \Sigma_t \quad \& \quad \Phi_t = G$

General result

$$\hat{\Lambda}_{\mathbb{CE}}^g = \frac{\left| \hat{\Sigma}_0 \right|^{TN}}{\prod_{t=1}^T \left| \hat{\Sigma}_t \right|^N} \prod_{k=1}^N \frac{g\left(\mathbf{x}_k^{t^H} \{ \hat{\Sigma}_t \}^{-1} \mathbf{x}_k^t \right)}{g\left(\mathbf{x}_k^{t^H} \{ \hat{\Sigma}_0 \}^{-1} \mathbf{x}_k^t \right)} \begin{matrix} \text{H}_1 \\ \gtrless \\ \text{H}_0 \end{matrix} \lambda, \quad (6)$$

where: $\hat{\Sigma}_t = f_t(\hat{\Sigma}_0)$, $\hat{\Sigma}_0 = \frac{1}{T} \sum_{t=1}^T f_t(\hat{\Sigma}_0)$ and $f_t(\Sigma) = \frac{1}{N} \sum_{k=1}^N \frac{-g'\left(\mathbf{x}_k^{t^H} \{ \hat{\Sigma} \}^{-1} \mathbf{x}_k^t \right)}{g\left(\mathbf{x}_k^{t^H} \{ \hat{\Sigma} \}^{-1} \mathbf{x}_k^t \right)} \mathbf{x}_k^t \mathbf{x}_k^{t^H}$

2nd approach: Statistical properties (2/3)

Proposition

The distribution of $2 \log(\hat{\Lambda}_{\mathbb{C}\mathcal{E}}^g)$ under H_0 is asymptotically equivalent to a $\chi_{(T-1)p^2}^2$ distribution.

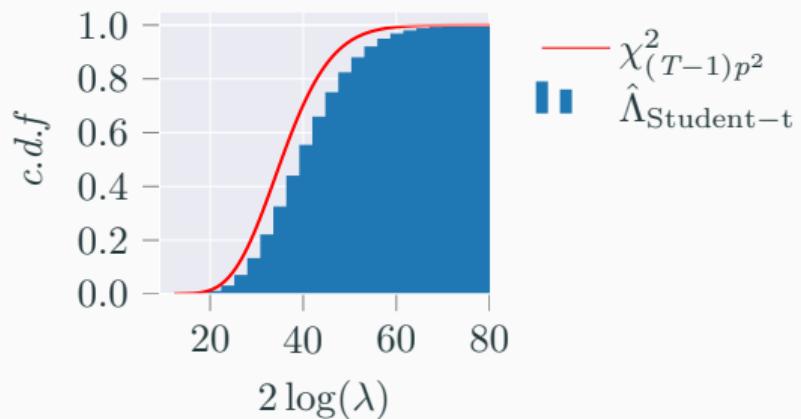
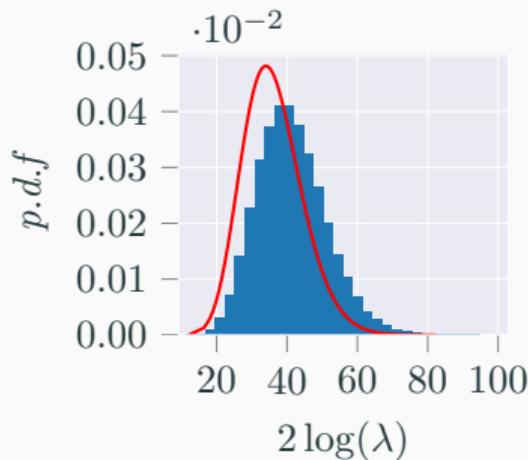


Figure 4: Distribution of statistic under small sample size ($p = 3, T = 5, N = 9, 10^5$ Trials)

2nd approach: Statistical properties (2/3)

Proposition

The distribution of $2 \log(\hat{\Lambda}_{\mathbb{C}\mathcal{E}}^g)$ under H_0 is asymptotically equivalent to a $\chi_{(T-1)p^2}^2$ distribution.

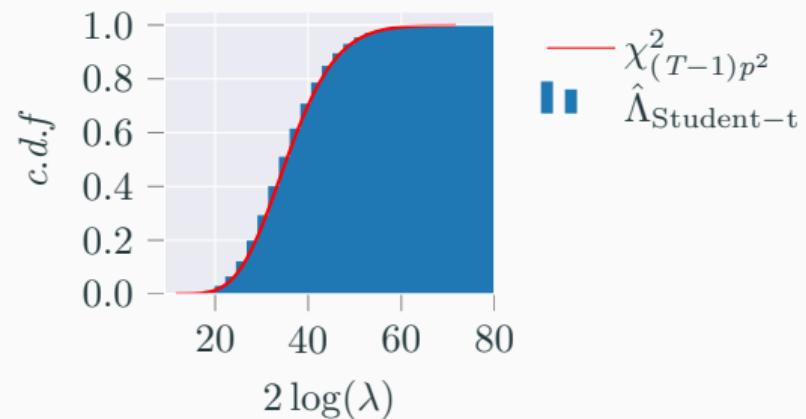
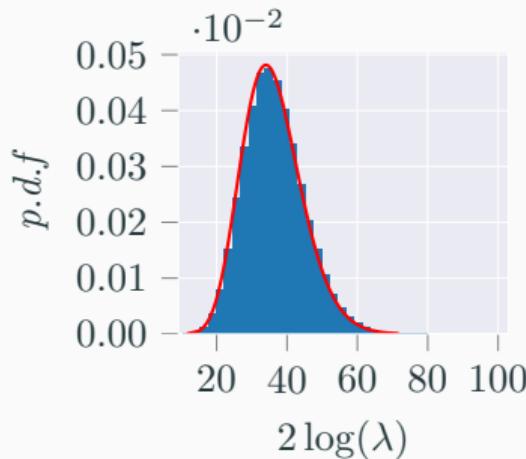
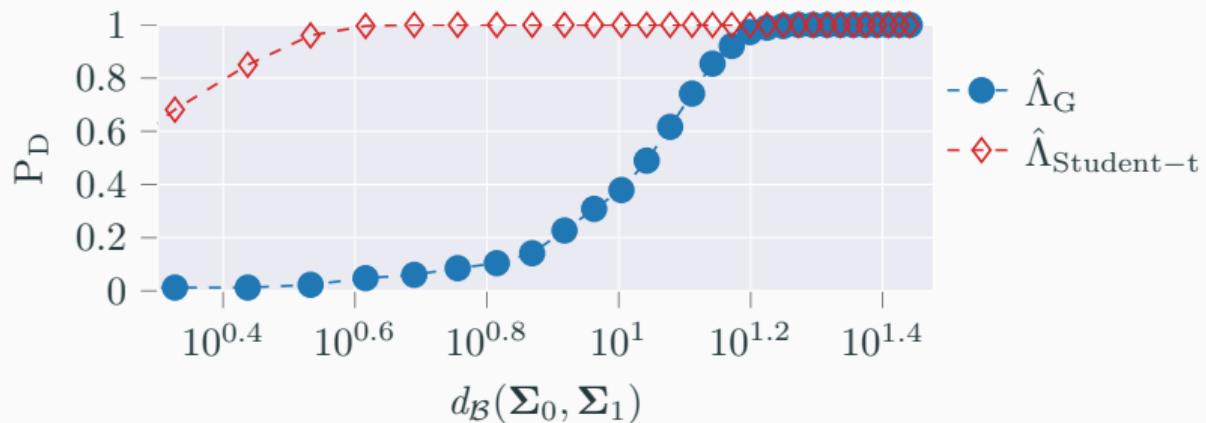


Figure 4: Distribution of statistic under large sample size ($p = 3$, $T = 5$, $N = 100, 10^5$ Trials)

2nd approach: Experimental simulations (3/3)

Monte-Carlo (600 trials) setup

- **Model:** Student-t distribution $\nu = 2, p = 5, N = 15, T = 10, P_{FA} = 10^{-3}$
- **Changes:** $t < 5: \Sigma_t = \Sigma_0, t \geq 5: \Sigma_t = \Sigma_1$
- **SNR:** $d_B(\Sigma_0, \Sigma_1) = \log \left(|\Sigma_1 + \Sigma_2|^2 / (|\Sigma_1| |\Sigma_2|) \right) - 2p \log(2).$



3rd approach: Compound Gaussian model (1/4)

Definition

A set of vectors $\{\mathbf{x}_1, \dots, \mathbf{x}_N\} \in \mathbb{C}^{p \times N}$ is said to follow a deterministic compound-Gaussian model with texture parameter $\boldsymbol{\tau} = [\tau_1, \dots, \tau_N]^T$ and shape matrix parameter $\boldsymbol{\xi} \in \mathbb{S}_{\mathbb{H}}^p$ with $\text{Tr}(\boldsymbol{\xi}) = p$, denoted $\mathbb{CCG}(\boldsymbol{\tau}, \boldsymbol{\xi})$ if each observation is distributed as follows:

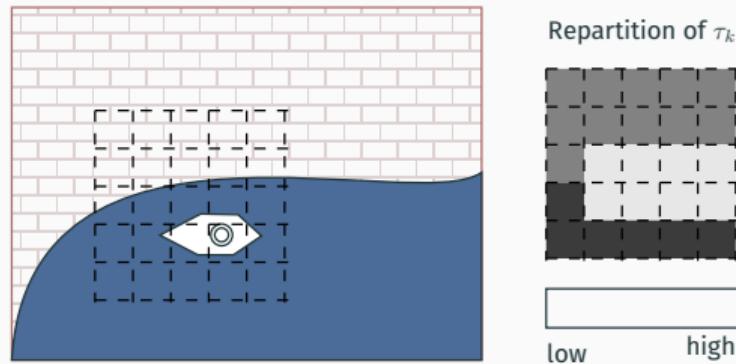
$$\mathbf{x}_k \sim \mathbb{CN}(\mathbf{0}_p, \tau_k \boldsymbol{\xi}) \quad (7)$$

This model is often found in radar applications [Greco and De Maio, 2016, Pascal et al., 2004]. In the context of SAR ITS we can use the following formulation:

$$\mathbf{x}_k^t \sim \mathbb{CCG}(\boldsymbol{\tau}_t, \boldsymbol{\xi}_t).$$

3rd approach: Compound Gaussian model (2/4)

Intuition behind texture:



Detection test: $\theta_t = \{\tau_t, \xi_t\}, \Phi_t = \emptyset$

3rd approach: Compound Gaussian model (3/4)

Proposition

$$\hat{\Lambda}_{\text{MT}} = \frac{\left| \hat{\xi}_0 \right|^{TN}}{\prod_{t=1}^T \left| \hat{\xi}_t \right|^N} \prod_{k=1}^N \frac{\left(\sum_{t=1}^T (\mathbf{x}_k^t)^H \{\hat{\xi}_0\}^{-1} \mathbf{x}_k^t \right)^{Tp}}{T^{Tp} \prod_{t=1}^T \left(\mathbf{x}_k^t \right)^H \{\hat{\xi}_t\}^{-1} \mathbf{x}_k^t} \stackrel{H_1}{\gtrless} \lambda, \quad (8)$$

where:

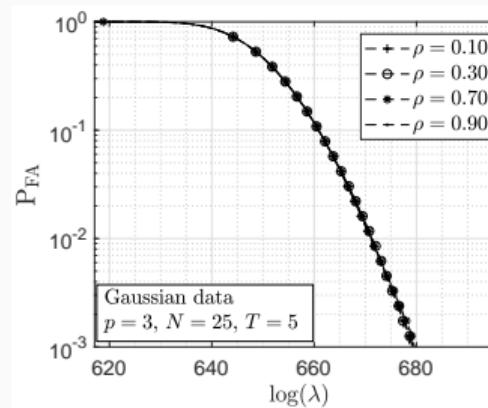
$$\hat{\xi}_0 = \frac{p}{N} \sum_{k=1}^N \frac{\sum_{t=1}^T \mathbf{x}_k^t (\mathbf{x}_k^t)^H}{\sum_{t=1}^T \mathbf{x}_k^t \mathbf{x}_k^t \{\hat{\xi}_0\}^{-1} \mathbf{x}_k^t}, \quad \hat{\xi}_t = \frac{p}{N} \sum_{k=1}^N \frac{\mathbf{x}_k^t (\mathbf{x}_k^t)^H}{\mathbf{x}_k^t \mathbf{x}_k^t \{\hat{\xi}_0\}^{-1} \mathbf{x}_k^t}. \quad (9)$$

3rd approach: Compound Gaussian model (4/4)

CFARness towards shape matrix

$\hat{\Lambda}_{MT}$ is CFAR matrix.

Simulation: $\mathbf{x}_k^t \sim \mathbb{C}\mathcal{N}(\mathbf{0}_p, (\rho^{|i-j|})_{ij})$.

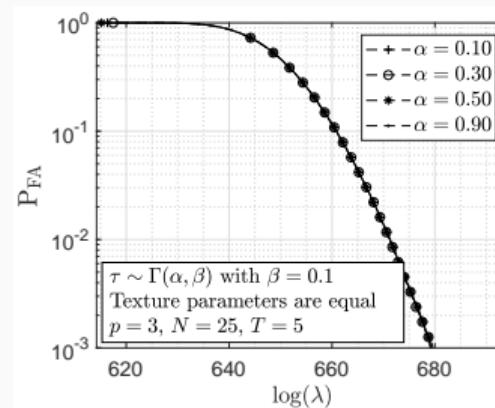


3rd approach: Compound Gaussian model (4/4)

CFARness towards texture parameters

$\hat{\Lambda}_{MT}$ is CFAR texture.

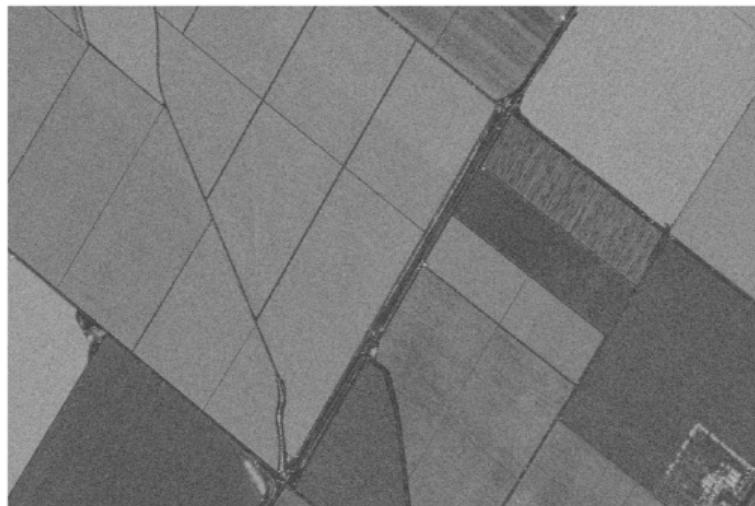
Simulation: $\mathbf{x}_k^t = \sqrt{\tau_k^t} \mathbf{z}_k^t$ where $\mathbf{z}_k^t \sim \mathcal{CN}(\mathbf{0}_p, (0.3^{|i-j|})_{ij})$ and $\forall(k, t), \tau_k^t = \tau_k$



Application to real data (1/3)

Setup

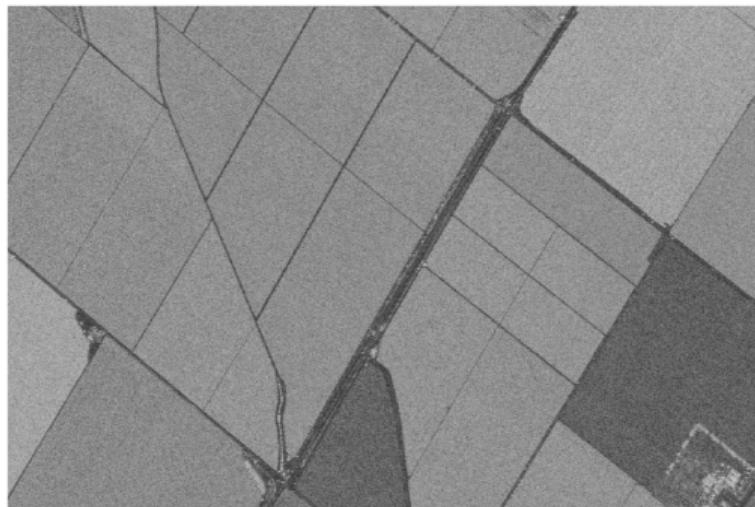
- **Data source:** UAVSAR (Courtesy NASA/JPL)
- **Parameters:** $p = 3, T = 17, N = 25$



Application to real data (1/3)

Setup

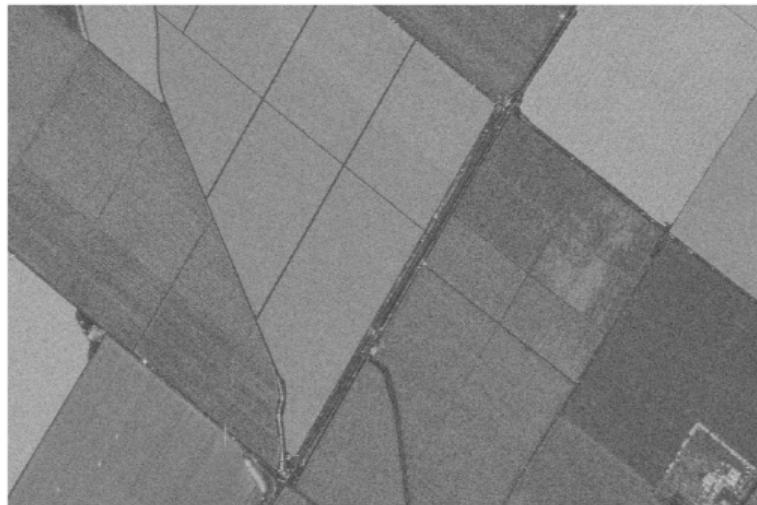
- **Data source:** UAVSAR (Courtesy NASA/JPL)
- **Parameters:** $p = 3, T = 17, N = 25$



Application to real data (1/3)

Setup

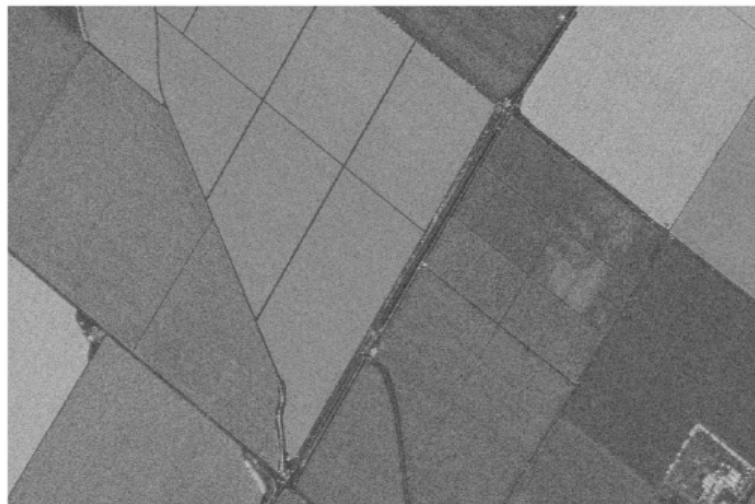
- **Data source:** UAVSAR (Courtesy NASA/JPL)
- **Parameters:** $p = 3, T = 17, N = 25$



Application to real data (1/3)

Setup

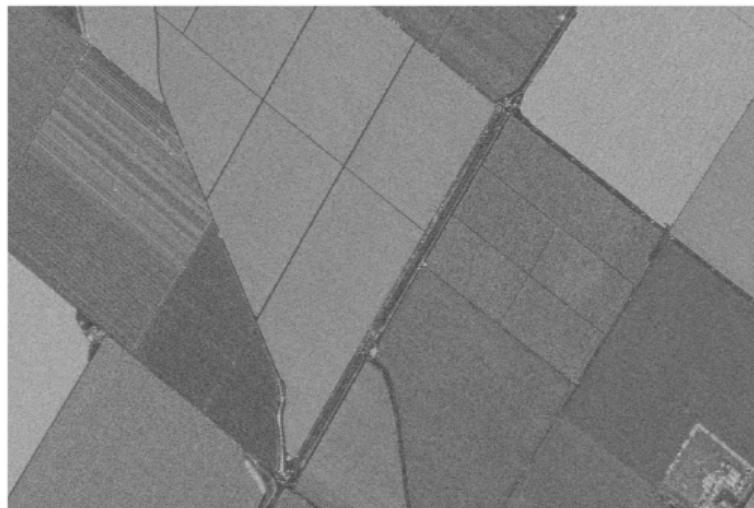
- **Data source:** UAVSAR (Courtesy NASA/JPL)
- **Parameters:** $p = 3, T = 17, N = 25$



Application to real data (1/3)

Setup

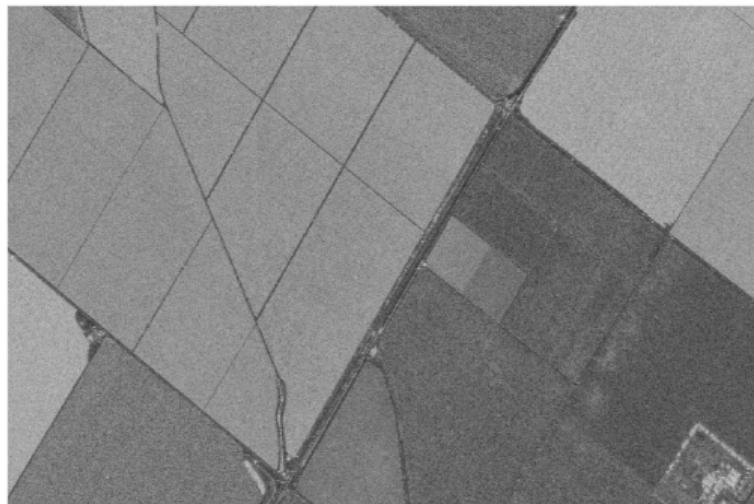
- **Data source:** UAVSAR (Courtesy NASA/JPL)
- **Parameters:** $p = 3, T = 17, N = 25$



Application to real data (1/3)

Setup

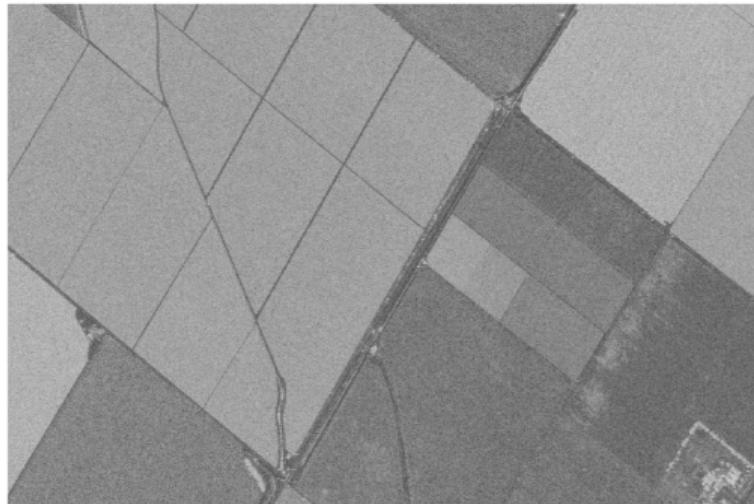
- **Data source:** UAVSAR (Courtesy NASA/JPL)
- **Parameters:** $p = 3, T = 17, N = 25$



Application to real data (1/3)

Setup

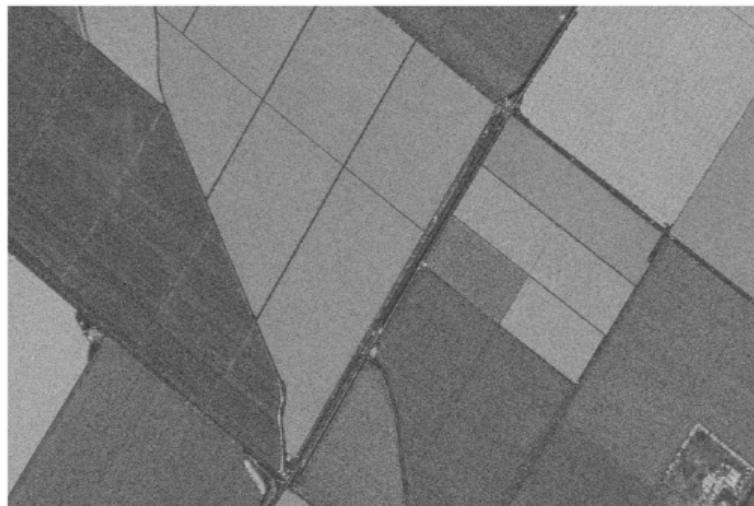
- **Data source:** UAVSAR (Courtesy NASA/JPL)
- **Parameters:** $p = 3, T = 17, N = 25$



Application to real data (1/3)

Setup

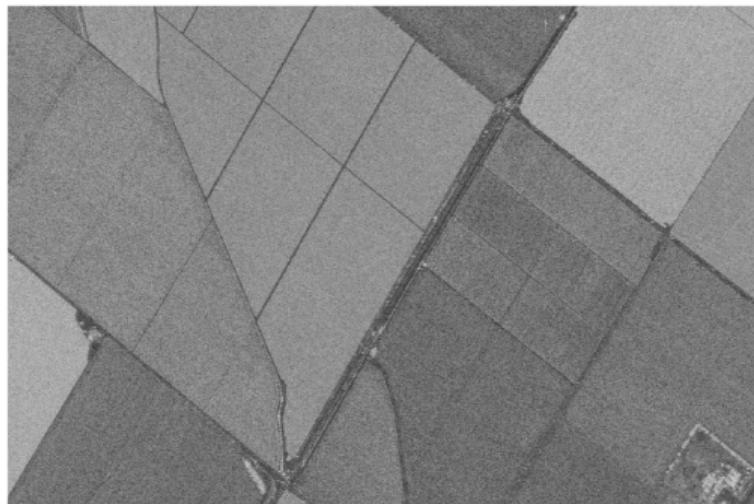
- **Data source:** UAVSAR (Courtesy NASA/JPL)
- **Parameters:** $p = 3, T = 17, N = 25$



Application to real data (1/3)

Setup

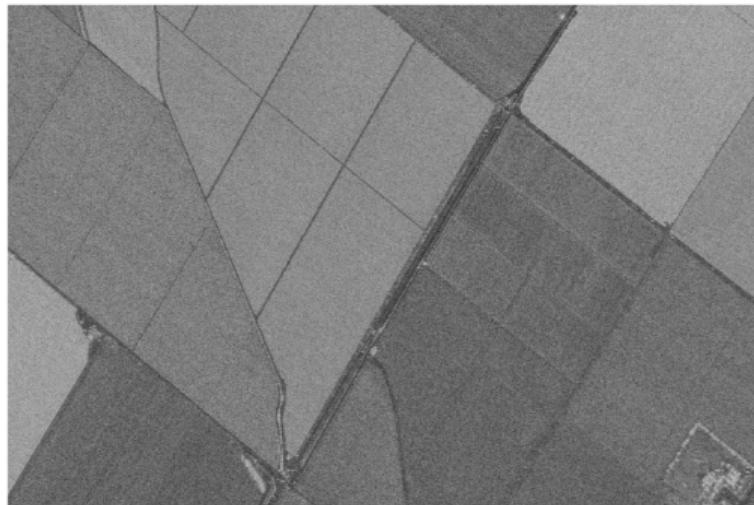
- **Data source:** UAVSAR (Courtesy NASA/JPL)
- **Parameters:** $p = 3, T = 17, N = 25$



Application to real data (1/3)

Setup

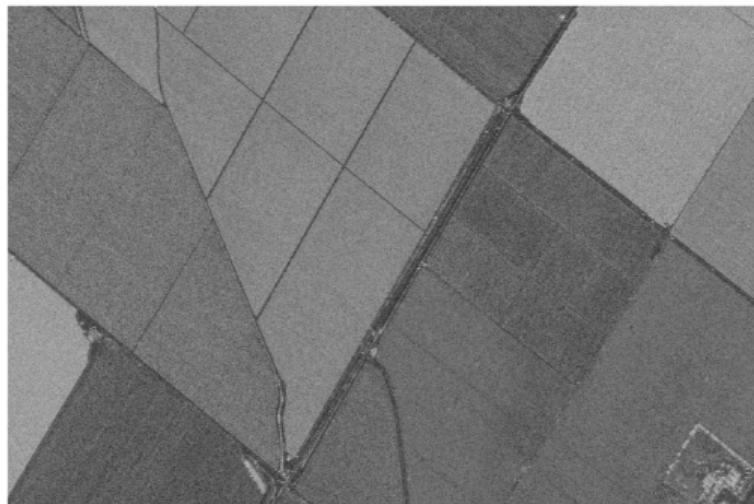
- **Data source:** UAVSAR (Courtesy NASA/JPL)
- **Parameters:** $p = 3, T = 17, N = 25$



Application to real data (1/3)

Setup

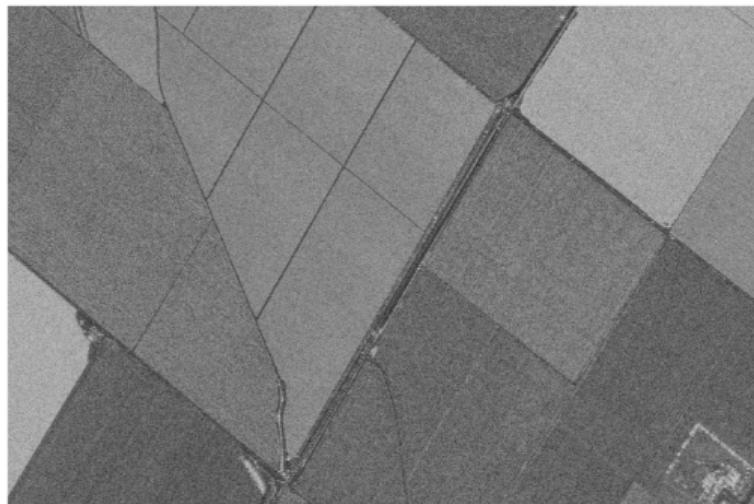
- **Data source:** UAVSAR (Courtesy NASA/JPL)
- **Parameters:** $p = 3, T = 17, N = 25$



Application to real data (1/3)

Setup

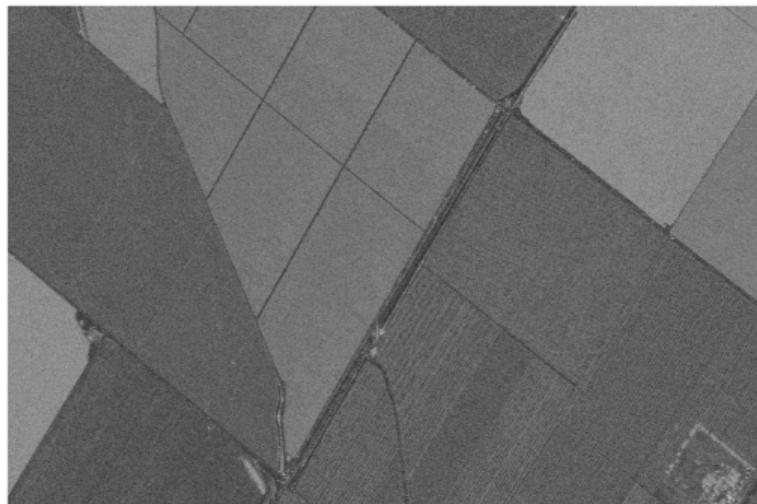
- **Data source:** UAVSAR (Courtesy NASA/JPL)
- **Parameters:** $p = 3, T = 17, N = 25$



Application to real data (1/3)

Setup

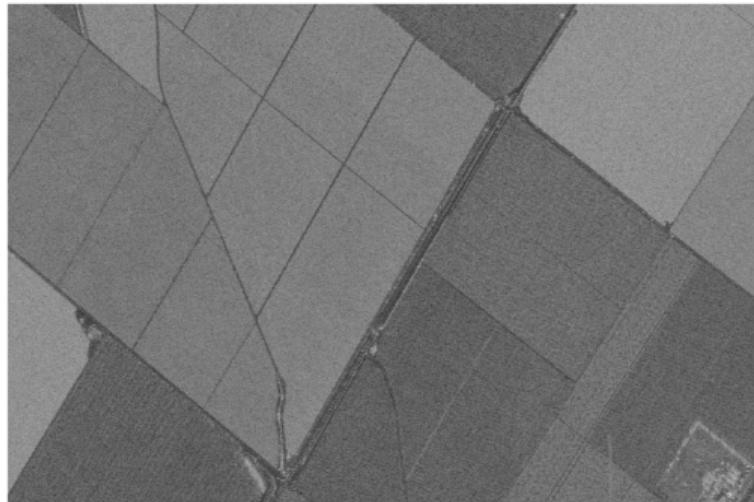
- **Data source:** UAVSAR (Courtesy NASA/JPL)
- **Parameters:** $p = 3, T = 17, N = 25$



Application to real data (1/3)

Setup

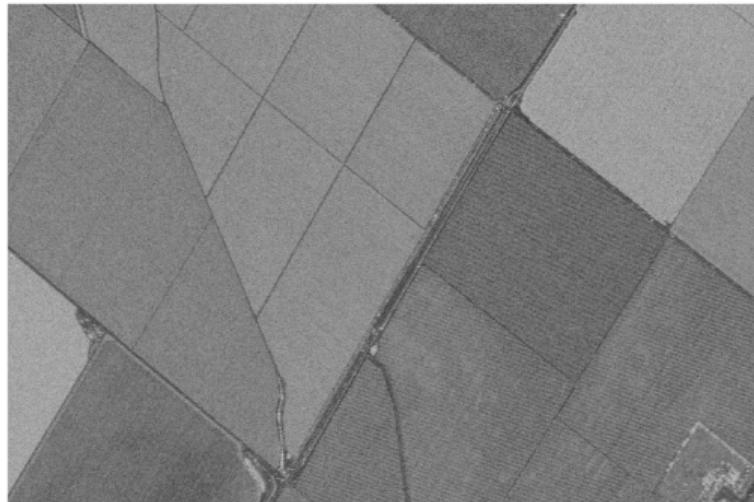
- **Data source:** UAVSAR (Courtesy NASA/JPL)
- **Parameters:** $p = 3, T = 17, N = 25$



Application to real data (1/3)

Setup

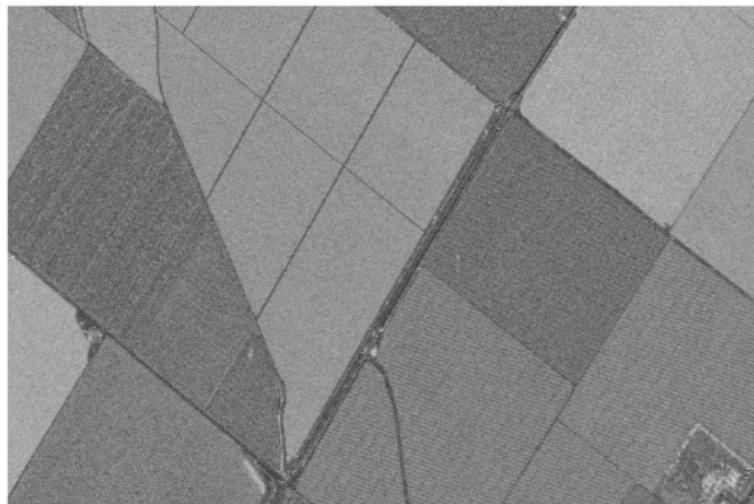
- **Data source:** UAVSAR (Courtesy NASA/JPL)
- **Parameters:** $p = 3, T = 17, N = 25$



Application to real data (1/3)

Setup

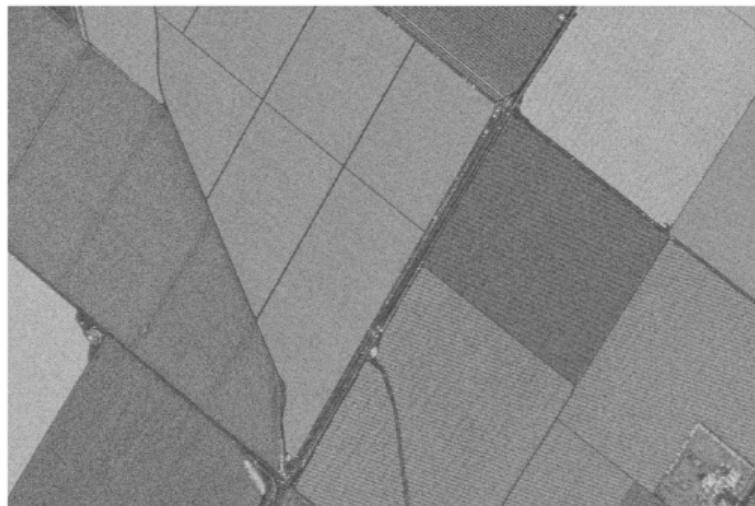
- **Data source:** UAVSAR (Courtesy NASA/JPL)
- **Parameters:** $p = 3, T = 17, N = 25$



Application to real data (1/3)

Setup

- **Data source:** UAVSAR (Courtesy NASA/JPL)
- **Parameters:** $p = 3, T = 17, N = 25$



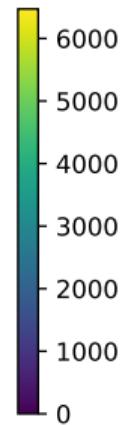
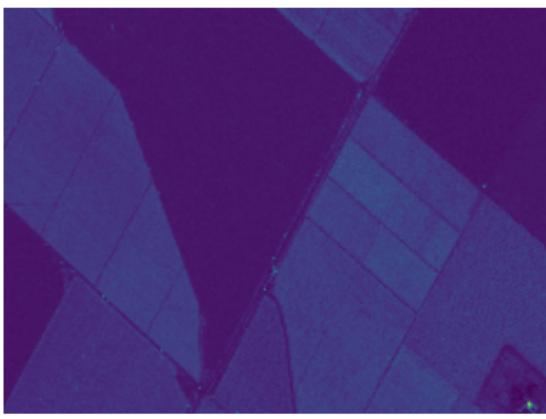
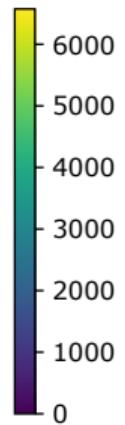
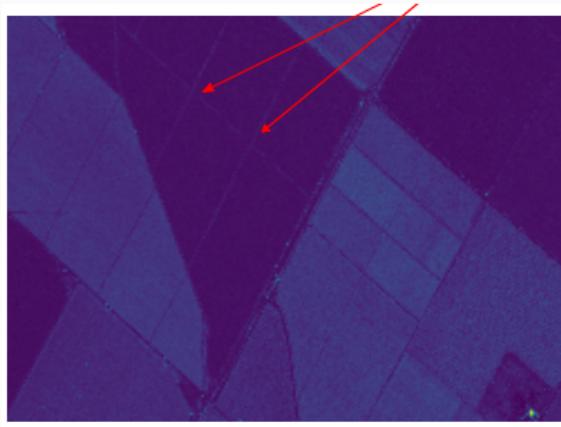
Application to real data (1/3)

Setup

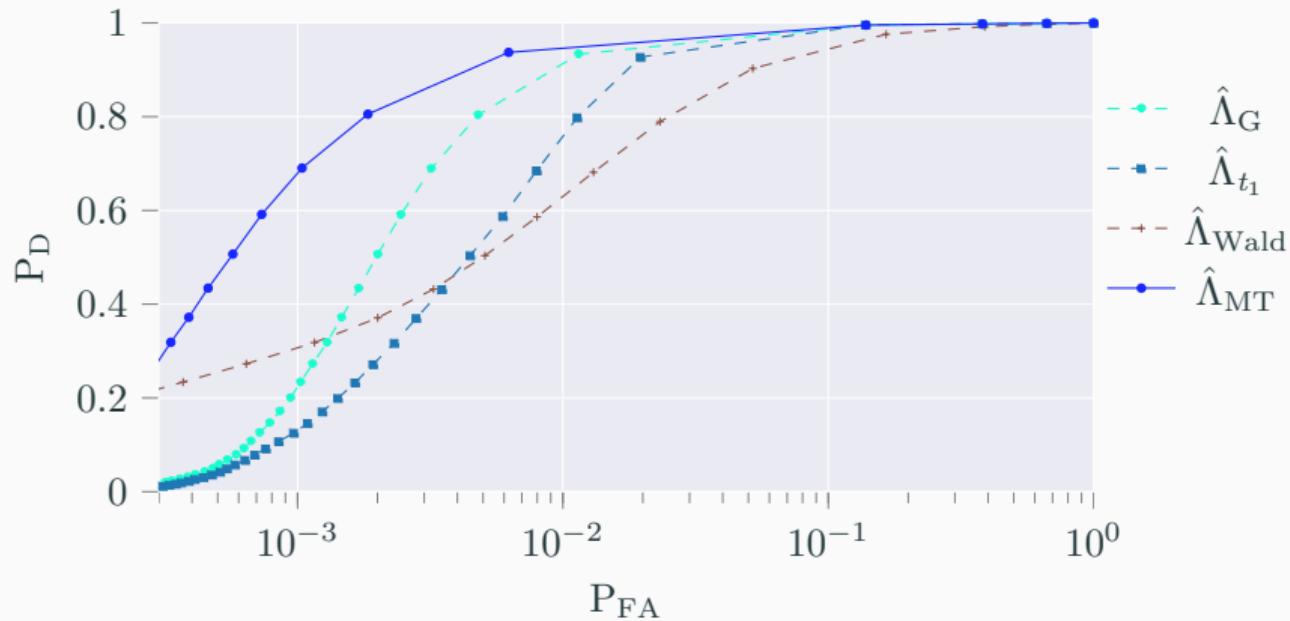
- **Data source:** UAVSAR (Courtesy NASA/JPL)
- **Parameters:** $p = 3, T = 17, N = 25$



Application to real data (2/3)



Application to real data (3/3)



Contents of the presentation

1 Robust Change Detection for SAR Images

- Motivations
- Statistical framework
- Gaussian Assumption
- Extension to robust models

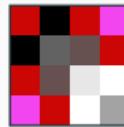
2 Riemannian Geometry and Applications

- Motivations
- Introduction to Riemannian Geometry
- Adapting Euclidean Learning algorithms to Riemannian manifolds
- Applications

Motivations

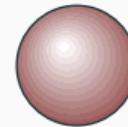
Many learning problems deal with data which is in a **non-Euclidean** space

SPD matrices \mathbb{S}_+^p



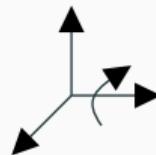
fMRI
Computer vision
SAR images

Hypersphere S_p



Geography
PolSAR

Rotations $SO(3)$

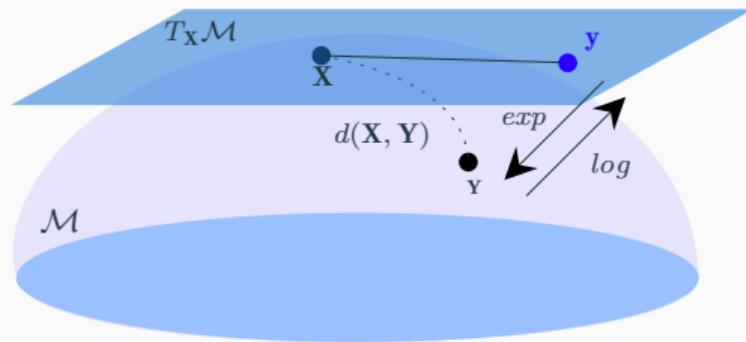


Camera position
3D objects

The data for some applications lie in spaces called manifolds

→ To take into account the nature of the data, we can consider differential geometry !

Intro to Riemannian Geometry

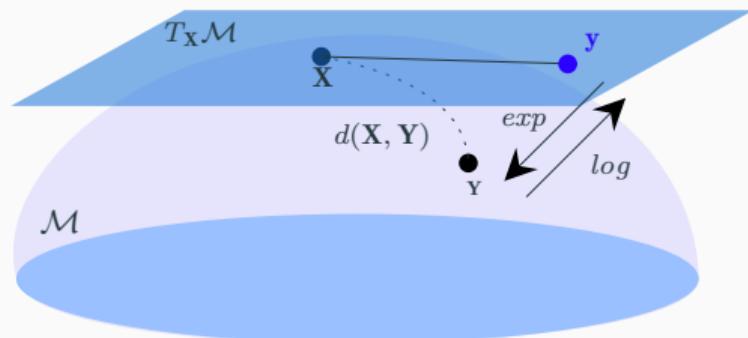


Riemannian manifold

A manifold where we can define a tangent space with an inner product (metric) that is *smoothly* varying.

→ Several tools are available depending on the metric

Intro to Riemannian Geometry



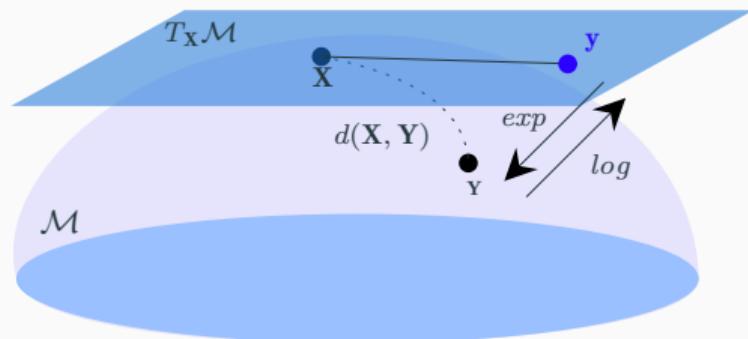
Geodesic distance

Length of the shortest curve between two points.

Example on SPD manifold with affine-invariant metric:

$$d_A(\mathbf{X}, \mathbf{Y}) = \|\log(\mathbf{X}^{-1/2}\mathbf{Y}\mathbf{X}^{-1/2})\|_{\mathbb{F}}. \quad (10)$$

Intro to Riemannian Geometry



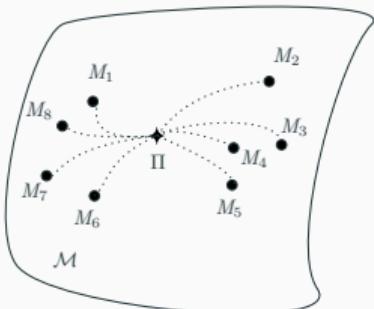
Logarithm mapping

Mapping a point to the tangent space of another point.

Example on SPD manifold with affine-invariant metric

$$\log_X(Y) = X^{\frac{1}{2}} \log(X^{-\frac{1}{2}} Y X^{-\frac{1}{2}}) X^{\frac{1}{2}}. \quad (10)$$

Riemannian mean



Riemannian mean

$$\Pi(\{\alpha_i, \mathbf{M}_i\}_{1 \leq i \leq N}) = \operatorname{argmin}_{\mathbf{M} \in \mathcal{M}} \sum_{i=1}^N \alpha_i d^2(\mathbf{M}, \mathbf{M}_i). \quad (11)$$

Example: SPD manifold with affine-invariant metric

$$\Pi_0 \leftarrow \Pi_E(\{\alpha_i, \mathbf{M}_i\}_{1 \leq i \leq N}), k \leftarrow 0$$

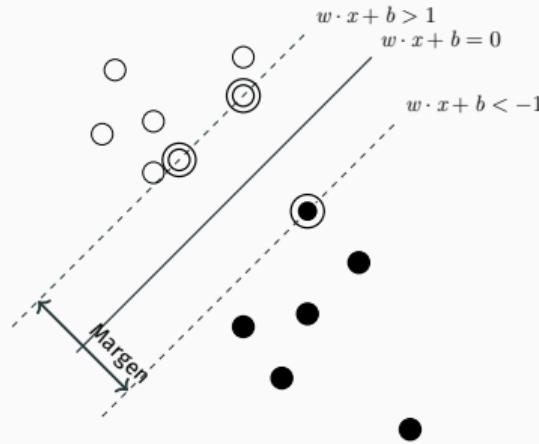
while not converged **do**

$$\left| \begin{array}{l} \Pi_{k+1} \leftarrow \Pi_k^{\frac{1}{2}} \exp \left(\gamma \sum_{i=1}^N \alpha_i \log \left(\Pi_k^{-\frac{1}{2}} \mathbf{M}_i \Pi_k^{-\frac{1}{2}} \right) \right) \Pi_k^{\frac{1}{2}} \\ k \leftarrow k + 1 \end{array} \right.$$

end

$$\Pi_A(\{\alpha_i, \mathbf{M}_i\}_{1 \leq i \leq N}) \leftarrow \Pi_{k+1}$$

Learning procedure



Supervised learning: Given data

$\{(\mathbf{x}_i, y_i) \in \mathbb{R}^d \times \mathcal{C} : 1 \leq i \leq N\}$, where \mathcal{C} is a discrete set of class labels. Find a mapping

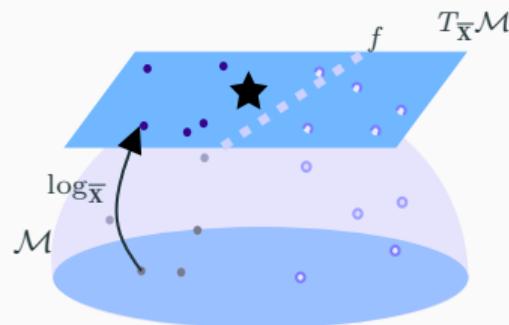
$$f: \mathbf{x} \in \mathbb{R}^d \rightarrow \mathcal{C},$$

that minimizes the empirical risk

$$R = N^{-1} \sum_i (f(\mathbf{x}_i) == y_i).$$

→ How to adapt to Riemannian manifolds ?

1st solution: Map the data to a tangent space



- $X_i / y_i = 1$
 - $X_i / y_i = -1$
- ★ $\bar{X} = \Pi_A (\{w_i, X_i\}_{1 \leq i \leq N})$

Consider the following procedure:

$$f \circ \log_{\Pi_A(\{1/N, X_i\}_{1 \leq i \leq N})} : \mathcal{M} \rightarrow \mathcal{C}. \quad (12)$$

Reference point \bar{X} : Riemannian mean
 $\Pi_A (\{\alpha_i, X_i\}_{1 \leq i \leq N})$ with weights $\alpha_i = 1/N$.

2nd solution: Use geodesic distances

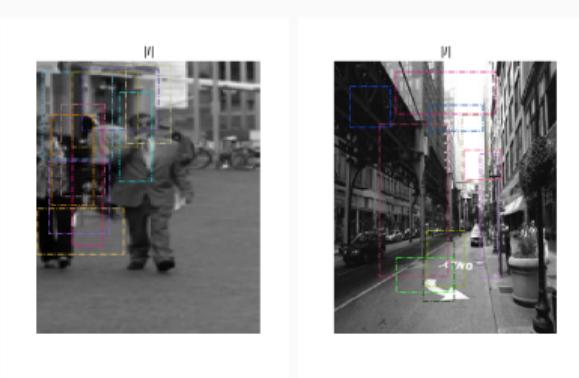
For algorithms based on distances:

- Minimum distance to mean:
 - Compute Riemannian mean of each class
 - Assign each test data to the closest mean (in the Riemannian sense)
- K-nearest neighbors:
 - For each test data compute distance to all training data
 - Assign class by majority of the K-nearest points

Works also for non-supervised classification:

- K-means
- Spectral clustering

Application: Pedestrian detection



Features:

$$\mathbf{z}(x, y) = \left[x, y, |I_x|, |I_y|, \sqrt{I_x^2 + I_y^2}, |I_{xx}|, |I_{yy}|, \arctan \frac{|I_x|}{|I_y|} \right].$$

The *covariance descriptor* of an arbitrary window W is a SPD matrix of the feature vectors $\mathbf{z}(x, y)$

$$\mathbf{C}_W = \frac{1}{n_x n_y - 1} \sum_{x, y \in W} (\mathbf{z}(x, y) - \bar{\mathbf{z}})(\mathbf{z}(x, y) - \bar{\mathbf{z}})^\top,$$

where $\bar{\mathbf{z}} = \frac{1}{n_x n_y} \sum_{x, y} \mathbf{z}(x, y)$ is the sample mean of $\mathbf{z}(x, y)$

Results on INRIA dataset

		Fold 1	Fold 2	Fold 3	Fold 4	mean
Euclidean	RBF SVM	0.819	0.823	0.819	0.820	0.820
	Logitboost	0.934	0.931	0.933	0.935	0.933
	KNN	0.780	0.781	0.780	0.783	0.781
	MDM	0.597	0.595	0.592	0.595	0.595
	LogisticRegression	0.831	0.831	0.832	0.831	0.831
Riemannian	RBF SVM	0.892	0.892	0.892	0.894	0.892
	Logitboost	0.948	0.947	0.946	0.950	0.948
	KNN	0.827	0.825	0.826	0.825	0.826
	MDM	0.692	0.698	0.701	0.699	0.697
	LogisticRegression	0.741	0.709	0.719	0.685	0.714

Application: H- α clustering of SAR images

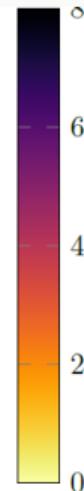
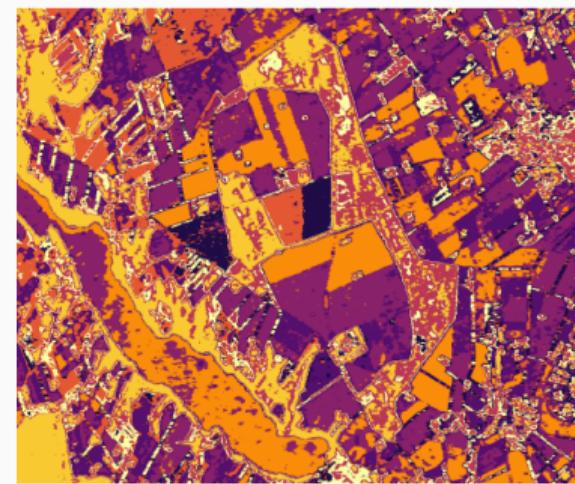
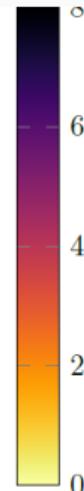
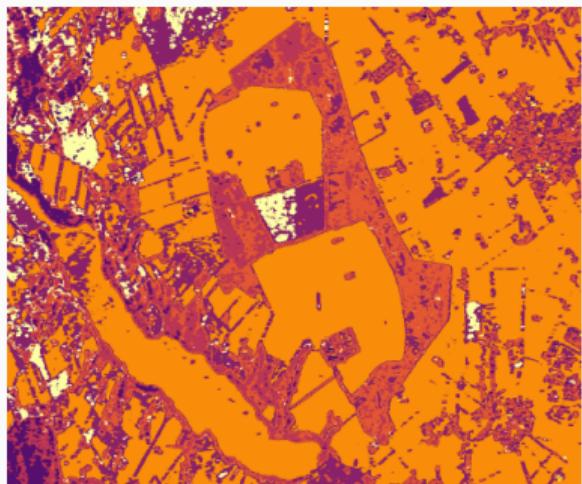
- Data: Emisar Foulom, polarimetric data $p = 3$
- Image size: 1750 rows x 997 columns
- Features: Covariances computed on 7×7 patches



Results

Left: K-means with Wishart distance and Euclidean mean

Right: K-means with Riemannian distance and mean



Merci pour votre attention !

References i

-  Anderson, T. (2003).
An Introduction to Multivariate Statistical Analysis.
Wiley Series in Probability and Statistics. Wiley.
-  BOX, G. E. P. (1949).
A GENERAL DISTRIBUTION THEORY FOR A CLASS OF LIKELIHOOD CRITERIA.
Biometrika, 36(3-4):317–346.
-  Ciuonzo, D., Carotenuto, V., and Maio, A. D. (2017).
On multiple covariance equality testing with application to SAR change detection.
IEEE Transactions on Signal Processing, 65(19):5078–5091.

References ii

-  Conradsen, K., Nielsen, A. A., Schou, J., and Skriver, H. (2003).
A test statistic in the complex wishart distribution and its application to change detection in polarimetric SAR data.
IEEE Transactions on Geoscience and Remote Sensing, 41(1):4–19.
-  Drašković, G., Pascal, F., and Tupin, F. (2019).
 m -nl: Robust nl-means approach for polsar images denoising.
IEEE Geoscience and Remote Sensing Letters, 16(6):997–1001.
-  Formont, P., Pascal, F., Vasile, G., Ovarlez, J.-P., and Ferro-Famil, L. (2011).
Statistical classification for heterogeneous polarimetric SAR images.
IEEE Journal of Selected Topics in Signal Processing, 5(3):567–576.

References iii

-  Greco, M. S. and De Maio, A., editors (2016).
Modern Radar Detection Theory.
SciTech Publishing.
-  Kay, S. M. and Gabriel, J. R. (2003).
An invariance property of the generalized likelihood ratio test.
IEEE Signal Processing Letters, 10(12):352–355.
-  Ollila, E., Tyler, D. E., Koivunen, V., and Poor, H. V. (2012).
Compound-Gaussian clutter modeling with an inverse Gaussian texture distribution.
IEEE Signal Processing Letters, 19(12):876–879.

-  Pascal, F., Ovarlez, J.-P., Forster, P., and Larzabal, P. (2004).
Constant false alarm rate detection in spherically invariant random processes.
In *European Signal Processing Conference (EUSIPCO)*, pages 2143–2146, Vienna.