Theoretical Analysis of an Improved Covariance Matrix Estimator in Non-Gaussian Noise

F. Pascal[123], P. Forster[2], J-P. Ovarlez[1], P. Larzabal[3]

[1] ONERA, Chemin de la Hunière, F-91761 Palaiseau Cedex, France
[2] GEA, 1 Chemin Desvallières, F-92410 Ville d'Avray, France
[3] ENS Cachan, 61 Avenue du Président Wilson, F-94235 Cachan Cedex, France

ICASSP - March 21, 2005

1 - Basic results of Detection Theory - Noise Characterization
2 - Detection Test and Covariance Matrix Estimation
3 - Theoretical Analysis of the Fixed Point Estimator
4 - Application : Adaptive Detection Performance of the GLRT-LQ
5 - Conclusions

Basic results of Detection Theory

- **In a m-dimensional observed vector y, the basic problem of detecting a complex signal $\mathbf{s} = \alpha\,\mathbf{p}$ (where p is a steering vector), embedded in an additive noise c, can be stated as the following binary hypothesis test:**

$$\begin{cases} \textbf{Hypothesis } H_0: & \mathbf{y} = \mathbf{c} & \mathbf{y}_i = \mathbf{c}_i \quad i = 1, \ldots, N \\ \textbf{Hypothesis } H_1: & \mathbf{y} = \mathbf{s} + \mathbf{c} & \mathbf{y}_i = \mathbf{c}_i \quad i = 1, \ldots, N \end{cases}$$

  where the $\mathbf{c}_i$'s are N signal-free independent measurements (called secondary data) used to estimate, for example, the clutter covariance matrix.

- **Detection: Neyman-Pearson criterion: Maximize the *probability of detection* P_d for a given *probability of false alarm* P_{fa}**
  - ***Probability of detection* P_d: Maximise the probability to decide H_1 when the signal is present.**
  - ***Probability of false alarm* P_{fa}: Probability to decide H_1 when the signal is missing.**

$$\Downarrow$$

When the noise PDF (probability density function) is known *a priori*, Maximum Likelihood Theory is used to decide the hypothesis.

Basic results of Detection Theory

Detection test: **Comparison of the Likelihood Ratio $\Lambda(\mathbf{y})$ with a given threshold η:**

$$\Lambda(\mathbf{y}) = \frac{p_{\mathbf{y}}(\mathbf{y}/H_1)}{p_{\mathbf{y}}(\mathbf{y}/H_0)} \underset{H_1}{\overset{H_0}{\lessgtr}} \eta\,,$$

to ensure $P_{fa} = \mathbb{P}(\Lambda(\mathbf{y}) > \eta/H_0)$.

Performance Analysis of the detection test for a given P_{fa} and when the target is present for different SNRs (Signal to Noise Ratio):

$$P_d = \mathbb{P}(\Lambda(\mathbf{y}) > \eta/H_1).$$

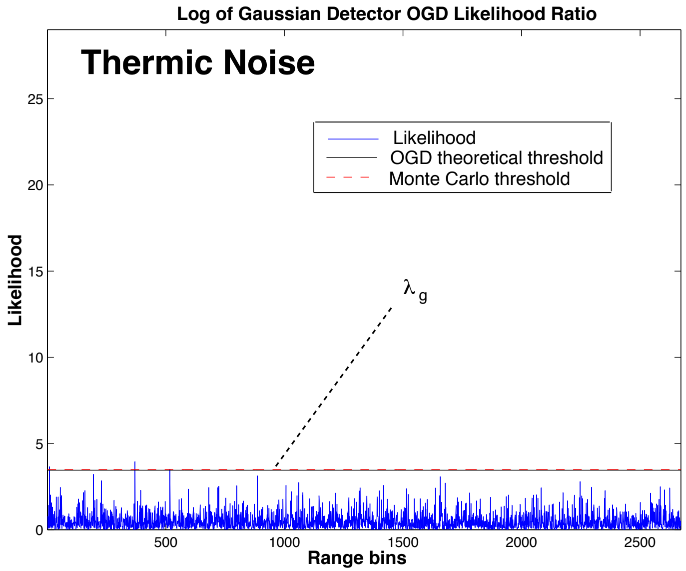
When some parameters θ (as range, velocity, clutter parameter, ...) have to be estimated, the associated detection test (GLRT) is said to be θ-CFAR if its statistics does not depend on θ

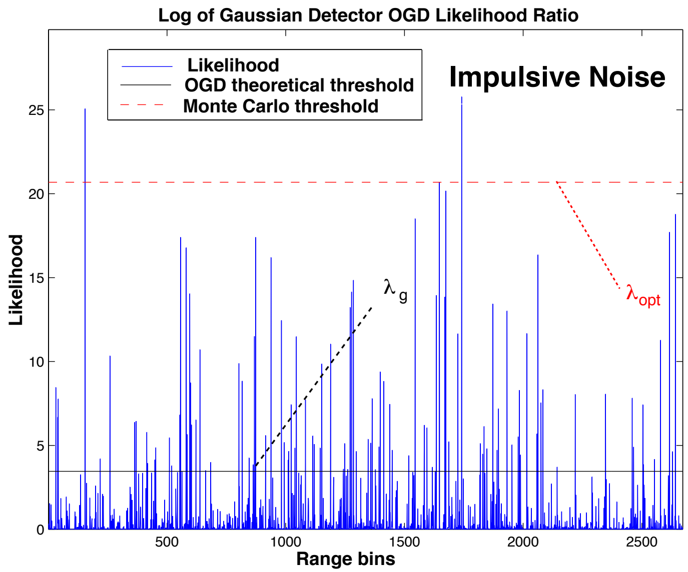
Basic results of Detection Theory

Failure of the basic detector with non-Gaussian background

- **In Gaussian case, Optimal Gaussian Detector OGD is a quadratic detector:**

$$\Lambda(\mathbf{y}) = \frac{|\mathbf{p}^{\dagger}\mathbf{M}^{-1}\mathbf{y}|^2}{\mathbf{p}^{\dagger}\mathbf{M}^{-1}\mathbf{p}} \underset{H_1}{\overset{H_0}{\lessgtr}} \lambda_g$$





- **The detection threshold $\lambda_g = \sqrt{-\ln P_{fa}}$ calculated for Gaussian assumption ensures a fixed probability of false alarm.**

- **Threshold adjustment is optimal for Gaussian noise but generates false alarms when the noise is non-Gaussian with same power. Increasing the detection threshold ($\lambda_g \rightarrow \lambda_{opt}$) for the noise allows to adjust the wanted probability of false alarm but corrupts the detection.**

⇒ OGD detection performance significantly decreases when noise hypothesis are not valid

⇒ Knowing the noise characterization is required.

Noise Characterization

◇ Characterization with Spherically Invariant Random Processes (SIRP)

- **Compound processes representation:** $\mathbf{c} = \mathbf{x}\sqrt{\tau}$
  - **$\mathbf{x}$ is a spherically complex Gaussian vector (*speckle*) with the covariance matrix $2\mathbf{M}$ which can modelize, for example, the temporal fluctuations of the clutter ,**
  - **τ is a positive random variable, independent from x, with statistic law $p(\tau)$ called the *texture* which can modelize the spatial fluctuations of the clutter power .**
- **Probability Density Function under H_0:** $p_{\mathbf{c}}(\mathbf{c}/H_0) = \displaystyle\int_0^{+\infty} \frac{1}{(2\,\pi\,\tau)^m |\mathbf{M}|} \exp\left(-\frac{\mathbf{c}^\dagger\,\mathbf{M}^{-1}\,\mathbf{c}}{2\,\tau}\right) p(\tau)\,d\tau$.

◇ Advantages:

- **Modelize a random walk ,**
- **The family of the SIRP includes an infinity of laws: Gauss, Rayleigh, Chi ^{2}, Laplace, Cauchy, Weibull, K-distribution, Alpha-Stable, ...**
- **The SIRP statistics are invariant by linear filtering (Matched Filter, Doppler Filtering),**
- **The SIRP kernel is Gaussian: the target parameters estimates with the Maximum Likelihood are given by the maximization of the traditional matched filter.**

Detection Test

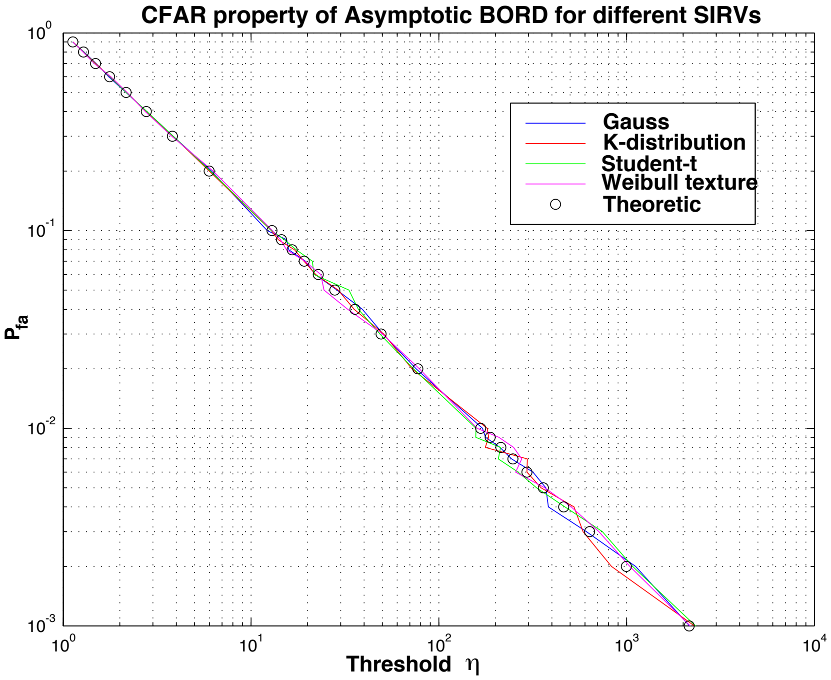
When M is known and τ is a random variable, the resulting detection test is the GLRT-LQ [Conte, Gini, Jay] :

$$\Lambda(\mathbf{M}) = \frac{|\mathbf{p}^\dagger \mathbf{M}^{-1}\mathbf{y}|^2}{(\mathbf{p}^\dagger \mathbf{M}^{-1}\mathbf{p})(\mathbf{y}^\dagger \mathbf{M}^{-1}\mathbf{y})} \underset{H_1}{\overset{H_0}{\lessgtr}} \lambda\,.$$

- **The Likelihood Ratio $\Lambda(\mathbf{M})$ statistic is independent of the texture statistic $p(\tau)$ under hypothesis H_0,**

- **Thus the relationship between P_{fa} and the detection threshold is independent of the texture statistic $p(\tau)$ under hypothesis H_0 and is expressed as:**

$$\lambda = P_{fa}^{\frac{m}{1-m}} \qquad \textbf{(1)}$$



Texture-CFAR property for the GLRT-LQ

LIMITATION: SIRP COVARIANCE MATRIX M IS GENERALLY UNKNOWN.

Covariance Matrix Estimation

Problem: **In practice, SIRP covariance matrix $\mathbf{M}$ is unknown. When $\widehat{\mathbf{M}}$ estimates $\mathbf{M}$ with N (finite) measurements, the relationship $\lambda = P_{fa}^{\frac{m}{1-m}}$ is not valid any more because $\widehat{\mathbf{M}}$ is a random matrix.**

$$\Downarrow$$

$$\textbf{New Likelihood Ratio: } \hat{\Lambda}(\widehat{\mathbf{M}}) = \frac{|\mathbf{p}^\dagger \widehat{\mathbf{M}}^{-1} \mathbf{y}|^2}{(\mathbf{p}^\dagger \widehat{\mathbf{M}}^{-1} \mathbf{p})(\mathbf{y}^\dagger \widehat{\mathbf{M}}^{-1} \mathbf{y})}$$

Consequence: As the distribution and the properties of $\hat{\Lambda}(\widehat{\mathbf{M}})$ depend on the nature of $\widehat{\mathbf{M}}$, the 3 following estimators will be analyzed:

- $\widehat{\mathbf{M}}_{\mathcal{W}} = \sum_{i=1}^{N} \mathbf{x}_i\,\mathbf{x}_i^\dagger$**: the Sample Covariance Matrix which is Wishart distributed and is only used in a theoretical work or for Gaussian process,**
- $\widehat{\mathbf{M}}_{NSCM} = \frac{m}{N} \sum_{i=1}^{N} \frac{\mathbf{c}_i\,\mathbf{c}_i^\dagger}{\mathbf{c}_i^\dagger\,\mathbf{c}_i} = \frac{m}{N} \sum_{i=1}^{N} \frac{\mathbf{x}_i\,\mathbf{x}_i^\dagger}{\mathbf{x}_i^\dagger\,\mathbf{x}_i}$**: the Normalized Sample Covariance Matrix [Conte 1994], classically used in the radar literature,**
- $\widehat{\mathbf{M}}_{fp}$ **, solution of** $\widehat{\mathbf{M}} = \frac{m}{N} \sum_{i=1}^{N} \left(\frac{\mathbf{c}_i \mathbf{c}_i^\dagger}{\mathbf{c}_i^\dagger \widehat{\mathbf{M}}^{-1} \mathbf{c}_i} \right)$**, equation resulting from the Maximum Likelihood.**

Wishart: Theoretical Results

Following results have been derived from works of Kraut and Scharf.

- **$\hat{\Lambda}(\widehat{\mathbf{M}}_{\mathcal{W}})$ distribution:**

$$g_{N,m}(x) = \frac{(N-m+1)(m-1)}{(N-1)} \frac{{}_2F_1(a,a;b;x)}{(1-x)^{N-m}} \Pi_{[0,1]}(x)$$

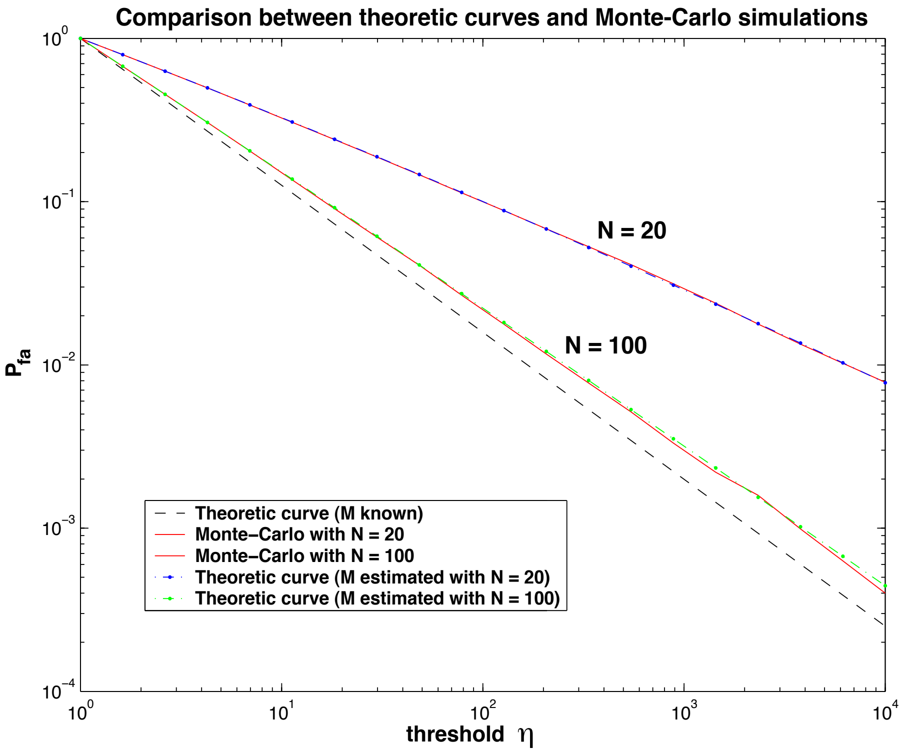
where $a = N - m + 2$, $b = N + 2$ and ${}_2F_1$ is the hypergeometric function.

- **Relationship P_{fa}- threshold:**

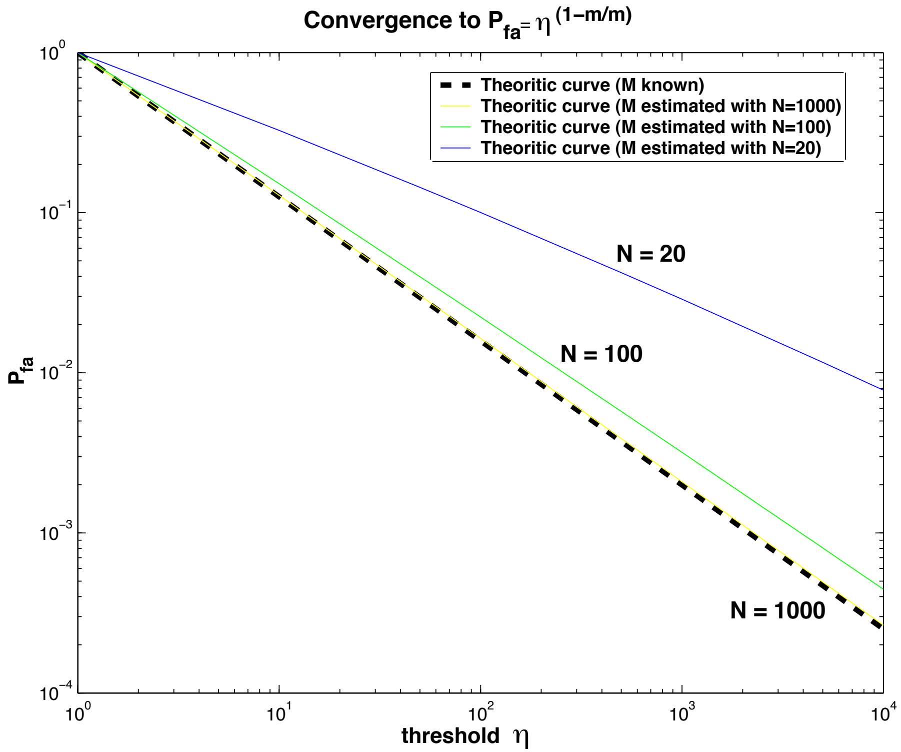
$$\begin{aligned} P_{fa} &= \eta^{-\frac{a-1}{m}} \, {}_2F_1\left(a, a-1; b-1; 1-\eta^{-\frac{1}{m}}\right) \\ &= (1-\lambda)^{a-1} \, {}_2F_1(a, a-1; b-1; \lambda) \end{aligned} \quad \textbf{(2)}$$

with $\lambda = 1 - \eta^{-\frac{1}{m}}$

Wishart: Monte-Carlo Simulations



Validation of the relation (2)



Convergence of (2) to (1)

- **Figure 1: The Monte Carlo simulations confirm the theoretical result given by (2) .**
- **Figure 2: Convergence of (2) to (1) when N tends to infinity (The estimate covariance matrix tends to the true one).**

Theoretical Analysis of the Fixed Point Estimator $\widehat{\mathbf{M}}_{fp}$

When the τ_i's are assumed to be deterministic in the N $\mathbf{c}_i$'s used to estimate the covariance matrix M, the Maximum Likelihood Estimator (called the FPE) is given as THE solution of the following equation:

$$\widehat{\mathbf{M}} = f(\widehat{\mathbf{M}})$$

where the function f is defined as follows: $f(\widehat{\mathbf{M}}) = \dfrac{m}{N}\displaystyle\sum_{i=1}^{N}\left(\dfrac{\mathbf{c}_i\mathbf{c}_i^\dagger}{\mathbf{c}_i^\dagger\widehat{\mathbf{M}}^{-1}\mathbf{c}_i}\right).$

The study of function f allows to establish the following results:

1. **the function f admits a single fixed point, called $\widehat{\mathbf{M}}_{fp}$;**
2. **The fixed point can be easily obtained with recursive algorithm;**
3. **$\widehat{\mathbf{M}}_{fp}$ is unbiased;**
4. **New Likelihood Ratio:** $\hat{\Lambda}(\widehat{\mathbf{M}}_{fp}) = \dfrac{|\mathbf{p}^\dagger\,\widehat{\mathbf{M}}_{fp}^{-1}\,\mathbf{y}|^2}{(\mathbf{p}^\dagger\,\widehat{\mathbf{M}}_{fp}\,\mathbf{p})(\mathbf{y}^\dagger\,\widehat{\mathbf{M}}_{fp}^{-1}\,\mathbf{y})}$
5. **The distribution of $\hat{\Lambda}(\widehat{\mathbf{M}}_{fp})$ has a closed-form expression which allows to find the value of the threshold λ for a given P_{fa}.**

Comparison of the 3 estimators

	$\widehat{\mathbf{M}}$	FPE: $\widehat{\mathbf{M}}_{fp}$	$\widehat{\mathbf{M}}_{NSCM}$	Wishart: $\widehat{\mathbf{M}}_{\mathcal{W}}$
Asymptotical	$\mathbb{E}\left(\mathbf{vec}(\widehat{\mathbf{M}})\ \mathbf{vec}(\widehat{\mathbf{M}})^\top\right)$	$\frac{m+1}{m}\ \mathbf{C}_{as}$	$\frac{m}{m+1}\ \mathbf{C}_{as}$	$\mathbf{C}_{as}$
Properties	$\mathbb{E}\left(\mathbf{vec}(\widehat{\mathbf{M}})\ \mathbf{vec}(\widehat{\mathbf{M}})^*\right)$	$\frac{m+1}{m}\ \mathbf{B}_{as}$	$\frac{m}{m+1}\ \mathbf{B}_{as}$	$\mathbf{B}_{as}$
	Bias of $\widehat{\mathbf{M}}$	Unbiased	Biased	Unbiased

- $\mathbf{C}_{as} = \mathbf{P} - \frac{1}{m}\left(\mathbf{vec}(\mathbf{I}_m)\right)\left(\mathbf{vec}(\mathbf{I}_m)\right)^\top$ and **P** is a given permutation matrix.
- $\mathbf{B}_{as} = \mathbf{I}_{m^2} - \frac{1}{m}\left(\mathbf{vec}(\mathbf{I}_m)\right)\left(\mathbf{vec}(\mathbf{I}_m)\right)^\top$.

The covariance matrix C of the asymptotic Gaussian law is perfectly defined by the two quantities :
$\mathbb{E}\left(\mathbf{vec}(\widehat{\mathbf{M}})\ \mathbf{vec}(\widehat{\mathbf{M}})^\top\right)$ **and** $\mathbb{E}\left(\mathbf{vec}(\widehat{\mathbf{M}})\ \mathbf{vec}(\widehat{\mathbf{M}})^*\right)$ **.**

Comparison of the 3 estimators

⋄ Table interpretation : Asymptotical properties

- **Estimators covariance matrix :**

$\widehat{\mathbf{M}}$	FPE: $\widehat{\mathbf{M}}_{fp}$	$\widehat{\mathbf{M}}_{NSCM}$	Wishart: $\widehat{\mathbf{M}}_{\mathcal{W}}$
$\mathbb{E}\left(\mathbf{vec}(\widehat{\mathbf{M}})\ \mathbf{vec}(\widehat{\mathbf{M}})^\top\right)$	$\frac{m+1}{m}\ \mathbf{C}_{as}$	$\frac{m}{m+1}\ \mathbf{C}_{as}$	$\mathbf{C}_{as}$
$\mathbb{E}\left(\mathbf{vec}(\widehat{\mathbf{M}})\ \mathbf{vec}(\widehat{\mathbf{M}})^*\right)$	$\frac{m+1}{m}\ \mathbf{B}_{as}$	$\frac{m}{m+1}\ \mathbf{B}_{as}$	$\mathbf{B}_{as}$

  – $\widehat{\mathbf{M}}_{\mathcal{W}}$, $\widehat{\mathbf{M}}_{fp}$ et $\widehat{\mathbf{M}}_{NSCM}$ **have the same covariance matrix up to a scaling factor.**
  Signification: **the relationship "P_{fa}-threshold" (2) established with $\widehat{\mathbf{M}}_{\mathcal{W}}$ and for N secondary data, is still valid with $\widehat{\mathbf{M}}_{fp}$ when the number of secondary data is $\frac{m}{m+1}\ N$.**

- **Estimators bias :**

$\widehat{\mathbf{M}}$	FPE : $\widehat{\mathbf{M}}_{fp}$	$\widehat{\mathbf{M}}_{NSCM}$	Wishart: $\widehat{\mathbf{M}}_{\mathcal{W}}$
Bias of $\widehat{\mathbf{M}}$	Unbiased	Biased	Unbiased

Comparison of the 3 estimators

- **Independence of $\hat{\Lambda}(\widehat{\mathbf{M}})$ with the texture :**

$\widehat{\mathbf{M}}$	Point Fixe: $\widehat{\mathbf{M}}_{fp}$	$\widehat{\mathbf{M}}_{NSCM}$	Wishart: $\widehat{\mathbf{M}}_{\mathcal{W}}$
CFAR-texture of $\hat{\Lambda}(\widehat{\mathbf{M}})$	**Yes**	**Yes**	**No**

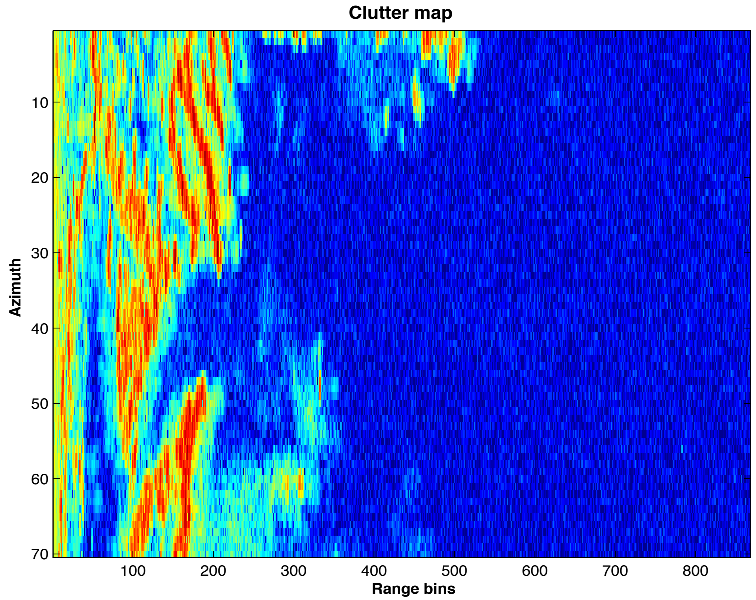
$\hat{\Lambda}(\widehat{\mathbf{M}}_{fp})$ **and** $\hat{\Lambda}(\widehat{\mathbf{M}}_{NSCME})$ **are both independent of the texture.**

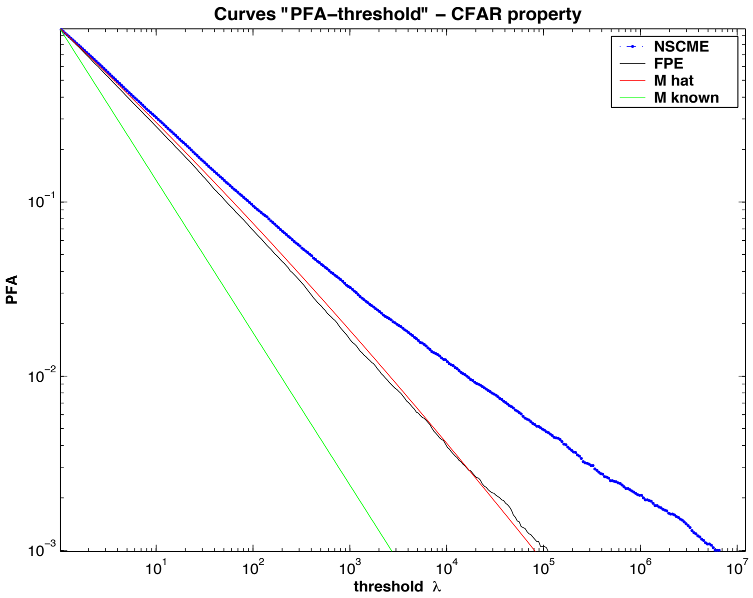
- **Independence of $\hat{\Lambda}(\widehat{\mathbf{M}})$ with the covariance matrix $\mathbf{M}$:**

$\widehat{\mathbf{M}}$	Point Fixe: $\widehat{\mathbf{M}}_{fp}$	$\widehat{\mathbf{M}}_{NSCM}$	Wishart: $\widehat{\mathbf{M}}_{\mathcal{W}}$
CFAR-matrix of $\hat{\Lambda}(\widehat{\mathbf{M}})$	**Yes**	**No**	**Yes**

$\hat{\Lambda}(\widehat{\mathbf{M}}_{fp})$ **is independent of $\mathbf{M}$ in opposite with** $\hat{\Lambda}(\widehat{\mathbf{M}}_{NSCM})$**.**

Application : Adaptive Detection Performances of the GLRT-LQ on radar data





Azimut/range bins map

Relationship "P_{fa}-threshold"

The set of parameters is $m = 8$ pulses and $N = 24$ secondary data.

THEORY AND REALITY PERFECTLY CORRESPOND.

Conclusions

- **Theoretical analysis of an improved estimator, the FPE :**
  - **Gaussian asymptotic distribution**
  - **unbiasedness**
  - **covariance matrix**
- **Same asymptotic distribution as Wishart matrix with $\frac{m}{m+1}\,N$ degrees of freedom.**
- **Very good agreements with real data.**