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

Rémi BEISSON[1], Pascal VALLET[1], Guillaume GINOLHAC[2], 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 $(\mathbf{y}_{n,1})_{n\in\mathbb{Z}}, \ldots, (\mathbf{y}_{n,L})_{n\in\mathbb{Z}}$, où $(\mathbf{y}_{n,\ell})$ 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 $(\mathbf{y}_{n,\ell})_{n=1,\ldots,N_\ell}$ $\forall \ell \in \{1, \ldots, L\}$ sont disponibles pour chaque série temporelle et on note :

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

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 $N = N_1 + \cdots + N_L$. La majorité des statistiques de tests utilisées dans la littérature (cf. [1]) sont de la forme [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 test du ratio de vraisemblance (GLRT) avec $\phi(x) = \log(x)$. De plus les statistiques de test basées sur (2) possèdent la propriété du taux de fausse alarme constant *CFAR* qui se traduit sous $\mathcal{H}_0$ par l'indépendance de S de $\mathbf{R}_1, \ldots, \mathbf{R}_L, \mathbf{R}$. Les approximations faites dans le *régime asymptotique standard* où $N_1, \ldots, N_L \to \infty$ et M, L sont fixes, permettent d'estimer la distribution des statistiques de type (2) sous $\mathcal{H}_0$ (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 $K < M$,

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.

d'un signal utile et $\sigma^2\mathbf{I}$ la matrice de covariance d'un bruit additif spatialement blanc. On se place alors dans 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 $\mathcal{H}_1$ ce qui empêche de distinguer les deux hypothèses. 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 une nouvelle expression que l'on donne en (7). Il est précisé dans [4] que dans le *régime asymptotique des grandes dimensions* sa loi sous $\mathcal{H}_0$ dépend de paramètres inconnus ce qui ne permet pas de valider la propriété *CFAR*.

Ainsi dans cet article, en se basant sur le modèle (3), nous proposons une statistique de test qui est une généralisation de celle présentée en [5], 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 qui nous permettra de justifier la propriété *CFAR* et enfin nous mettrons en avant ses performances sur des données simulées puis réelles (*SAR*).

2 Statistique de Test

Avant de pouvoir présenter la statistique de test, il convient de définir le *régime asymptotique des grandes dimensions* nécessaire à sa construction.

Hypothèse 1. *$N_1 = N_1(M), \ldots, N_L = N_L(M)$ sont fonctions de M telles que :*

$$\frac{M}{N_\ell} = c_\ell + o\left(\frac{1}{\sqrt{M}}\right)$$

où $c_1, \ldots, c_L \in]0, 1]$ avec L indépendant de M. On représentera par la suite ce régime asymptotique par $\xrightarrow[M\to\infty]{}$.

Considérons également les notations et l'hypothèse qui suivent.
Notations. $c := \left(\sum_{\ell=1}^{L} \frac{1}{c_\ell}\right)^{-1}, \ \ \mathbf{\Gamma} := \sum_{\ell=1}^{L} \frac{N_\ell}{N}\mathbf{\Gamma}_\ell$

Hypothèse 2. *Pour tout $k \in \{1, \ldots, K\}$, $\ell \in \{1, \ldots, L\}$:*

$$\lambda_k(\mathbf{\Gamma}_\ell) = \gamma_{k,\ell} + o\left(\frac{1}{\sqrt{M}}\right)$$

et $\forall k \in \{1, \ldots, KL\}$

$$\lambda_k(\mathbf{\Gamma}) = \gamma_k + o\left(\frac{1}{\sqrt{M}}\right)$$

On peut ensuite, reformuler (1) de manière équivalente sous la forme suivante (cf. [4]) :

$$\begin{array}{ll} \mathcal{H}_0 : & \forall k, \ell, \ \lambda_k(\mathbf{\Gamma}_\ell) = \lambda_k(\mathbf{\Gamma}) \\ \mathcal{H}_1 : & \exists k, \ell : \lambda_k(\mathbf{\Gamma}_\ell) \neq \lambda_k(\mathbf{\Gamma}) \end{array} \tag{4}$$

Par conséquent, il est possible de discriminer les hypothèses $\mathcal{H}_0$ et $\mathcal{H}_1$ en exploitant uniquement les valeurs propres des matrices $\mathbf{\Gamma}_1, \ldots, \mathbf{\Gamma}_L, \mathbf{\Gamma}$.
En se basant sur les résultats du Théorème 1 de [5] on développe un nouveau test de détection en utilisant des estimateurs consistants, dans le *régime asymptotique des grandes dimensions*, des valeurs propres $\gamma_k, \gamma_{k,\ell}$. On considérera $\hat{\sigma}^2$ l'estimateur de σ^2 suivant :

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

puis $\hat{\gamma}_k$ la plus grande solution de l'équation $\frac{(\gamma_k+\hat{\sigma}^2)(\gamma_k+\hat{\sigma}^2 c)}{\gamma_k} = \lambda_k(\hat{\mathbf{R}})$ si $\lambda_k(\hat{\mathbf{R}}) > \hat{\sigma}^2(1+\sqrt{c})^2$ ou $\hat{\gamma}_k = \hat{\sigma}^2\sqrt{c}$ sinon. On suivra le même raisonnement pour $\hat{\gamma}_{k,\ell}$ en prenant soin de remplacer c, $\hat{\mathbf{R}}$ par c_ℓ, $\hat{\mathbf{R}}_\ell$. En posant par la suite $\hat{\boldsymbol{\gamma}} = (\hat{\gamma}_k - \hat{\gamma}_{k,\ell})_{\substack{k=1,\ldots,K \\ \ell=1,\ldots,L}}$, on propose la statistique de test suivante :

$$T_\epsilon = \mathbb{1}_{(\epsilon,+\infty)}\left(\|\hat{\boldsymbol{\gamma}}\|_2^2\right) \tag{5}$$

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

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

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

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

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

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

Sa loi sous $\mathcal{H}_0$ est également connue et détaillée ci-après.

Théorème 2. *Sous $\mathcal{H}_0$,*

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

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

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

et avec

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

où

$$\omega_{k,\ell}^2 = \frac{c_\ell \gamma_k^2(\gamma_k+\sigma^2)^2}{\gamma_k^2 - \sigma^4 c_\ell} \text{ et } \xi_k = \frac{c_0 \gamma_k^2(\gamma_k+\sigma^2)^2}{\gamma_k^2 - \sigma^4 c_0} \tag{6}$$

Ce nouveau test est intéressant en pratique. En effet, étant en mesure d'obtenir $\hat{\mathbf{\Omega}}$ un estimateur consistant dans le *régime des grandes dimensions* de $\mathbf{\Omega}$, en remplaçant $\gamma_k, \gamma_{k,\ell}, \sigma^2$ par $\hat{\gamma}_k, \hat{\gamma}_{k,\ell}, \hat{\sigma}^2$, il est possible par le biais les Théorèmes 1 et 2 d'échantillonner la loi gaussienne de covariance $\mathbf{H}\hat{\mathbf{\Omega}}\mathbf{H}^T$. Il est alors envisageable de fixer un seuil $\hat{\epsilon}$ de telle sorte que $\mathbb{P}_{\mathcal{H}_0}(T_{\hat{\epsilon}} = 1) \xrightarrow[M\to\infty]{} \beta$ avec β le taux de fausse alarme.

3 Résultats sur données simulées et *SAR*

On présente tout d'abord les résultats sur données simulées. Afin de se rapprocher le plus possible de ce que l'on peut retrouver dans les données réelles qui vont suivre, on prend $L = 2$ puis $M = 12$, $N = 50$, $N_1 = N_2 = 25$, $\sigma^2 = 1$ et $K = 5$. On construit ensuite, pour chaque série temporelle, une matrice *Toeplitz* symétrique $\mathbf{T}_\ell$ de taille $(M \times M)$ telle que $T_{\ell_{i,j}} = \alpha\rho_\ell^{|i-j|}$ avec α une constante pour agir sur le ratio signal sur bruit. On en prend sa décomposition en valeurs singulières $\mathbf{T}_\ell = \alpha\sum_{m=1}^{M}\gamma_{m,l}\mathbf{u}_{m,l}\mathbf{u}_{m,l}^*$. On définit alors les matrices de covariance $\mathbf{\Gamma}_\ell$ comme suit :

$$\forall \ell \in \{1,2\} \;\; \mathbf{\Gamma}_\ell = \sum_{k=1}^{K}\gamma_{k,l}\mathbf{u}_{k,l}\mathbf{u}_{k,l}^*$$

On pose sous $\mathcal{H}_0$ $\rho_1 = \rho_2 = 0.1$ puis sous $\mathcal{H}_1$ $\rho_1 = 0.1$ et $\rho_2 = 0.5$. On comparera les performances de T_ϵ avec le $GLRT_{LR}$ évoqué en introduction dont on fournit ci-après une expression :

$$\begin{aligned} GLRT_{\mathrm{LR}} = &-\sum_{\ell=1}^{L}N_\ell\sum_{k=1}^{K}\log\left(\frac{\lambda_k(\hat{\mathbf{R}}_\ell)}{\lambda_k(\hat{\mathbf{R}})}\right) \\ &- N(M-K)\log\left(\frac{\frac{1}{M-K}\sum\limits_{\ell=1}^{L}\frac{N_\ell}{N}\sum\limits_{k=K+1}^{M}\lambda_k(\hat{\mathbf{R}}_\ell)}{\frac{1}{M-K}\sum\limits_{k=K+1}^{M}\lambda_k(\hat{\mathbf{R}})}\right) \end{aligned} \quad (7)$$

La Figure 1 illustre l'impacte du ratio signal sur bruit α. On constate des performances plus avantageuses pour la statistique de test T_ϵ et une dégradation des performances plus lente avec α qui diminue. Dans la Figure 2 on surestime ou sous-estime 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 très peu sensible à une mauvaise estimation du rang qui ne change que très peu les performances contrairement au $GLRT_{LR}$ qui subit fortement la sous-estimation de K.

On s'intéresse maintenant au *dataset* qui est une série d'images *SAR* ([6], [7]) mise à disposition par *NASA/JPL-Caltech* [2]. On considère 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

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

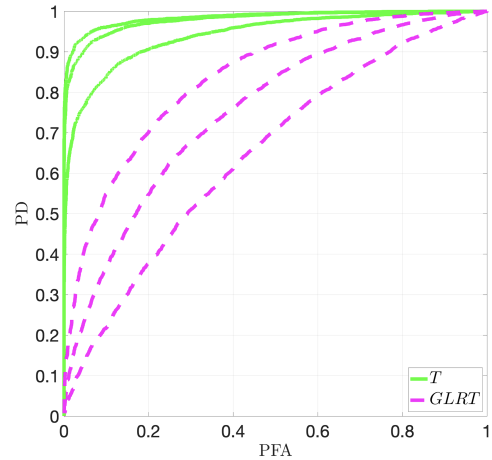

FIGURE 1 – Courbes ROC, $\alpha_{dB} \in \{3, 7, 10\}$

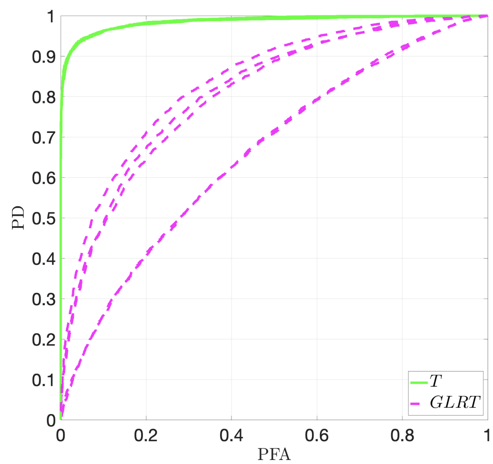

FIGURE 2 – Courbes ROC, $\hat{K} \in \{3, 4, 5, 6, 7\}$ et $\alpha_{dB} = 10$

d'images pour chaque scène est $L = 2$ et chaque image est en polarisation totale (HH, VV et HV). Étant en haute résolution, il est possible d'utiliser une décomposition en ondelette [8] pour mettre en valeur les propriétés spectro-angulaires des diffuseurs. Finalement, la dimension 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, le rang $K = 5$ a été choisi. On considère 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) où 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 justifiée 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 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. Il est important de rappeler que la statistique de test T_ϵ est plus avantageuse du fait que sa loi asymptotique sous $\mathcal{H}_0$ soit connue, ce qui permet en pratique de fixer un seuil de détection en fonction d'une PFA.

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

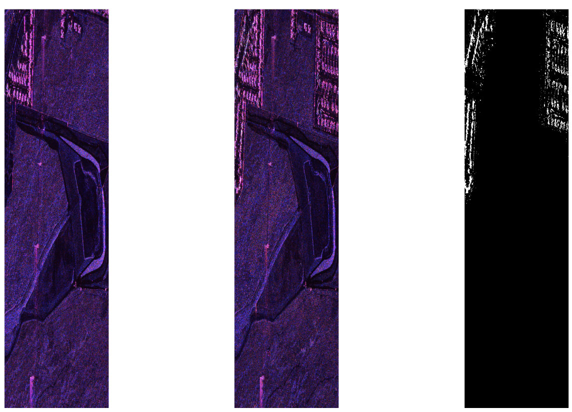

FIGURE 4 – Scène 2 : 23 avril 2009 (gauche), 15 mai 2011 (milieu) et vérité terrain associée (droite).

4 Conclusion

Dans cet article nous avons proposé une statistique de test pour détecter un changement dans une covariance rang réduit, alternative à celle du ratio de vraisemblance. Cette statistique a été étudiée dans le régime des grandes dimensions, en terme de consistance et de loi asymptotique, qui permet entre autre de contrôler son taux de fausse alarme asymptotique. Les performances observées sur données synthétiques et réelles montrent une amélioration par rapport au ratio de vraisemblance, ainsi qu'une certaine robustesse à la mauvaise connaissance de certains paramètres comme le rang, et tout ceci pour des dimensions modérées.

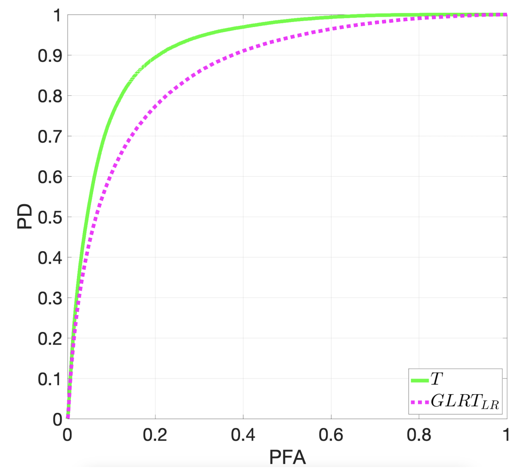

FIGURE 5 – Courbes *ROC* pour la Scène 1

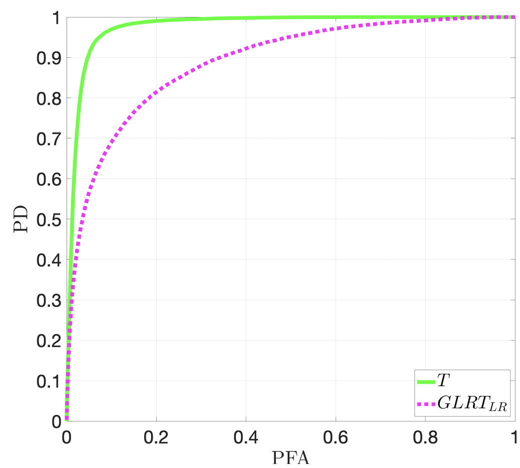

FIGURE 6 – Courbes *ROC* pour la Scène 2

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. Change detection in the covariance structure of high-dimensional gaussian low-rank models. In *2021 IEEE Statistical Signal Processing Workshop (SSP)*, pages 421–425. IEEE, 2021.

[5] 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.

[6] 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.

[7] 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.

[8] 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.