HAL open science

Détection de changements dans les modèles gaussiens de rang K appliquée à l'imagerie SAR

Rémi Beisson, Guillaume Ginolhac, Pascal Vallet, Audrey Giremus

To cite this version:

Rémi Beisson, Guillaume Ginolhac, Pascal Vallet, Audrey Giremus. Détection de changements dans les modèles gaussiens de rang K appliquée à l'imagerie SAR. 29° COLLOQUE SUR LE TRAITEMENT DU SIGNAL ET DES IMAGES (GRETSI 2023), Aug 2023, GRENOBLE, France. hal-04234480

HAL Id: hal-04234480

https://hal.science/hal-04234480

Submitted on 10 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Détection de changements dans les modèles gaussiens de rang K appliquée à l'imagerie SAR

Rémi BEISSON[1], Guillaume GINOLHAC[2], Pascal VALLET[1], Audrey GIREMUS[1]

[1]Laboratoire IMS (CNRS, Univ. Bordeaux, Bordeaux INP)
351 Cours de la Libération, 33400 Talence, France

[2]LISTIC (Univ. Savoie/Mont-Blanc, Polytech Annecy)
5 chemin de Bellevue, 74940 Annecy, France

{remi.beisson, pascal.vallet, audrey.giremus}@ims-bordeaux.fr,
guillaume.ginolhac@univ-smb.fr

Résumé – Dans cet article, nous proposons l'application d'une nouvelle statistique de test, basée sur un test d'égalité entre les matrices de covariance de L séries temporelles gaussiennes multivariées de dimensions M, à l'imagerie *Synthetic-aperture radar* (SAR). Cette statistique de test se place dans le cas où les L matrices de covariances sont une perturbation de rang K de l'identité, assimilable à un modèle signal plus bruit. Elle possède les propriétés de consistance et de taux de fausse alarme asymptotique constant dans le régime des grandes dimensions, lorsque la dimension des pixels M et le nombre de pixels à l'étude $N_1, \ldots, N_L$ tendent vers l'infini au même rythme. Des simulations sur des données réelles SAR, permettront de mettre en avant les bonnes performances du test développé.

Abstract – In this article we propose the application of a new test statistic, based on a test of equality between the covariance matrices of L multivariate Gaussian time series of dimensions M, to the imagery resulting from *Synthetic-aperture radar*(SAR). This test statistic is placed in the case where the L covariance matrices are a disturbance of rank K of the identity, comparable to a signal plus noise model. It has the properties of consistency and a constant asymptotic false alarm rate in the high-dimensional regime, when the dimensions of the pixels M and the number of pixels under study $N_1, \ldots, N_L$ tend towards infinity at the same rate. Simulations on SAR real data will make it possible to highlight interesting performances.

1 Introduction

Dans cette étude, on considère L séries temporelles multidimensionnelles $(\mathbf{y}_{1,n})_{n\in\mathbb{Z}}, \ldots, (\mathbf{y}_{L,n})_{n\in\mathbb{Z}}$, où $((\mathbf{y}_{\ell,n})$ est un vecteur de taille M, et que l'on suppose mutuellement indépendantes :

$$\forall \ell \in \{1, \ldots, L\}, \quad (\mathbf{y}_{n,\ell})_{n\in\mathbb{Z}} \overset{\text{i.i.d.}}{\sim} \mathcal{N}_{\mathbb{C}^M}(\mathbf{0}, \mathbf{R}_\ell)$$

Avec pour objectif de détecter les changements à travers la loi de $(\mathbf{y}_{n,\ell})_{n\in\mathbb{Z}}$ $\forall \ell \in \{1, \ldots, L\}$, on considère le test d'hypothèse binaire suivant portant sur l'égalité de matrices de covariance :

$$\begin{aligned} \mathcal{H}_0 : &\quad \mathbf{R}_1 = \ldots = \mathbf{R}_L \\ \mathcal{H}_1 : &\quad \exists (i,j) \in \{1, \ldots, L\}^2 \ / \ \mathbf{R}_i \neq \mathbf{R}_j \end{aligned} \tag{1}$$

On suppose que les observations $(\mathbf{y}_{\ell,n})_{n=1,\ldots,N_1}$ $\forall \ell \in \{1, \ldots, L\}$ sont disponibles pour chaque série temporelle et on note :

$$\hat{\mathbf{R}}_\ell := \frac{1}{N_\ell} \sum_{n=1}^{N_\ell} \mathbf{y}_{n,\ell} \mathbf{y}_{n,\ell}^*, \ \ \hat{\mathbf{R}} := \sum_{\ell=1}^{L} \frac{N_\ell}{N} \hat{\mathbf{R}}_\ell$$

où $\hat{\mathbf{R}}_\ell$ est l'estimateur empirique standard de $\mathbf{R}_\ell$ et $\hat{\mathbf{R}}$ la matrice de covariance groupée sur les L séries temporelles, avec $N = N_1 + \cdots + N_L$. On retrouve dans la littérature (cf. [1]) l'écriture suivante pour définir la forme standard des statistiques de test [1] pour (1) :

$$S = \sum_{\ell=1}^{L} \frac{N_\ell}{N} \frac{1}{M} \sum_{k=1}^{M} \varphi\left(\lambda_k(\hat{\mathbf{R}}_\ell \hat{\mathbf{R}}^{-1})\right) \tag{2}$$

où φ est une fonction continue sur $]0, +\infty[$. On retrouve notamment à partir de cette expression (cf. [1]) le *Generalised Likelihood Ratio Test* (GLRT) avec $\phi(x) = \log(x)$. On peut également souligner, dans le cas du régime asymptotique standard où $N_1, \ldots, N_L \to \infty$ et M, L sont fixes, la propriété du taux de fausse alarme constant (*CFAR*) pour les statistiques (2) qui se traduit sous $\mathcal{H}_0$ par l'indépendance de S de $\mathbf{R}_1, \ldots, \mathbf{R}_L, \mathbf{R}$ (cf. [2] pour le cas du GLRT). Cependant, un certain nombre d'applications porte sur des données de grandes dimensions où $M \gg 1$, et où les changements potentiels dans les matrices de covariance $\mathbf{R}_\ell$ se font sur une composante rang réduite [3]. On considère ainsi le modèle suivant :

$$\mathbf{R}_\ell = \mathbf{\Gamma}_\ell + \sigma^2 \mathbf{I} \tag{3}$$

avec $\mathbf{\Gamma}_\ell$ représentant la matrice de covariance, de rang $K < M$, d'un signal utile et $\sigma^2\mathbf{I}$ la matrice de covariance d'un bruit addi-

1. Pour une matrice hermitienne $\mathbf{A}$ de taille $(M \times M)$, on note $\lambda_1(\mathbf{A}) \geq \ldots \geq \lambda_M(\mathbf{A})$ ses valeurs propres par ordre décroissant.

tif spatialement blanc. On se place alors alors le régime asymptotique des grandes dimensions où $M, N_1, \ldots, N_L \to \infty$ alors que K, L restent fixes. Dans ce régime asymptotique et sous le modèle (3), $\forall \ell \neq \ell'$ les matrices $\mathbf{R}_\ell^{-1}\mathbf{R}_{\ell'}$ sont une perturbation de rang fixe de l'identité, il est alors possible de montrer que (2) converge vers une même limite sous $\mathcal{H}_0$ et $\mathcal{H}_1$. Il est alors probable que ses performances soient fortement dégradées. Dans ce cas, il peut être plus intéressant de développer un nouveau test adapté au modèle (3). En dérivant le *GLRT* ($\varphi = log$ dans (2)) pour ce modèle, on obtient la forme suivante :

$$\begin{aligned} GLRT_{\mathrm{LR}} = &-\sum_{\ell=1}^{L} N_\ell \sum_{k=1}^{K} \log\left(\frac{\lambda_k(\hat{\mathbf{R}}_\ell)}{\lambda_k(\hat{\mathbf{R}})}\right) \\ &- N(M-K)\log\left(\frac{\frac{1}{M-K}\sum\limits_{\ell=1}^{L}\frac{N_\ell}{N}\sum\limits_{k=K+1}^{M}\lambda_k(\hat{\mathbf{R}}_\ell)}{\frac{1}{M-K}\sum\limits_{k=K+1}^{M}\lambda_k(\hat{\mathbf{R}})}\right) \end{aligned} \tag{4}$$

Bien que cette statistique soit adaptée au modèle (3), sa loi sous $\mathcal{H}_0$, difficile à étudier, n'est pas connue ce qui rend très compliqué le choix d'une valeur de seuil de détection.

Dans cet article, en se basant sur le modèle (3), nous proposons donc une statistique de test qui est une généralisation, de celle présentée en [4], au cas $K > 1$ qui va se baser sur l'étude des plus grandes valeurs propres des matrices $\hat{\mathbf{R}}_1, \ldots, \hat{\mathbf{R}}_L, \hat{\mathbf{R}}$. Nous mettrons en avant le fait que cette statistique est consistante dans le *régime asymptotique des grandes dimensions*, on étudiera son comportement au second ordre et finalement nous montrerons enfin ses performances lorsque l'on utilise des données réelles *SAR*.

2 Statistique de Test

En repartant des travaux sur le spectre de $\hat{\mathbf{R}}$ (cf. Partie 2 [4]), nous pouvons généraliser le Théorème 2 au cas $K > 1$. On établit ainsi un théorème centrale limite conjoint sur les plus grandes valeurs propres de $\hat{\mathbf{R}}_1, \ldots, \hat{\mathbf{R}}_L, \hat{\mathbf{R}}$:

Théorème 1. *Sous $\mathcal{H}_0$ et l'hypothèse 2 de [4], on a :*

$$\sqrt{M}\begin{pmatrix} \lambda_k\left(\hat{\mathbf{R}}\right) - \phi_c(\gamma_k, \sigma^2) \\ \left(\lambda_k\left(\hat{\mathbf{R}}_\ell\right) - \phi_{c_\ell}(\gamma_{k,\ell}, \sigma^2)\right)_{\ell=1,\ldots,L} \end{pmatrix}_{k=1,\ldots,K} \xrightarrow[M\to\infty]{\mathcal{D}} \mathcal{N}_{\mathbb{R}^{K(L+1)}}\left(\mathbf{0}, \mathbf{\Theta}\right),$$

où $\mathbf{\Theta} := \mathrm{bdiag}\left(\mathbf{\Theta}_1, \ldots, \mathbf{\Theta}_K\right)$ avec :

$$\mathbf{\Theta}_k := \begin{pmatrix} \theta_{k,0}^2 & \vartheta_{k,1} & \ldots & \vartheta_{k,L} \\ \vartheta_{k,1} & \ddots & & \\ \vdots & & \ddots & \\ \vartheta_{k,L} & & & \theta_{k,L}^2 \end{pmatrix}$$

et en notant $c_0 = c$,

$$\theta_{k,\ell}^2 = c_\ell \frac{(\gamma_k^2 - \sigma^4 c_\ell)(\gamma_k + \sigma^2)^2}{\gamma_k^2}, \quad \ell \geq 0$$

$$\vartheta_{k,\ell} = c_0 \frac{(\gamma_k^2 - \sigma^4 c_\ell)(\gamma_k + \sigma^2)^2}{\gamma_k^2}, \quad \ell \geq 1$$

Nous allons maintenant exploiter les résultats de la partie 2 de [4] et du Théorème 1 afin de construire notre statistique de test.

Il est possible sous le modèle (3) de changer la nature du test (1) en utilisant le résultat du Lemme suivant.

Lemme 1. *Les assertions suivantes sont équivalentes :*
(a) $\mathbf{\Gamma}_1 = \ldots = \mathbf{\Gamma}_L$
(b) For all $k = 1, \ldots, K$, $\ell = 1, \ldots, L$, $\lambda_k\left(\mathbf{\Gamma}_\ell\right) = \lambda_k\left(\mathbf{\Gamma}\right)$.
Donc sous $\mathcal{H}_1$, il existe k, ℓ tel que $\gamma_k \neq \gamma_{k,\ell}$.

Ce Lemme souligne qu'un changement dans le signal et donc dans les valeurs propres de $\mathbf{\Gamma}_1, \ldots, \mathbf{\Gamma}_L$ implique un changement dans celles de $\mathbf{\Gamma}$. On peut donc ramener notre étude au test suivant :

$$\begin{aligned} \mathcal{H}_0 : &\quad \forall k, \ell,\ \lambda_k\left(\mathbf{\Gamma}_\ell\right) = \lambda_k\left(\mathbf{\Gamma}\right) \\ \mathcal{H}_1 : &\quad \exists k, \ell : \lambda_k\left(\mathbf{\Gamma}_\ell\right) \neq \lambda_k\left(\mathbf{\Gamma}\right) \end{aligned} \tag{5}$$

Ainsi en se basant sur les résultats du Théorème 1 de [4] on développe un nouveau test de détection en utilisant des estimateurs consistant des valeurs propres $\gamma_{k,\ell}, \gamma_k$ $\forall k \in \{1, \ldots, K\}, \ell \in \{1, \ldots, L\}$. On posera donc $\hat{\gamma}_k$ la plus grande solution de l'équation $\phi_c(\gamma_k, \sigma^2) = \lambda_k(\hat{\mathbf{R}})$ si $\lambda_k(\hat{\mathbf{R}}) > \sigma^2(1 + \sqrt{c})^2$ ou $\hat{\gamma}_k = \sigma^2 c$ sinon. On suivra le même raisonnement pour $\hat{\gamma}_{k,\ell}$ en prenant soin de remplacer c, $\hat{\mathbf{R}}$ par c_ℓ, $\hat{\mathbf{R}}_\ell$ et on considérera $\hat{\sigma}^2$ l'estimateur de σ^2 suivant :

$$\hat{\sigma}^2 := \sum_{\ell=1}^{L} \frac{N_\ell}{N} \frac{1}{M-K} \sum_{k=K+1}^{M} \lambda_k\left(\hat{\mathbf{R}}_\ell\right).$$

En posant par la suite : $\hat{\boldsymbol{\gamma}} = (\hat{\gamma}_k - \hat{\gamma}_{k,\ell})_{\substack{k=1,\ldots,K \\ \ell=1,\ldots,L}}$ et en appliquant une *Delta-Methode* à partir des résultats de la partie 2 de [4] et du Théorème 1 on montre sous $\mathcal{H}_0$:

Corollaire 1.

$$\sqrt{M}\hat{\boldsymbol{\gamma}} \xrightarrow[M\to\infty]{\mathcal{D}} \mathcal{N}_{\mathbb{R}^{KL}}\left(\mathbf{0}, \mathbf{H}\mathbf{\Omega}\mathbf{H}^T\right)$$

où $\mathbf{H} = \mathrm{bdiag}\left(\tilde{\mathbf{H}}, \ldots, \tilde{\mathbf{H}}\right)$, $\mathbf{\Omega} = \mathrm{bdiag}\left(\mathbf{\Omega}_1, \ldots, \mathbf{\Omega}_K\right)$ avec $\tilde{\mathbf{H}}$ la matrice $L \times (L+1)$ donné par :

$$\tilde{\mathbf{H}} = \begin{pmatrix} 1 & -1 & 0 & \ldots & \ldots & 0 \\ 1 & 0 & -1 & \ddots & & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & & \ddots & \ddots & 0 \\ 1 & 0 & \ldots & \ldots & 0 & -1 \end{pmatrix}$$

et avec

$$\mathbf{\Omega}_k = \begin{pmatrix} \omega_{k,0}^2 & \xi_k & \ldots & \xi_k \\ \xi_k & \ddots & & \\ \vdots & & \ddots & \\ \xi_k & & & \omega_{k,L}^2 \end{pmatrix}$$

où $\omega_{k,\ell}^2 = \frac{c_\ell \gamma_k^2 (\gamma_k + \sigma^2)^2}{\gamma_k^2 - \sigma^4 c_\ell}$ et $\xi_k = \frac{c_0 \gamma_k^2 (\gamma_k + \sigma^2)^2}{\gamma_k^2 - \sigma^4 c_0}$.

A partir des résultats évoqués précédemment, on propose la statistique de test suivante :

$$T_\epsilon = \mathbb{1}_{(\epsilon,+\infty)}\left(\|\hat{\boldsymbol{\gamma}}\|_2^2\right) \tag{6}$$

avec ϵ le seuil de contrôle sur la probabilité de fausse alarme asymptotique choisi en conséquence du résultat suivant :

$$\|\hat{\boldsymbol{\gamma}}\|_2^2 \xrightarrow[M\to\infty]{\text{a.s.}} \begin{cases} 0 & \text{sous } \mathcal{H}_0 \\ \epsilon_1 := \|\boldsymbol{\gamma}\|_2^2 > 0 & \text{sous } \mathcal{H}_1, \end{cases}$$

avec $\boldsymbol{\gamma} = (\gamma_k - \gamma_{k,l})_{\substack{k=1,\dots,K \\ \ell=1,\dots,L}}$. La statistique de test proposée est donc consistante comme l'atteste le Théorème suivant.

Théorème 2. *Sous les Hypothèses 1, 2 (cf. [4]), et* $\forall \epsilon \in]0, \epsilon_1[$,

$$\mathbb{P}_i\left(\lim_{M\to\infty} T_\epsilon = i\right) = 1,$$

ou $\mathbb{P}_i$ *est la mesure de probabilité sous l'hypothèse* $\mathcal{H}_i$.

Ce nouveau test est intéressant en pratique car puisque nous sommes en mesure d'obtenir $\hat{\boldsymbol{\Omega}}$ un estimateur consistant de $\boldsymbol{\Omega}$, en utilisant le Corollaire 1 et le Théorème 2 il est possible de fixer un seuil de détection en fonction d'une probabilité de fausse alarme.

3 Résultats sur données *SAR*

Nous présentons tout d'abord les résultats sur données simulées. Afin de nous rapprocher le plus possible de ce que l'on peut retrouver dans les données réelles qui vont suivre, nous prenons $L = 2$ puis $M = 12$, $N = 50$, $N_1 = N_2 = 25$ et $K = 5$. En ce qui concerne la définition des matrices de covariance $\boldsymbol{\Gamma}_\ell$ on pose :

$$\forall \ell \in \{1,2\} \;\; \boldsymbol{\Gamma}_\ell = SNR \times \mathbf{G}_\ell + \mathbf{I}_M$$

avec SNR une constante pour agir sur le ratio signal sur bruit et la décomposition en valeur singulière de la matrice *Toeplitz* symétrique $\mathbf{T}_\ell$ ($T_{\ell_{i,j}} = \rho_\ell^{|i-j|}$) suivante $\mathbf{T}_\ell(\rho_\ell) = \mathbf{U}_\ell\mathbf{S}_\ell\mathbf{V}_\ell$ qui donne ensuite $\mathbf{G}_\ell = [\mathbf{U}_\ell(:,1:K)\mathbf{S}_\ell(1:K,1:K)\mathbf{U}_\ell(:,1:K)^*]$. Nous considérons sous $\mathcal{H}_0$, $\rho_1 = \rho_2 = 0.1$ puis sous $\mathcal{H}_1$ $\rho_1 = 0.1$ et $\rho_2 = 0.5$.
Dans la Figure 1 nous faisons varier le ratio SNR puis dans la Figure 2 nous sur-estimons ou sous-estimons K dans l'expression des statistiques afin de tester la robustesse face à une mauvaise estimation du rang. La statistique de test T_ϵ s'avère plus résistante au bruit et très peu sensible à une mauvaise estimation du rang ce qui n'est pas le cas du $GLRT_{LR}$.

On s'intéresse maintenant au *dataset* qui est une série d'images *SAR* ([5], [6]) mise à disposition par *NASA/JPL-Caltech*[2]. Nous considérons dans ce papier 2 scènes différentes de taille (2360 × 600) pixels pour la première et (2300 × 600) pixels pour la seconde. Pour chacune de ces scènes la résolution en azimuth est de 0.6m et en distance de 1.67m. Le nombre d'images pour chaque scène est $L = 2$ et chaque image est en polarisation totale (HH, VV et HV). Comme nous sommes en

2. UAVSAR (SanAnd 26524 03 Segment 4), https ://uavsar.jpl.nasa.gov

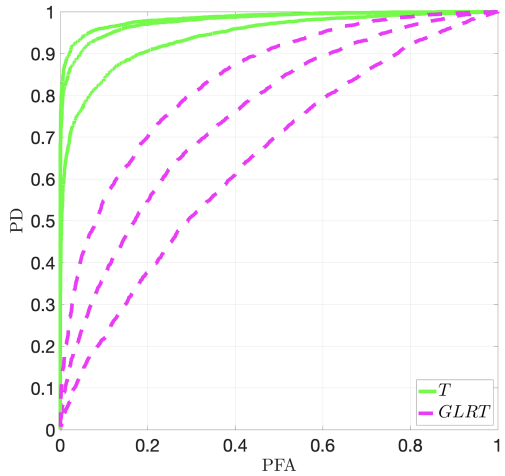

FIGURE 1 – Courbes ROC, $SNR_{dB} \in \{3, 7, 10\}$

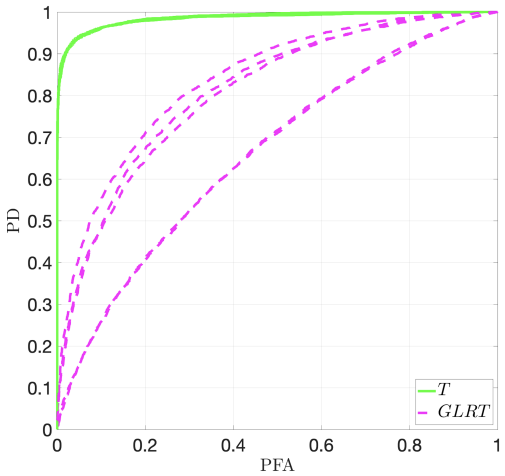

FIGURE 2 – Courbes ROC, $K \in \{3, 4, 5, 6, 7\}$

haute résolution, il est possible d'utiliser une décomposition en ondelette [7] pour mettre en valeur les propriétés spectro-angulaires des diffuseurs. Finalement, la taille de chaque pixel sera de $M = 12$. La mise en évidence des performances des tests $GLRT_{LR}$ et T_ϵ se fait à l'aide du calcul des courbes *ROC* (*Receiver Operating Characteristic*).
Pour toutes les simulations qui suivent, après avoir étudié les données en détails, nous avons décidé de prendre pour le rang $K = 5$. Nous considérons également $N_1 = N_2 = 25$ et $N = 50$.

On présente en Figures 3 et 4 la représentation en *RGB Pauli* des 2 scènes pour la date la plus ancienne (gauche) et la plus récente (milieu) ainsi que la vérité terrain associée (droite) ou les pixels ayant subi un changement sont en blanc.
On observe sur les Figures 5 et 6 les performances pour les deux statistiques avec un avantage net pour T_ϵ. On peut remarquer une différence de performance entre les deux simulations pour la statistique T_ϵ qui peut être justifier par la justesse de la vérité terrain qui peut faire varier les performances de la statistique. On notera également que ces résultats sont obtenus pour

FIGURE 3 – Scène 1 : 23 avril 2009 (gauche), 15 mai 2011 (milieu) et vérité terrain associée (droite).

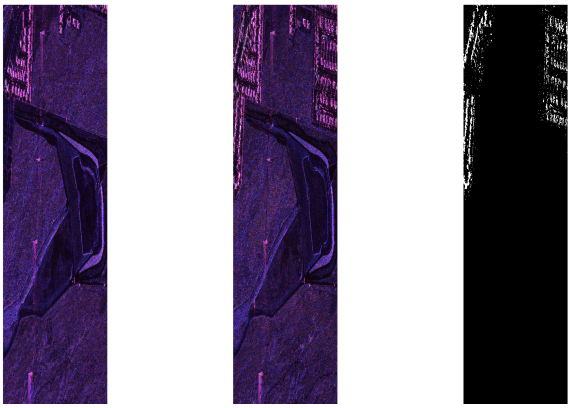

FIGURE 4 – Scène 2 : 23 avril 2009 (gauche), 15 mai 2011 (milieu) et vérité terrain associée (droite).

des valeurs M, N_1, N_2, N faibles à modérées alors que l'analyse statistique conduisant au nouveau détecteur suppose être en grande dimension.

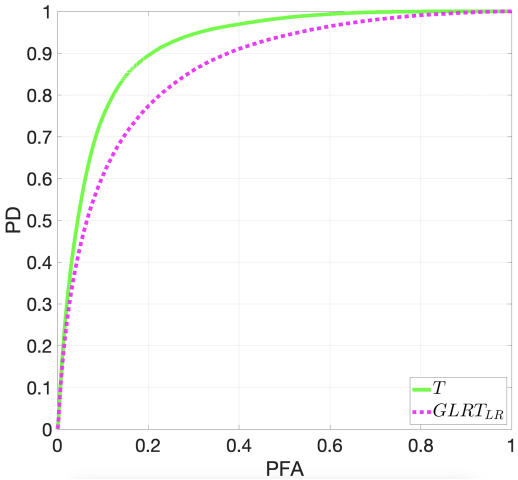

FIGURE 5 – Courbes *ROC* pour la Scène 1

En plus de son bon comportement, il est important de rappeler que la statistique de test T_ϵ est plus avantageuse du fait que nous connaissions sa loi sous $\mathcal{H}_0$, ce qui nous permet en pratique de fixer un seuil de détection en fonction d'une probabilité de fausse alarme.

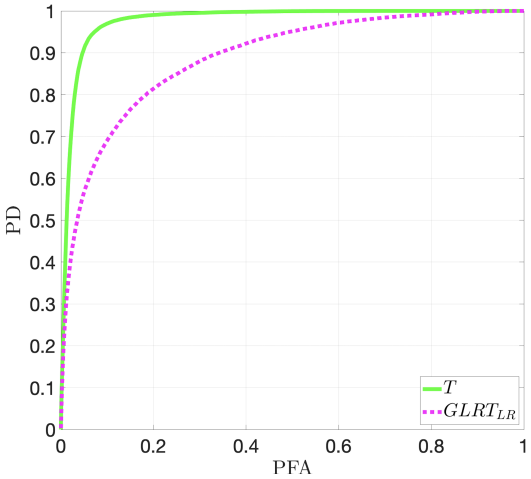

FIGURE 6 – Courbes *ROC* pour la Scène 2

4 Conclusion

Dans cet article nous avons proposé une statistique de test alternative au $GLRT_{LR}$, adaptée à un modèle de matrice de covariance à rang réduit. Celle-ci s'avère plus performante sur des données réelles issues de *SAR* que le test classique $GLRT_{LR}$. Il est intéressant de noter que les résultats sont obtenus pour des dimensions modérées voir faibles alors que son étude théorique est basée sur le *régime des grandes dimensions*.

Références

[1] D. Ciuonzo, V. Carotenuto, and A. De Maio. On multiple covariance equality testing with application to SAR change detection. *IEEE Trans. on Signal Process.*, 65(19) :5078–5091, 2017.

[2] R.J. Muirhead. *Aspects of multivariate statistical theory*. Wiley Online Library, 1982.

[3] I.M. Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.*, 29(2) :295–327, 04 2001.

[4] Rémi BEISSON, Pascal VALLET, Audrey GIREMUS, and Guillaume GINOLHAC. Détection de changements dans les modèles gaussiens de rang 1. Number 001-0242, Nancy, 2022. GRETSI - Groupe de Recherche en Traitement du Signal et des Images.

[5] A. Mian, G. Ginolhac, J.-P. Ovarlez, and A. Atto. New robust statistics for change detection in time series of multivariate SAR images. *IEEE Trans. Signal Process.*, 67(2) :520–534, 2018.

[6] R. Abdallah, A. Mian, A. Breloy, A. Taylor, M. El Korso, and D. Lautru. Detection methods based on structured covariance matrices for multivariate SAR images processing. *IEEE Geosci. Remote. Sens. Lett.*, 16(7) :1160–1164, 2019.

[7] A. Mian, J-P. Ovarlez, A. M. Atto, and G. Ginolhac. Design of new wavelet packets adapted to high-resolution SAR images with an application to target detection. *IEEE Transactions on Geoscience and Remote Sensing*, 57(6) :3919–3932, June 2019.