

# Apprentissage explicable d'un ensemble de divergences pour la similarité inter-classe de données SAR

Matthieu GALLET<sup>1,2</sup> Abdourrahmane ATTO<sup>1</sup> Emmanuel TROUVÉ<sup>1</sup> Fatima KARBOU<sup>2</sup>

<sup>1</sup>Université Savoie Mont-Blanc, LISTIC, Annecy, France

<sup>2</sup>Météo France, CEN, Grenoble, France

**Résumé** – Dans ce travail, nous proposons une méthode de classification de données bivariées positives en utilisant un ensemble de divergences basées sur des statistiques descriptives, une modélisation paramétrique des marginales et la modélisation des dépendances via des copules. Ces divergences sont combinées via des opérateurs de réseaux neuronaux pour déterminer la classe du couple d'entré considéré. Cette approche est testée sur un dataset d'images radars à ouverture synthétique (SAR) du satellite PAZ.

**Abstract** – In this work, we propose a method for classification of positive bivariate data using a set of divergences based on statistics, parametric marginal modeling and dependence modeling via copulas. These divergences are combined using neural network operators to determine the class of the input couple being considered. This approach is tested on a dataset of synthetic aperture radar (SAR) images from the PAZ satellite.

## 1 Introduction

La tâche de reconnaissance d'états de surface ou classification d'images SAR reste un challenge tant en termes de performance que sur la capacité des modèles et schémas d'analyse à être interprétables. En effet, le bruit et les distorsions présents dans les données SAR rendent difficiles l'interprétation des résultats et la labellisation des données. De nombreux travaux se sont intéressés à cette tâche en utilisant notamment les réseaux de neurones ou des approches non paramétriques, qui, en dépit de leur efficacité dans de nombreux domaines, restent difficilement explicables et explorent un espace de paramètres très large [10]. Les modèles paramétriques ou semi-paramétriques, quant à eux, offrent une interprétabilité plus aisée et une capacité à être adaptés à des problématiques spécifiques [12]. Dans le domaine de l'imagerie, les moments statistiques [2] ou la modélisation de marginales [11] ont été largement utilisés, en particulier pour la tâche de classification. Cette dépendance peut être modélisée de manière plus complète à l'aide de copules. Les travaux de [8] ont montré que les copules peuvent capturer les dépendances entre différentes modalités de données de télédétection pour détecter des changements. Dans ce contexte, nous proposons de réaliser une classification de couples d'images en utilisant un ensemble de divergences basées sur des modèles paramétriques estimés sur des données positives de type SAR. Les divergences sont ensuite combinées dans une étape d'apprentissage neuronale superficiel, en contraignant la positivité des poids. Pour cela, on propose deux nouvelles divergences de Kullback Leibler, l'une basée sur la troncation de distributions gaussiennes et l'autre sur les copules Eyraud-Farlie-Gumbel-Morgenstern (EFGM) [3]. Nous construisons dans la Section 2 l'ensemble de divergences. Nous utilisons dans un premier temps dans la Partie 2.1 des distances entre statistiques descriptives usuelles pour l'analyse de données SAR. Nous nous intéressons dans la Partie 2.2, à l'ajout de divergences de Kullback Leibler obtenue par une modélisation des marginales des entrées. Enfin, nous complétons notre ensemble, par la modélisation de la dépendance des

canaux via les copules dans la Partie 2.3. Nous présentons dans la Partie 3.1, l'apprentissage de la combinaison optimale de cet ensemble à partir d'opérateurs de réseaux neuronaux tels que les couches denses et les activations dans un cadre de classification supervisé. Finalement, nous testons cette approche sur des données SAR en bande X du satellite PAZ dans la Partie 3.2. Nous nous concentrons sur la classification des paires d'images, afin de déterminer l'appartenance du couple à l'une des deux classes étudiées ou d'identifier une divergence de classes entre les images de la paire.

## 2 Ensemble de divergences

### 2.1 Mesures de statistiques descriptives

En considérant un ensemble de  $n$  observations  $x_1, \dots, x_n$ , appartenant à une variable aléatoire  $X$  de loi  $f_X$  et de support  $[0; +\infty)$ , on s'intéresse aux mesures de statistiques descriptives de  $X$  suivantes : la moyenne arithmétique  $\mu$ , l'écart-type  $\sigma$  et le coefficient de variation CV. Ce dernier est défini par :  $CV = \sigma/\mu$ . L'utilisation de ces mesures de statistiques descriptives est motivée par le fait qu'elles sont significatives dans l'étude du signal SAR. La moyenne est ainsi un coefficient de rétrodiffusion moyen à l'échelle de la fenêtre considérée, qui peut servir à discriminer des zones de nature différente [2], de même que le coefficient de variation qui permet de caractériser l'homogénéité de la rétrodiffusion dans une zone. Dans ce cadre, on caractérise la distance  $\mathcal{D}$  entre deux distributions  $f_X$  et  $f_Y$  par la norme  $\ell_2$  entre chacune de ces mesures de statistiques descriptives.

### 2.2 Distribution marginales positives

Soit  $\mathbf{1}_S$  la fonction indicatrice du set  $S$ ,  $\Gamma(\cdot)$  et  $\Gamma(\cdot, \cdot)$  les fonctions spéciales respectivement appelées Gamma et Gamma incomplète supérieure.

Dans ce travail, nous considérons des fonctions de densité de probabilité (PDF) paramétriques ayant un support sur  $\mathbb{R}^+$ . Ces

modèles de PDF seront utilisés pour décrire des données positives associées à l'intensité ou l'amplitude de données SAR, ou encore à des caractéristiques de modèle CNN. On peut considérer l'opération de rectification propre aux fonctions d'activation des réseaux de neurones, qui vise à mettre à 0 les parties négatives, ou celle de troncation qui réduit l'intervalle d'étude entre 2 bornes  $a$  et  $b$ . Dans la suite de ce travail nous considérons uniquement la troncation d'une fonction de densité de probabilité  $g$  sur le support  $]0; +\infty)$ . Cette transformation se fait en deux étapes :

- considérer la restriction  $g_+$  de  $g$  définie par  $g_+ = g \mathbf{1}_{\mathbb{R}^+}$  ;
- calculer la PDF tronquée comme  $f = \lambda g_+$  où  $\lambda$  correspond à un paramètre de normalisation spécifié pour satisfaire la condition de somme unitaire sur  $f$ .

À partir de cette formulation de troncation, et en considérant la distribution Gaussienne généralisée (GGau) comme  $g$ , nous obtenons :

**Définition 1** (GGau tronquée : RGGau). *La RGGau associée avec un paramètre de position ou décalage  $\mu \in \mathbb{R}$ , un paramètre d'échelle  $\sigma > 0$  et un paramètre de forme  $\beta > 0$  est définie par :*

$$f_{\mu, \sigma, \beta}^{RGGau}(x) = \frac{1}{\Lambda(\mu, \sigma, \beta)} e^{-\left(\frac{x-\mu}{\sigma}\right)^\beta} \mathbf{1}_{\mathbb{R}^+}(x), \quad (1)$$

où  $\Lambda(\mu, \sigma, \beta)$  est la fonction de normalisation de la RGGau définie par :

**Proposition 1** (Constante de normalisation de la RGGau). *Pour le modèle RGGau défini par (1), on a :*

$$\Lambda(\mu, \sigma, \beta) = \begin{cases} \sigma \Gamma \left[ 1 + \frac{1}{\beta} \right] + \frac{\sigma}{\beta} \left( \Gamma \left[ \frac{1}{\beta} \right] - \Gamma \left[ \frac{1}{\beta}, \left( \frac{\mu}{\sigma} \right)^\beta \right] \right) & \text{si } \mu \geq 0 \\ \frac{\sigma}{\beta} \left( \Gamma \left[ \frac{1}{\beta}, \left( \frac{-\mu}{\sigma} \right)^\beta \right] \right) & \text{sinon} \end{cases} \quad (2)$$

En plus de (1), nous considérons la classe des distributions Weibull (Wb) dont le support par construction est  $\mathbb{R}^+$ .

**Définition 2** (Weibull PDF). *Une distribution de Weibull avec un paramètre d'échelle  $\lambda > 0$  et un paramètre de forme  $\beta > 0$  est définie par :*

$$f_{\lambda, \beta}^{Wbl}(x) = \frac{\beta}{\lambda} \left( \frac{x}{\lambda} \right)^{\beta-1} e^{-\left(\frac{x}{\lambda}\right)^\beta} \mathbf{1}_{\mathbb{R}^+}(x) \quad (3)$$

Ce choix de distribution est motivé par le fait que la distribution de Weibull donne la distribution exponentielle décroissante lorsque  $\beta = 1$ , et la distribution de Rayleigh lorsque  $\beta = 2$  et  $\lambda = \sqrt{2}\alpha^1$ . Ces deux distributions sont respectivement utilisées pour décrire l'intensité et l'amplitude des données SAR.

Il est intéressant de noter que lorsque  $\mu < 0$  dans (1), alors RGGau se réduit à une courbe exponentielle décroissante qui coïncide avec certaines forme de Weibull. Dans ce cadre joint où nous considérons également cette distribution, nous excluons ces fonctions redondantes en spécifiant  $\mu \geq 0$ . En ce sens (1) se réduit à :

$$\Lambda(\mu, \sigma, \beta) = \sigma \Gamma \left[ 1 + \frac{1}{\beta} \right] + \frac{\sigma}{\beta} \left( \Gamma \left[ \frac{1}{\beta} \right] - \Gamma \left[ \frac{1}{\beta}, \left( \frac{\mu}{\sigma} \right)^\beta \right] \right) \quad (4)$$

<sup>1</sup> $\alpha$  étant le paramètre de forme de la distribution de Rayleigh

Par la suite, on considère le paramètre de localisation entre deux distributions RGGau  $p$  et  $q$ ,  $\bar{\mu} = 0.5(\mu_p + \mu_q)$  comme le centroïde des paramètres de localisation  $\mu_{(\cdot)}$ . En effet, sans cette contrainte, les formes suivantes de divergences n'admettent pas de formulation analytique. Afin de garder l'information de localisation, on ajoutera la divergence sur les paramètres décrit dans la section 2.1 :  $\mathcal{D}_{loc} = \|\mu_p - \mu_q\|_2$ . Pour la divergence entre deux RGGau, on supposera un alignement de variable latente, comme si les variables RGGau étaient distribuées par rapport à un paramètre de localisation de centroïde  $\bar{\mu}$ .

Considérons deux variables aléatoires  $X_p$  et  $X_q$  ayant des fonctions de densité de probabilité  $f_{X_p}$  et  $f_{X_q}$  respectivement. Nous utiliserons la divergence de Kullback-Leibler (KLD) pour mesurer leur dissimilarité, notée  $\mathcal{K}$  dans la suite et définie par  $\mathcal{K}(X_p||X_q) = \mathcal{H}(X_p||X_q) - \mathcal{H}(X_p)$ , où  $\mathcal{H}(\cdot)$  est l'entropie d'une variable aléatoire  $Z$  admettant une densité de probabilité  $f_Z$ , et  $\mathcal{H}(X_p||X_q)$  est l'entropie croisée de  $X_p$  et  $X_q$ .

On considère dans ce travail, la version symétrique de la KLD (aussi appelée divergence de Jeffreys), définie par :

$$\mathcal{J}(X_p, X_q) = \mathcal{K}(X_p||X_q) + \mathcal{K}(X_q||X_p), \quad (5)$$

Dans la suite, nous considérons les variables aléatoires  $X_k^{RGGau}$  associées à la PDF  $f_{\mu_k, \sigma_k, \beta_k}^{RGGau}$ .

**Proposition 2.** *La divergence de Kullback-Leibler (KLD) entre deux variables aléatoires  $X_p$  et  $X_q$  suivant une loi RGGau est donnée par :*

$$\mathcal{K}(X_p^{RGGau}||X_q^{RGGau}) = \frac{\sigma_p}{\beta_p \cdot \Lambda(\bar{\mu}, \sigma_p, \beta_p)} \times (M - N) + \log \left[ \frac{\Lambda(\bar{\mu}, \sigma_q, \beta_q)}{\Lambda(\bar{\mu}, \sigma_p, \beta_p)} \right],$$

$$\text{avec } M = \left( \frac{\sigma_p}{\sigma_q} \right)^{\beta_q} \left[ 2\Gamma \left( \frac{1 + \beta_q}{\beta_p} \right) - \Gamma \left( \frac{1 + \beta_q}{\beta_p}, \left( \frac{\bar{\mu}}{\sigma_p} \right)^{\beta_p} \right) \right]$$

$$\text{et } N = \frac{2}{\beta_p} \Gamma \left( \frac{1}{\beta_p} \right) - \Gamma \left( 1 + \frac{1}{\beta_p}, \left( \frac{\bar{\mu}}{\sigma_p} \right)^{\beta_p} \right) \quad (6)$$

La divergence de Jeffreys  $\mathcal{J}(X_p^{RGGau}, X_q^{RGGau})$ , est obtenue utilisant l'équation (5) et (6).

La deuxième divergence symétrique considérée est celle portant sur les marginales suivant une distribution de Weibull donnée par (2). Elle est définie par ([6]) :

$$\begin{aligned} \mathcal{J}(X_p^{Wbl}, X_q^{Wbl}) &= \Gamma \left( 1 + \frac{\beta_q}{\beta_p} \right) \left( \frac{\alpha_p}{\alpha_q} \right)^{\beta_q} \\ &+ \Gamma \left( 1 + \frac{\beta_p}{\beta_q} \right) \left( \frac{\alpha_q}{\alpha_p} \right)^{\beta_p} \\ &+ (\beta_p - \beta_q) \cdot \log \left( \frac{\alpha_p}{\alpha_q} \right) + \varepsilon \cdot \left( \frac{\beta_p}{\beta_q} + \frac{\beta_q}{\beta_p} - 2 \right) - 2 \end{aligned}$$

avec  $\varepsilon = -\psi(1)$  est la constante d'Euler-Mascheroni.

### 2.3 Copules sur données bivariées

Nous proposons d'utiliser des modèles de copules pour modéliser les dépendances les données bivariées. Ces dernières

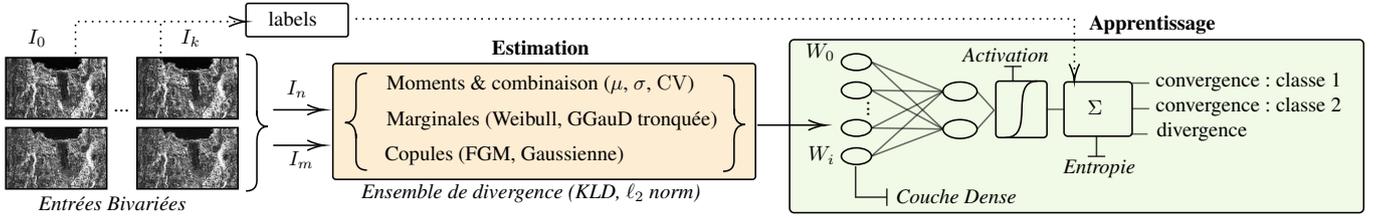


FIGURE 1 : Schéma de l'architecture de l'apprentissage d'ensembles de divergence.

permettent de séparer les contributions de chaque marginale et de la dépendance entre elles. Une copule  $C$  est une fonction de distribution cumulative conjointe (CDF) définie sur le cube unité  $d$ -dimensionnel  $[0; 1]^d$ , telle que chaque marginale soit uniforme sur  $[0; 1]$ . La restriction à des copules bivariées vient de l'utilisation dans ce travail, de données SAR à deux composantes polarimétriques, comme Sentinel-1 ou PAZ. Pour cette modélisation de dépendance, nous avons choisi de travailler avec des copules de type EFGM, pour sa forme simple et sa capacité à approximer au premier ordre des copules plus complexes comme les copules de type Frank et Plackett [9]. Pour l'estimation et par extension le calcul de divergence entre les copules, nous avons choisi d'utiliser la méthode non paramétrique basée sur les histogrammes [4] pour transformer les marginales en variables aléatoires uniformes sur  $[0, 1]$ . Soit deux variables aléatoires uniformes  $U$  et  $V$  sur  $[0, 1]$ , sous l'hypothèse d'une copule bivariée de type EFGM, on a la PDF du couple  $(U, V)$  suivante :

$$c(u, v) = 1 + \rho(1 - 2u)(1 - 2v) \mathbb{1}_{[0,1] \times [0,1]}(u, v)$$

On pose :

$$\mathcal{Q}_1(\rho_p, \rho_q) = |\rho_p - \rho_q| \left| \tanh^{-1}(\rho_p) - \tanh^{-1}(\rho_q) \right|, \quad (7)$$

$$\mathcal{Q}_2(\rho_p, \rho_q) = \left(1 - \frac{\rho_q}{\rho_p}\right) \frac{\tanh^{-1}(\rho_p)}{\rho_p} + \left(1 - \frac{\rho_p}{\rho_q}\right) \frac{\tanh^{-1}(\rho_q)}{\rho_q}, \quad (8)$$

$$\mathcal{Q}_3(\rho_p, \rho_q) = \left(1 - \frac{\rho_q}{\rho_p}\right) \frac{\text{Li}_2(\rho_p) - \text{Li}_2(-\rho_p)}{\rho_p} + \left(1 - \frac{\rho_p}{\rho_q}\right) \frac{\text{Li}_2(\rho_q) - \text{Li}_2(-\rho_q)}{\rho_q}, \quad (9)$$

avec  $\text{Li}_2$  la fonction polylogarithme de degré 2 (Dilogarithme). On propose ainsi la divergence suivante :

**Proposition 3** (Jeffreys divergence EFGM).

$$\mathcal{J}(\mathbf{X}_p^{EFGM}, \mathbf{X}_q^{EFGM}) = \frac{3}{2} \left( \frac{\rho_p}{\rho_q} + \frac{\rho_q}{\rho_p} - 2 \right) + \frac{1}{2} \sum_{k=1}^3 \mathcal{Q}_k(\rho_p, \rho_q), \quad (10)$$

Nous complétons cette divergence avec la distance  $l_2$  entre les distributions de FGM définie par :

**Proposition 4** (EFGM  $l_2$  distance).

$$l_2(\mathbf{X}_p^{EFGM}, \mathbf{X}_q^{EFGM}) = \frac{1}{9} (\rho_p - \rho_q)^2 \quad (11)$$

Afin de parfaire notre ensemble de divergence, nous proposons d'utiliser la divergence de Kullback-Leibler entre 2 copules Gaussiennes bivariée [7]. Ce choix est motivé par le fait que la copule Gaussienne est une copule de référence

simple à mettre en oeuvre, mais qui permet de couvrir un grand nombre de cas [1]. Cette divergence permet de séparer la contribution des marginales de celle de la copule, comme indiqué dans l'équation 14 de [7]. En considérant uniquement la contribution de la copule sous la forme de Jeffreys on obtient :

$$\mathcal{J}(\mathbf{X}_p^{\text{GauC}}, \mathbf{X}_q^{\text{GauC}}) = (1 - \rho_p \rho_q) \left( \frac{1}{1 - \rho_p^2} + \frac{1}{1 - \rho_q^2} \right), \quad (12)$$

avec  $\rho_p, \rho_q$  la corrélation entre les marginales de la copule respectivement pour les variables,  $\mathbf{X}_p^{\text{GauC}}$  et  $\mathbf{X}_q^{\text{GauC}}$ .

## 3 Simulations

### 3.1 Apprentissage de l'ensemble de divergence

On considère un cadre d'apprentissage supervisé pour l'agrégation des divergences estimées dans la partie 2, tel qu'illustré dans la Figure 1. À partir de l'ensemble de divergences, estimées sur les paires d'images d'entrée, nous cherchons ainsi à minimiser la fonction suivante

$$\mathcal{L} = H[\mathbf{y}, \mathcal{A} \circ f_\theta(\mathbf{d})]$$

avec  $H$  l'entropie croisée,  $\mathcal{A}$  une fonction d'activation softmax,  $\mathbf{d}$  le vecteur d'entrée des divergences et  $f_\theta$  une architecture neuronale caractérisé par un jeu de paramètre  $\theta$ .

Nous explorons 3 architectures neuronales  $f_\theta$  différentes :

- une couche dense ou fully connected ( $\mathbf{y} = \mathbf{X}\mathbf{w} + b$ ,  $\mathbf{w}$  et  $b$  étant les paramètres d'apprentissages),
- une couche dense avec activation suivi d'un minimum pooling et d'une couche dense,
- deux couches denses successives.

L'objectif est d'apprendre la combinaison optimale de l'ensemble de divergence dans le cadre d'une classification multi-classe. Le résultat de l'apprentissage permet de déterminer soit la classe à laquelle appartient le couple d'images en entrée, soit s'il existe une différence de classe entre les deux images. Nous contraignons la positivité des paramètres  $\theta$  afin de garder une explicabilité dans la décision finale de classification. L'opérateur de minimum pooling permet de réduire la dimensionnalité d'un vecteur en donnant la valeur minimale sur une fenêtre glissante de taille fixe. Le choix du minimum se justifie par l'utilisation de distances ou de leurs combinaisons.

### 3.2 Résultats

Les codes de ce travail sont donnés sur Github. Nous réalisons l'évaluation de cette approche sur un ensemble de données SAR en bande X du satellite PAZ, acquises entre juillet 2020 et août 2020 sur le massif du Mont-Blanc. Au total, 774 images

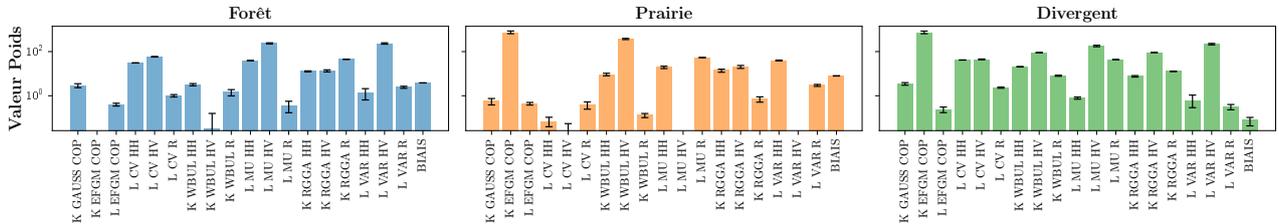


FIGURE 2 : Valeur des poids et du biais pour les 3 classes considérées en fonction des distance considérées. La première lettre représentant la nature de la distance (Kl ou  $\ell_2$ ), le deuxième bloc le modèle paramétrique et le troisième bloc le canal considéré (HH, HV ou R= HH/HV)

de dimension 32 par 32 pixels sur 2 canaux polarimétriques (HH et HV) ont été sélectionnées. Les canaux HH, HV sont en amplitude linéaire. Nous complétons les canaux par l'ajout de la valeur absolue du log-ratio HH/HV. Ces échantillons sont répartis en 2 classes : forêt et prairie, respectivement (390 et 384 échantillons). La labellisation des images a été réalisée en utilisant le modèle de couverture de sol CORINE fournit par Copernicus [5]. La construction de l'ensemble de divergences se fait en estimant 18 divergences ou distances, détaillées dans le tableau 1, pour chaque paire d'images du dataset. Le label associé à un couple d'images étant : *forêt*, *prairie* ou *divergent*. Les résultats des différentes architectures sont données dans le tableau 2 en réalisant une validation croisée par la méthode du K-Fold (avec K=5) permettant de parcourir l'ensemble du dataset avec la métrique F1-score. Nous remarquons que les 3 architectures donnent des résultats similaires ( $\sim 91.7\%$ ). L'architecture **b** reste la plus performante avec une précision de près de 5% supérieure à celle de l'architecture **a**, de par l'utilisation d'un nombre plus important de paramètres (195). Cependant, l'architecture **a** est intéressante par une variabilité plus faible que les autres architectures sur l'ensemble de folds et un faible nombre de paramètres (57). De plus, la simple combinaison des divergences permet de comprendre l'importance des divergences dans les résultats de classification, telle qu'illustrée dans la Figure 2. Nous remarquons que la classe *divergence* utilise l'ensemble des distance pour rendre sa décision. Les autres classes rejettent certaines distances, ainsi la classe *prairie* se concentre moins sur les distances liées au canal HV et se concentre sur la divergence entre copules EFGM, là où la classe *forêt* utilise majoritairement les statistiques descriptives.

## 4 Conclusion

Dans ce papier, nous avons montré la capacité d'un système composé d'un ensemble de divergences sur des modèles paramétriques, suivi de couches neuronales à apprendre une combinaison cohérente des divergences dans le cadre d'une classification supervisée multiclasse. Cette approche permet

SAR	stat. desc.	marginales	copules
HH	$\mu, \sigma, CV$	RGGau, Wbl	EFGM, Gaussienne
HV	$\mu, \sigma, CV$	RGGau, Wbl	EFGM
HH/HV	$\mu, \sigma, CV$	RGGau, Wbl	

TABLE 1 : Récapitulatif des modèles utilisés en fonction des données polarimétriques SAR, en bleu les distances basées sur les divergences de Jeffreys et en noir celle sur les normes  $\ell_2$

Architecture	a	b	c
F1-score (%)	$89.4 \pm 0.1$	$94.7 \pm 0.4$	$91.2 \pm 0.2$

TABLE 2 : Score d'apprentissage des différentes combinaisons à la fois d'éviter un sur-apprentissage sur les données, mais aussi de garder une interprétabilité du modèle. Les combinaisons plus complexes permettent de gagner en performance, au détriment d'un nombre de paramètres plus important. Les résultats obtenus sont préliminaires et visent à démontrer la capacité de l'approche proposée à classifier et à comprendre ce choix sur des données SAR en conditions réelles. Cette approche flexible pourra être mise en oeuvre dans le cadre d'apprentissage de natures différentes, non ou semi-supervisé, détection de changements, par sa capacité à rendre compte du résultat proposé.

## Références

- [1] M. Ishaq BHATTI et Hung Quang DO : Recent development in copula and its applications to the energy, forestry and environmental sciences. 44(36):19453–19473.
- [2] M.C. DOBSON, L. PIERCE, J. KELLNDORFER et F. ULABY : Use of sar image texture in terrain classification. In *IGARSS'97. 1997 IEEE International Geoscience and Remote Sensing Symposium Proceedings. Remote Sensing - A Scientific Vision for Sustainable Development*, volume 3, pages 1180–1183 vol.3, 1997.
- [3] Henri EYRAUD : Les principes de la mesure des corrélations. *Ann. Univ. Lyon, III. Ser., Sect. A*, 1(30-47):111, 1936.
- [4] Piotr JAWORSKI, Fabrizio DURANTE, Wolfgang Karl HÄRDLE et Tomasz RYCHLIK : Copula theory and its applications. 2010.
- [5] B. KOSZTRA, G. BÜTTNER, G. HAZEU et S. ARNOLD : Updated CLC illustrated nomenclature guidelines. titleTranslation : titleTranslation : titleTranslation : titleTranslation : .
- [6] Roland KWITT et Andreas UHL : Image similarity measurement by kullback-leibler divergences between complex wavelet subband statistics for texture retrieval. In *2008 15th IEEE International Conference on Image Processing*, pages 933–936, 2008.
- [7] Nour-Eddine LASMAR et Yannick BERTHOUMIEU : Gaussian copula multivariate modeling for texture image retrieval using wavelet transforms. 23(5):2246–2261. Number : 5.
- [8] G. MERCIER, G. MOSER et S.B. SERPICO : Conditional copulas for change detection in heterogeneous remote sensing images. 46(5):1428–1441.
- [9] Roger B. NELSEN : *An Introduction to Copulas*, volume 139 de *Lecture Notes in Statistics*. Springer.
- [10] Hemani PARIKH, Samir PATEL et Vibha PATEL : Classification of SAR and PolSAR images using deep learning : a review. 11(1):1–32. Number : 1.
- [11] Peter PLANINSIC, Jagmal SINGH et Dušan GLEICH : Sar image categorization using parametric and nonparametric approaches within a dual tree cwt. *IEEE Geoscience and Remote Sensing Letters*, 11(10):1757–1761, 2014.
- [12] Dong-Xiao YUE, Feng XU, Alejandro C. FRERY et Ya-Qiu JIN : Synthetic aperture radar image statistical modeling : Part one-single-pixel statistical models. 9(1):82–114. Number : 1.