

Une approche pour la tomographie SAR des zones forestières par apprentissage profond supervisé

Zoé BERENGER¹ Loïc DENIS^{2,1} Florence TUPIN¹ Laurent FERRO-FAMIL³

¹LTCI, Télécom Paris, Institut polytechnique de Paris, Palaiseau, France

²Laboratoire Hubert Curien, UMR 5516, CNRS, Institut d'Optique Graduate School, Univ. Lyon, UJM-Saint-Étienne, Saint-Étienne, France

³ISAE-SUPAERO and CESBIO, University of Toulouse, Toulouse, France

Résumé – L'imagerie tomographique par radar à synthèse d'ouverture (SAR) reconstruit les informations tridimensionnelles de réflectivité d'une scène à partir d'un ensemble d'acquisitions cohérentes réalisées dans une configuration interférométrique. Dans les zones forestières, cette réflectivité est modélisée sous la forme d'un profil vertical pour des coordonnées distance azimut arbitraires. Pour reconstruire ce profil, des techniques d'estimation spectrale non paramétrique à faible résolution ou une inversion régularisée mise en œuvre sous la forme d'algorithmes de minimisation itératifs, qui sont très coûteux en temps, sont utilisées. Nous présentons ici une approche d'apprentissage profond supervisée à partir de données simulées proposée dans [3]. Dans cet article nous étudions la variabilité inter-entraînement sur données simulées et réelles du réseau de neurones développé.

Abstract – Synthetic Aperture Radar (SAR) tomographic imaging reconstructs three-dimensional reflectivity information of a scene from a set of coherent acquisitions made in an interferometric configuration. In forested areas, this reflectivity is modeled as a vertical profile for arbitrary range azimuth coordinates. To reconstruct this profile, low resolution non parametric spectral estimation techniques or regularized inversion implemented as iterative minimization algorithms, which are very time consuming, are used. We present here a supervised deep learning approach based on simulated data proposed in [3]. In this paper, we study the inter-training variability on simulated and real data of the developed neural network.

1 Introduction

La tomographie par radar à synthèse d'ouverture (SAR) est une technique qui permet de reconstruire des images 3D d'environnements volumiques à partir d'un ensemble d'images SAR 2D cohérentes. La réflectivité d'environnements naturels comme les forêts étant affectée par le phénomène de speckle, les procédures d'imagerie tomographique estiment des intensités à variance réduite à travers des moments incohérents d'ordre deux. L'objectif de l'étape de reconstruction tomographique est de reconstruire en chaque pixel des images un profil de réflectivité dans la direction verticale. Ce profil peut être modélisé à l'aide de fonctions continues [1].

Dans la littérature, des techniques d'estimation spectrale non paramétriques, telles que la formation de faisceaux (*beam-forming*), correspondant à la transformée de Fourier discrète dans le cas de signaux à bande étroite, et le filtre de *Capon* [7] sont appliquées pour effectuer de la tomographie SAR sur des forêts. Ces estimateurs sont rapides et présentent des expressions simples, mais sont cependant limités par leur résolution verticale grossière. D'autres méthodes de pointe, telles que l'inversion régularisée d'un modèle linéaire de la matrice de covariance (par exemple des algorithmes inspirés de *Compressive Sensing* (CS) utilisant un petit nombre de fonctions de base paramétriques [2, 5]), permettent d'obtenir des reconstructions beaucoup plus précises. Elles nécessitent néanmoins des algorithmes de minimisation itératifs qui sont coûteux en temps et qui limitent leur application à des ensembles de données à grande échelle.

Les approches d'apprentissage profond, encore rares dans

ce domaine, utilisent surtout des entraînements supervisés à partir de données LiDAR qui servent à estimer la hauteur du sommet des arbres et celle du sol [?]. Faute d'autres données pouvant servir de vérité terrain, toute autre méthode supervisée doit utiliser des profils simulés.

L'objectif de cet article est de proposer une approche supervisée d'apprentissage profond pour la reconstruction tomographique SAR dans les zones forestières. Des données sont simulées à partir d'un modèle génératif validé dans des travaux antérieurs afin de remédier à ce manque de vérité terrain.

2 Contexte

Le modèle d'acquisition linéaire du vecteur d'amplitudes complexes $\mathbf{y} \in \mathbb{C}^N$ collecté en un pixel 2D fixé par chacune des N images SAR peut être exprimé comme suit :

$$\mathbf{y} = \sum_{k=1}^{N_s} s_k \mathbf{a}(z_k) + \epsilon = \mathbf{A}\mathbf{s} + \epsilon \quad (1)$$

où $\mathbf{s} = [s_1, \dots, s_{N_s}]^T$ contient le coefficient de réflexion complexe des N_s diffuseurs observés, $\mathbf{a}(z) = [1, \dots, e^{jk_z^{(N)}z}]^T$ est un vecteur de mesure, ou vecteur directionnel (steering vector), modélisant la différence de phase interférométrique entre la première et la n -ième image pour un diffuseur situé à la hauteur z . $\mathbf{A} = [\mathbf{a}(z_1), \dots, \mathbf{a}(z_{N_s})]$ est la matrice de mesure, ou matrice directionnelle (steering matrix), tandis que ϵ représente le bruit de mesure, qui suit une distribution gaussienne complexe circulaire centrée de covariance $\sigma_\epsilon^2 \mathbf{I}_N$.

Dans les zones forestières, la densité verticale de la réflectivité s peut être représentée par un grand nombre $N_s \gg N$ de réflectivités de sources indépendantes, en pratique réparties sur un vecteur $\mathbf{z} \in \mathbb{R}^{N_z}$ de N_z hauteurs discrètes. La matrice de covariance de s est diagonale et est donnée par $\Sigma_{ss} = \mathbb{E}[ss^H] = \text{diag}(\mathbf{p})$, avec \cdot^H l'opérateur de transposition hermitienne [7]. La matrice de covariance du signal mesuré peut alors être exprimée comme suit :

$$\Sigma = \mathbb{E}[\mathbf{y}\mathbf{y}^H] = \mathbf{A}\text{diag}(\mathbf{p})\mathbf{A}^H + \sigma_c^2\mathbf{I}_N \quad (2)$$

Cette quantité est généralement estimée à partir de L réalisations indépendantes :

$$\widehat{\Sigma} = \frac{1}{L} \sum_{l=1}^L \mathbf{y}_l \mathbf{y}_l^H \quad (3)$$

L'objectif de l'imagerie tomographique SAR en forêt est l'estimation de $\mathbf{p} \in \mathbb{R}_+^{N_z}$ à partir de $\widehat{\Sigma}$.

2.1 Beamforming

Les estimateurs spectraux non paramétriques sont généralement utilisés comme référence pour la reconstruction tomographique des forêts, car ils sont bien adaptés à l'estimation de profils de réflectivités continues. Le plus classique est obtenu avec la *formation de faisceaux (beamforming)* qui équivaut à une synthèse d'ouverture multi-vues dans la direction de l'élévation [7] :

$$p_i^{(B)} = \frac{\mathbf{a}^H(z_i)\widehat{\Sigma}\mathbf{a}(z_i)}{N^2} \quad (4)$$

où \mathbf{a} est un vecteur colonne directionnel de la matrice \mathbf{A} . Dans l'approche proposée, ces profils *beamforming* ont été utilisés comme données d'entrée du réseau de neurones.

Le *beamforming* présente l'avantage d'être rapide, mais rencontre des difficultés lorsqu'il est confronté à des diffuseurs étroitement espacés, créant des lobes secondaires et des valeurs aberrantes.

Le filtre de *Capon* [7] est une version adaptative du *beamforming*, ayant pour objectif de minimiser la puissance en sortie tout en assurant un gain unitaire à la hauteur d'analyse. Ce filtre linéaire permet d'atténuer les lobes secondaires ainsi que les interférences entre différentes composantes du profil, au prix de la précision radiométrique. Ce filtre est aussi fortement affecté par le bruit thermique.

3 Approche proposée

L'approche proposée consiste à simuler des profils simples pour entraîner un réseau de neurones à reconstruire ce profil par apprentissage supervisé. Un modèle génératif est utilisé pour créer des profils de vérité terrain. Des réalisations de mesures SAR \mathbf{y} peuvent ensuite être simulées à l'aide de la matrice de mesure \mathbf{A} et de la matrice de covariance Σ définie en (2), pour une configuration géométrique préalablement définie. Pour faciliter l'apprentissage du réseau de neurones présenté dans cet article, des profils calculés avec *beamforming* à partir d'une matrice de corrélation multi-look $\widehat{\mathbf{R}} = \text{diag}(\mathbf{q})\widehat{\Sigma}\text{diag}(\mathbf{q})$ (où $q_i = 1/\sqrt{\widehat{\Sigma}_{ii}}$) pour L mesures \mathbf{y}_1

à \mathbf{y}_L , à la place de $\widehat{\Sigma}$ dans l'équation (4), sont utilisés comme données en entrée du réseau. Le passage par l'opération de *beamforming* permet de manipuler des profils verticaux d'intensité au lieu de matrices de covariance à valeurs complexes et de maintenir l'entrée et la sortie du réseau dans un espace réel. Cette approche permet aussi d'inclure la physique d'acquisition via la matrice de mesure et garantit une unité de représentation en entrée et sortie du réseau.

La matrice de corrélation, $\widehat{\mathbf{R}}$ est utilisée à la place de la matrice de covariance $\widehat{\Sigma}$ pour rendre la méthode insensible aux changements d'échelle liés à la valeur absolue de l'intensité. La reconstruction finale est donc $\widehat{\mathbf{p}} = \text{Tr}(\widehat{\Sigma}/N)\widehat{\mathbf{p}}$, avec $\widehat{\mathbf{p}}$ la sortie du réseau, pour récupérer la valeur d'intensité réelle du profil.

3.1 Modèle génératif

La distribution de la réflectivité dans une cellule de zone forestière est principalement composée de deux pics, un pour le sol et un pour la couronne des arbres, comme validé par les dernières méthodes de reconstruction tomographique [2, 5]. Ces pics peuvent être représentés par de nombreuses fonctions de base, par exemple deux gaussiennes. Les paramètres de ces gaussiennes sont adaptés aux types de forêts observées dans nos ensembles de données SAR pour chaque entraînement, c'est-à-dire des forêts boréales mesurées en bande L ($\lambda \approx 30$ cm) et des forêts tropicales observées en bande P ($\lambda \approx 70$ cm). Les plages de ces paramètres sont choisies pour être larges, afin de couvrir tous les profils possibles rencontrés dans la pratique et d'être représentatives de l'ensemble de la forêt.

En pratique, 10000 profils ont été générés, échantillonnés sur $N_z = 512$ hauteurs uniformément espacées, entre -10 m et 30 m en bande L, et -20 m et 60 m en bande P.

3.2 Simulation de mesures

Pour chaque profil généré avec le modèle décrit, une matrice de mesure \mathbf{A} choisie au hasard parmi les matrices de mesure d'un ensemble de données tomographiques réelles a été associée. Les profils ont été normalisés par $\alpha = \frac{1}{N} \sum_i \frac{1}{\sum_k |A_{ik}|^2 p_k}$ pour que la matrice $\Sigma = \mathbf{A}\text{diag}(\mathbf{p})\mathbf{A}^H$ corresponde directement à une matrice de corrélation. Des échantillons gaussiens complexes circulaires $\{\mathbf{y}_l\}_{l=1..L} \sim \mathcal{N}(0, \mathbf{A}\text{diag}(\mathbf{p})\mathbf{A}^H)$ ont ensuite été obtenus comme $\mathbf{y}_l = \mathbf{A}\text{diag}(\sqrt{\mathbf{p}}\mathbf{w}_l)$, avec \mathbf{w}_l un bruit blanc complexe centré sur 0 de variance unitaire. Les matrices de corrélation des échantillons $\widehat{\mathbf{R}}$ ont été calculées à partir de chaque ensemble de L échantillons $\{\mathbf{y}_l\}$, avec ici $L = 60$ vues.

3.3 Architecture

Une structure d'encodeur-décodeur de petite taille, validée empiriquement, est choisie pour l'architecture du réseau. L'encodeur entraîné consiste en 4 couches avec un nombre décroissant de neurones, un décodeur symétrique de 4 couches et un espace latent de taille 5. Cette dernière valeur est sélectionnée de manière à correspondre au nombre de paramètres utilisés dans la génération des profils de réflectivité simulés. Chaque couche est non biaisée et dispose d'une fonction d'activation leaky ReLU. L'apprentissage a été effectué sur 75% des données composées des profils *beamforming* simulés, les 25%

restants servant d'ensemble de validation. Des mini-batches de taille 32 avec un optimiseur Adam et un taux d'apprentissage de 10^{-3} sur 100 époques sont également fixés pour l'entraînement. La fonction de perte utilisée est une erreur quadratique moyenne entre le profil vérité terrain simulé et le profil prédit en sortie du réseau, ce qui est un choix standard pour les problèmes de régression. D'autres types de fonction de perte ont également été évalués, tels que la similarité cosinus, qui favorise les formes semblables plutôt que les valeurs proches, ou une fonction de transport optimal, produisant tous des résultats similaires.

4 Expériences

4.1 Données en bande L

Les premières données utilisées pour valider cette approche proviennent d'une pile tomographique de 6 images SAR aéroportées en bande L acquises lors de la campagne BioSAR-2 menée par le DLR en 2008 [6]. Ces images couvrent une forêt boréale du nord de la Suède, avec une résolution verticale variant de 6 m en distance (en début de fauchée) à 25 m (en fin de fauchée). Une fenêtre locale d'environ 60 vues est utilisée pour calculer la matrice de corrélation utilisée par l'algorithme *beamforming* pour calculer les tomogrammes. Le réseau étant entraîné sur des données simulées pour une plage fixe de hauteurs, il est nécessaire de compenser un modèle numérique d'élévation (DEM) grossier afin que les profils reconstruits présentent un pic correspondant au sol à une hauteur similaire.

La figure 1 montre un exemple de tomogramme en polarisation HH reconstruit avec la méthode proposée, dont la comparaison avec les résultats obtenus avec *beamforming* et *Capon*, indique que le réseau entraîné améliore effectivement la résolution du tomogramme par rapport aux méthodes d'estimation spectrale simple, tout en conservant un volume représentatif de la couronne de l'arbre, ce qui est attendu sur les données en bande L.

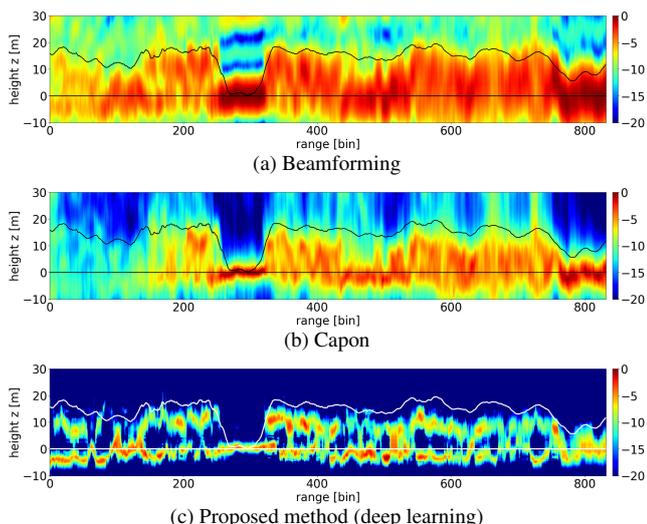


FIGURE 1 : Tomogramme pour une valeur d'azimut spécifique dans une forêt boréale en bande L, calculé avec la méthode : (a) Beamforming ; (b) Capon ; (c) Réseau de neurones proposé. Les estimations LiDAR des hauteurs (compensées) du sol et de la cime des arbres sont représentées en lignes continues.

4.2 Incertitudes

Afin de valider la stabilité de la méthode proposée, 10 entraînements à partir d'initialisations aléatoires du réseau ont été réalisés puis testés sur des données simulées et des données réelles. Les figures 2 et 3 montrent les profils moyens ainsi que la variance des résultats obtenus, et confirmer la fiabilité de notre approche de reconstruction tomographique.

Il est possible d'améliorer les résultats obtenus sur les données simulées en augmentant le nombre d'époques d'apprentissage du réseau, et d'obtenir ainsi des profils moyens reconstruits plus proches des profils simulés avec un écart interquartile réduit. Cependant, cela dégrade l'application de ce réseau à des données réelles, qui se spécialise trop au modèle génératif très simple proposé. Une telle amélioration serait possible avec un modèle parfaitement adapté à la forêt à reconstruire. De tels modèles n'étant pas disponibles, ceci n'est pas envisageable actuellement.

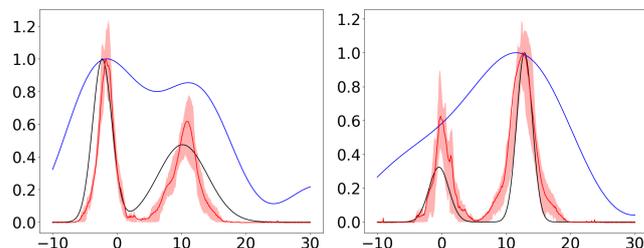


FIGURE 2 : Profils simulés (noir), profils beamforming calculés à partir de mesures simulées sur ces premiers profils (bleu) et reconstructions des profils par le réseau de neurones à partir des profils beamforming (rouge). Pour ce dernier, le profil moyen ainsi que l'écart interquartile sur 10 entraînements différents est affiché. Tous les profils ont été normalisés pour pouvoir être affichés sur la même échelle.

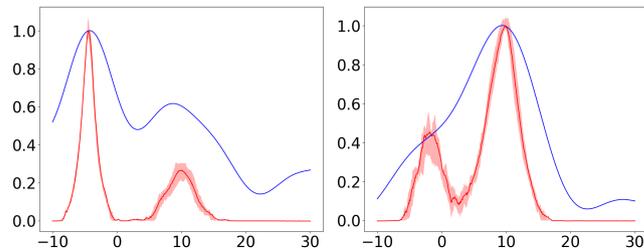


FIGURE 3 : Profils beamforming calculés sur les données réelles décrites dans la partie 4.1 (bleu) et reconstructions des profils par le réseau de neurones (rouge). Pour ce dernier, le profil moyen ainsi que l'écart interquartile sur 10 entraînements différents est affiché. Tous les profils ont été normalisés pour pouvoir être affichés sur la même échelle.

4.3 Données en bande P

Le deuxième jeu de données testées comprend des mesures SAR tomographiques aéroportées acquises en bande P par l'ONERA au-dessus du site d'essai de Paracou en Guyane française, lors de la campagne TropiSAR en 2009 [4]. Ces données comprennent également 6 images, avec une résolution verticale de l'ordre de 15 m. Le DEM a de nouveau été compensé avant de calculer la matrice de corrélation utilisée

pour comparer les différentes méthodes avec 56 vues, de calculer la reconstruction *beamforming* puis de la fournir en entrée au réseau. Un échantillon des résultats et des comparaisons est présenté dans la figure 4, indiquant que l’approche proposée convient particulièrement à ce type de données en produisant des profils reconstruits à haute résolution.

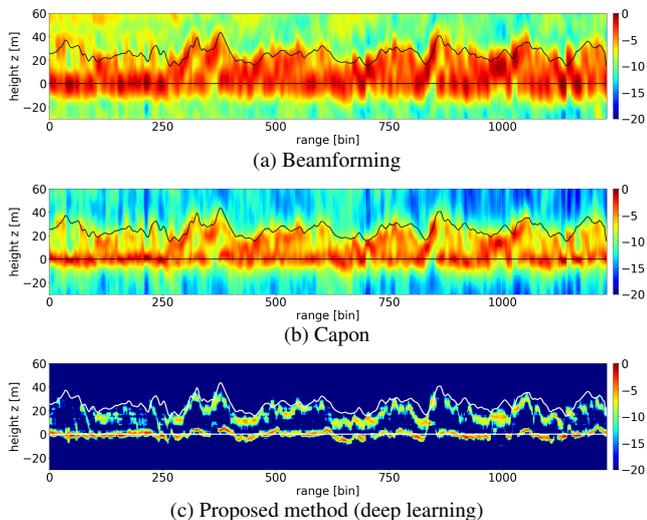


FIGURE 4 : Tomogramme pour une valeur d’azimut spécifique dans une forêt boréale en bande P, réalisé avec la méthode : (a) Beamforming ; (b) Capon ; (c) Réseau de neurones proposé. Les estimations LiDAR des hauteurs (compensées) du sol et de la cime des arbres sont représentées en lignes continues.

5 Conclusion

L’approche proposée utilise un réseau de neurones de faible dimension pour effectuer une inversion tomographique SAR sur les forêts, avec des résultats validés sur des données réelles en bande L et en bande P. Ces derniers révèlent des performances prometteuses, tant en termes de qualité que de temps de calcul nécessaire à la reconstruction d’une grande image, avec une accélération de plusieurs ordres de grandeur par rapport aux algorithmes d’inversion régularisée itératifs. Bien qu’obtenus avec un modèle génératif simple, ils confirment le fort potentiel de l’application de l’apprentissage profond dans ce domaine, avec des améliorations possibles notamment sur les données d’entrée ou sur le choix de la vérité terrain utilisée.

Remerciements : Ces travaux ont été financés par la fondation Futur et Ruptures de l’IMT (Institut Mines Télécom) et partiellement financés par l’ANR et la DGA dans le cadre du projet ASTRAL ANR-21-ASTR-0011.

Références

[1] Hossein AGHABABAEI, Giampaolo FERRAIOLI, Laurent FERRO-FAMIL, Yue HUANG, Mauro MARIOTTI D’ALESSANDRO, Vito PASCAZIO, Gilda SCHIRINZI et Stefano TEBALDINI : Forest SAR tomography : Principles and applications. *IEEE Geoscience and Remote Sensing Magazine*, 8(2):30–45, février 2020.

[2] Esteban AGUILERA, Matteo NANNINI et Andreas REIGBER : Wavelet-based compressed sensing for SAR tomography of forested areas. *IEEE Transactions on Geoscience and Remote Sensing*, 51(12):5283–5295, 2013.

[3] Zoé BERENGER, Loïc DENIS, Florence TUPIN, Laurent FERRO-FAMIL et Yue HUANG : A Deep Learning Approach for SAR Tomographic Imaging of Forested Areas. *In arXiv.2301.08605*, 2023.

[4] Pascale DUBOIS-FERNANDEZ, Thuy Le TOAN, Sandrine DANIEL, Hélène M. ORIOT, Jérôme CHAVE, Lilian BLANC, Ludovic VILLARD, Malcolm DAVIDSON et Michel PETIT : The TropiSAR airborne campaign in French Guiana : objectives, description, and observed temporal behavior of the backscatter signal. *IEEE Transactions on Geoscience and Remote Sensing*, 50:3228–3241, 2012.

[5] Laurent FERRO-FAMIL, Yue HUANG et N. GE : Estimation of the vertical structure of a tropical forest using basis functions and parametric SAR tomography. *In IGARSS*, 2022.

[6] Irena HAJNSEK, Rolf SCHEIBER, Martin KELLER, Ralf HORN, Seungkuk LEE, Lars ULANDER, Anders GUSTAVSSON, Gustaf SANDBERG, Thuy Le TOAN, Stefano TEBALDINI, Andrea MONTI-GUARNIERI et Fabio ROCCA : *BIOSAR 2008 technical assistance for the development of airborne SAR and geophysical measurements during the BioSAR 2008 experiment, final report*, 2009.

[7] Petre STOICA et Randolph L. MOSES : *Spectral analysis of signals*. Prentice Hall, Upper Saddle River, NJ, 2005.