A Deep Learning Approach for SAR Tomographic Imaging of Forested Areas

Zoé Berenger, Loïc Denis, Florence Tupin, Laurent Ferro-Famil, Yue Huang

A Deep Learning Approach for SAR Tomographic Imaging of Forested Areas

Zoé Berenger, Loïc Denis, *Senior Member, IEEE,* Florence Tupin, *Senior Member, IEEE,* Laurent Ferro-Famil, *Member, IEEE,* and Yue Huang

***Abstract*—Synthetic aperture radar tomographic imaging reconstructs the three-dimensional reflectivity of a scene from a set of coherent acquisitions performed in an interferometric configuration. In forest areas, a high number of elements backscatter the radar signal within each resolution cell. To reconstruct the vertical reflectivity profile, state-of-the-art techniques perform a regularized inversion implemented in the form of iterative minimization algorithms. We show that light-weight neural networks can be trained to perform this inversion with a single feed-forward pass, leading to fast reconstructions that could better scale to the amount of data provided by the future BIOMASS mission. We train our encoder-decoder network using simulated data and validate our technique on real L-band and P-band data.**

***Index Terms*—SAR tomography (TomoSAR), deep learning, forests, inverse problems**

I. Introduction

SYNTHETIC aperture radar (SAR) tomography (TomoSAR) uses a 2D aperture to perform 3D imaging. In the case of a narrow-band radar waveform and under the widely adopted Born approximation at order 1 [1], the imaging process simplifies to the 1D spectral analysis of a set of co-registered 2D SAR images [2]. It aims at reconstructing, for each 2D location, reflectivity profiles in the direction orthogonal to the radar line-of-sight. As illustrated in [3], parametric spectral estimation approaches, which have been widely used for the characterization of urban areas, such as high-resolution techniques [3] or *Compressive Sensing (CS)*-inspired regularized least squares minimization [4], [5], fail to adequately reconstruct the response of forested environments. They indeed estimate a small set of discrete point-like scattering sources, instead of a continuous function, known to represent well the reflectivity of such volumetric media [6].

Among the wide range of existing non-parametric spectral estimation techniques [7], the *beamformer*, i.e. the discrete Fourier transform, and *Capon*'s filter, also called the adaptive *beamformer*, are the most widely used to perform TomoSAR focusing over forests. The *beamformer* has a coarse resolution and creates sidelobes, whereas *Capon*'s sidelobe reduction capability comes at the cost of radiometric accuracy [7]. A parametric solution based on the use of a sparsifying basis able to approximate a continuous function using a small set of coefficients was proposed in [8]. This approach, named *wavelet-based CS*, used a *CS*-inspired optimization to determine a reflectivity profile constructed using an orthogonal wavelet matrix and a regularized number of wavelet coefficients. Another approach, using a small number of parametric basis functions, was proposed in [9]. The regularized inversion of a linear model of the covariance matrix leads to much more accurate reconstructions, nevertheless the latter estimators, which are themselves non-linear, require costly iterative minimization algorithms that impede their application to large-scale datasets.

In many imaging domains, deep learning has enabled computation time to be reduced while retaining high-quality results. This potential has been applied to SAR tomography in [10], [11], where Iterative Shrinkage Thresholding Algorithms (ISTA) have been unrolled to solve the $L_2 - L_1$ norm minimization problem posed by *CS* in urban areas.Yet most deep learning techniques for SAR tomography over forests focus on height estimation using LiDAR data as a reference [12].

This paper presents a supervised deep learning method for tomographic SAR reconstruction in forested areas. Its objective is to recover the reflectivity profile from the coarse profile obtained by the *beamformer*. We first use a physics-inspired generation model and interferometric baselines matching our SAR dataset to simulate reflectivity distributions and associated measurements. We then train a network with a light-weight architecture to learn a low-dimensional latent representation of these simulated profiles and to recover the original profiles free from *beamforming* artifacts. Our network does not perform an explicit inversion of the profiles' parametric model, but rather reconstructs deconvolved tomographic profiles. As our training set is generated from physically-plausible profiles, the network implicitly includes knowledge of our profile models, similarly to prior knowledge in Bayesian estimators. Finally, the neural network is evaluated on real *beamforming* profiles at L-band and P-band and compared to several methods, showing promising performances both in terms of reconstruction quality and computation time.

II. Methodology

A. Problem formulation

The tomographic signal measured over N SAR acquisitions, $\mathbf{y} \in \mathbb{C}^N$, may be formulated as the sum of the N_s contributions

This project has been funded by the Futur & Ruptures PhD program of the Fondation Mines-Telecom, and partially funded by ASTRAL project (ANR-21-ASTR-0011).

Z. Berenger and F. Tupin are with LTCI, Télécom Paris, Institut polytechnique de Paris, Paris, France (e-mail: name.surname@telecom-paris.fr).

L. Denis is with the Laboratoire Hubert Curien, UMR 5516, CNRS, Institut d'Optique Graduate School, Univ. Lyon, UJM-Saint-Étienne, Saint-Étienne 42023, France (e-mail: loic.denis@univ-st-etienne.fr).

L. Ferro-Famil is with ISAE-SUPAERO and CESBIO, University of Toulouse, Toulouse, France (e-mail: laurent.ferro-famil@isae-supaero.fr).

Y. Huang is with CESBIO, University of Toulouse, Toulouse, France (e-mail: yhuang228@gmail.com).

originating from the considered 2D resolution cells as:

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

where $\mathbf{s} = [s_1, \ldots, s_{N_s}]^T$ contains the complex reflection coefficient of the observed scatterers, $\mathbf{a}(z) = [1, \ldots, \mathrm{e}^{jk_z^{(N)}z}]^T$ is a steering vector and models the interferometric phase with $k_z^{(n)}z$ corresponding to the phase seen on the n-th image for a scatterer located at height z, $\mathbf{A} = [\mathbf{a}(z_1), \ldots, \mathbf{a}(z_{N_s})]$ is the sensing matrix, while $\boldsymbol{\epsilon}$ stands for the system noise, distributed according to a circular complex Gaussian distribution of covariance $\sigma_\epsilon^2 \mathbf{I}_N$.

As mentioned earlier, over forested areas, the vertical density of reflectivity $\mathbf{s}$ is well modelled by a speckle-affected continuous function, and hence may be represented in (1) by a large number $N_s \gg N$ of uncorrelated source reflectivities distributed over a vector $\mathbf{z} \in \mathbb{R}^{N_z}$ of N_z discrete heights. The covariance matrix of $\mathbf{s}$ is diagonal and is given by $\boldsymbol{\Sigma}_\mathbf{s} = \mathbb{E}[\mathbf{s}\mathbf{s}^H] = \mathrm{diag}(\mathbf{p})$, with $\cdot^H$ the Hermitian transpose operator. The measured signal covariance matrix then becomes:

$$\boldsymbol{\Sigma} = \mathbb{E}[\mathbf{y}\mathbf{y}^H] = \mathbf{A}\mathrm{diag}(\mathbf{p})\mathbf{A}^H + \sigma_\epsilon^2 \mathbf{I}_N \tag{2}$$

In practice, this quantity is estimated using L independent realizations of the measured vector, $\{\mathbf{y}_l\}_{l=1}^L$, sampled in the neighborhood of a 2D location. The corresponding sample covariance matrix is given by $\widehat{\boldsymbol{\Sigma}} = \frac{1}{L}\sum_{l=1}^L \mathbf{y}_l \mathbf{y}_l^H$. The objective of forest TomoSAR imaging is the estimation, or at least the characterization, of $\mathbf{p} \in \mathbb{R}_+^{N_z}$ from $\widehat{\boldsymbol{\Sigma}}$. The sample correlation matrix, $\widehat{\mathbf{R}} = \mathrm{diag}(\mathbf{q})\widehat{\boldsymbol{\Sigma}}\mathrm{diag}(\mathbf{q})$, with $q_i = 1/\sqrt{\widehat{\boldsymbol{\Sigma}}_{ii}}$, represents a version of $\widehat{\boldsymbol{\Sigma}}$, in which the intensity of each image is scaled to 1. Such a representation, independent of the observed absolute reflectivity, may also be used during specific steps of the 3D imaging process.

B. From fixed dictionaries to learned representations: strategies for vertical profile reconstruction

As shown in [3], [8], forest reflectivity profiles $\mathbf{p}$ can typically be approximated as a linear combination of a few basis functions, $\mathbf{p} \approx \boldsymbol{\Psi}\boldsymbol{\alpha}$, where $\boldsymbol{\Psi} \in \mathbb{R}^{N_z \times N_\alpha}$ is the dictionary whose columns are the basis functions and $\boldsymbol{\alpha} \in \mathbb{R}^{N_\alpha}$ is the vector of weights. For an adequately chosen basis, only a few functions at a time are necessary to approximate a given profile. The vector $\boldsymbol{\alpha}$ is then sparse: most coefficients in $\boldsymbol{\alpha}$ are zero and the number of non-zeros $\|\boldsymbol{\alpha}\|_0$ is small, as is the L1 norm $\|\boldsymbol{\alpha}\|_1$ which is often used as a proxy to measure sparsity. The reflectivity profile $\widehat{\mathbf{p}}^{(\mathrm{CS})} = \boldsymbol{\Psi}\widehat{\boldsymbol{\alpha}}$ can be estimated using the following minimization problem [8]:

$$\widehat{\mathbf{p}}^{(\mathrm{CS})} = \underset{\mathbf{p}\geq\mathbf{0}}{\arg\min} \; \|\mathbf{A}\mathrm{diag}(\mathbf{p})\mathbf{A}^H - \widehat{\boldsymbol{\Sigma}}\|_\mathrm{F}^2 + \lambda\|\boldsymbol{\Psi}^\dagger\mathbf{p}\|_1 \tag{3}$$

where λ is a hyper-parameter responsible for balancing the weight of the sparsity constraint, with respect to the data-fidelity term defined by the Frobenius norm. The pseudo inverse matrix, $\boldsymbol{\Psi}^\dagger$, in the L1 norm term verifies $\boldsymbol{\Psi}\boldsymbol{\Psi}^\dagger\mathbf{p} = \boldsymbol{\alpha}$, and may be replaced with $\boldsymbol{\Psi}^H$ in the case of a unitary orthogonal basis. Solving this problem typically requires several tens or hundreds of iterations of a minimization algorithm and the proper tuning of the hyper-parameter λ.

Rather than using a linear model for the profiles $\mathbf{p}$, non-linear models such as the generative model discussed in paragraph II-D can be used (e.g., Gaussian mixture models, or exponential profiles [9]). However, estimating the parameters of such models can be difficult due to identifiability issues and require additional constraints such as order constraints as recently proposed in a different modality of remote sensing for the estimation of phenological parameters from NDVI time series [13]. Our early experiments have shown that it was preferable to jointly learn a non-linear model for $\mathbf{p} = m(\boldsymbol{\alpha})$ and an estimator $\widehat{\boldsymbol{\alpha}}$ by training a deep neural network with an encoder-decoder architecture. Since the encoder and decoder are trained end-to-end, and no constraint is imposed on the latent space, the representation $\boldsymbol{\alpha}$ learned by the encoder and then transformed into a profile by the decoder is more flexible. The profiles estimated by the network can thus be considered as deconvolved versions of the input profiles, rather than the result of the inversion of a profile parametric model. The use of a simple feed-forward pass of a light-weight neural architecture leads to an extremely efficient tomographic reconstruction method, as described in the following paragraphs.

C. Proposed method

The different steps of our method are presented in Fig. 1. Starting from a ground truth profile of our generative model, L SAR measurement vectors $\mathbf{y}_1 \in \mathbb{C}^N$ to $\mathbf{y}_L \in \mathbb{C}^N$ are simulated using the steering matrix $\mathbf{A}$ of a specific geometric configuration with N antennas. To ease the task of the network, we provide it with the *beamforming* profile associated to the L-look correlation matrix $\widehat{\mathbf{R}}$ computed from the L measurement vectors $\mathbf{y}_1$ to $\mathbf{y}_L$. In this way, the network input and output are in the same space, i.e. both positive-valued vectors, easier to manipulate than complex-valued vectors. Besides, the *beamforming* profile depends less on the interferometric baselines than the correlation matrix $\widehat{\mathbf{R}}$. It includes the physics of acquisition through the steering matrix.

The correlation matrix has been used rather than the covariance matrix $\widehat{\boldsymbol{\Sigma}}$ to make the method invariant to differing intensity values. The final reconstruction will therefore be $\widehat{\mathbf{p}} = \mathrm{Tr}(\widehat{\boldsymbol{\Sigma}}/N)\,\tilde{\mathbf{p}}$, with $\tilde{\mathbf{p}}$ the output of the network, to recover the actual intensity value of the profile.

We have chosen to use a simple network architecture with an encoder-decoder creating a compressed representation of each profile in a low-dimensional latent space, less than 10, described in section III. This dimension is based on the number of parameters used in the profile generation model. A quadratic loss is applied between the ground-truth simulated profile and the output profile, a standard choice for regression problems.

D. Generative model

Given the poor vertical resolution of the reflectivity profiles that can be achieved with the tomographic stacks considered in this paper, only two peaks are considered: one for the ground and the other for the canopy. This is consistent with the observations made with the latest tomographic reconstruction

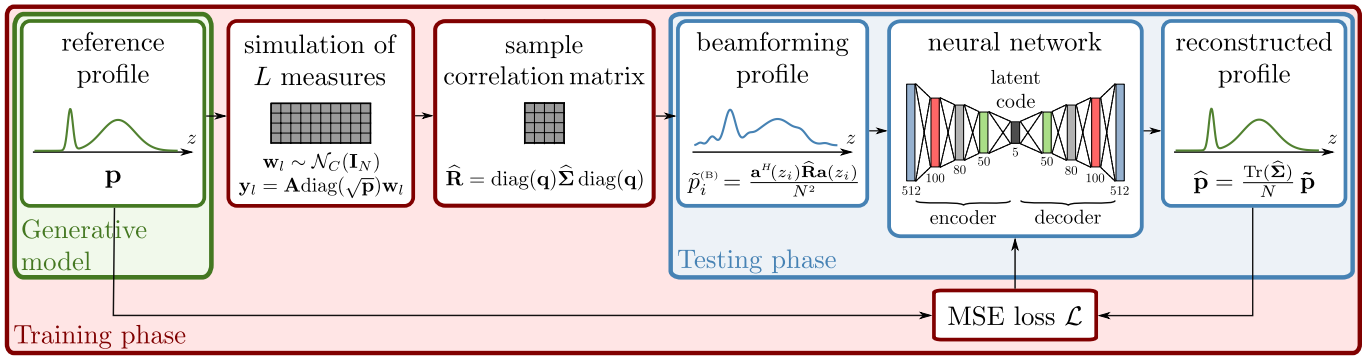


Fig. 1. Pipeline of the data simulation, training and testing process of the proposed method.

methods [8], [9]. These peaks can be represented by many basis functions, of which the simplest case is that of two Gaussian functions. The proposed approach therefore aims at learning a latent representation of a forest reflectivity profile from simulated profiles composed of two Gaussian terms. The parameters of these Gaussian shapes are adapted to the type of forests observed in our SAR datasets for each training, i.e. boreal forests in L band and tropical forests in P band. The ranges of these parameters, presented in Table I, are deliberately large, in order to cover all the variations present in forests as well as potential pre-processing errors, and are further discussed in section IV.

III. Experiments

A. Simulated data

10000 simulated profiles $\mathbf{p}$ sampled on N_z uniformly spaced heights were generated with a mixture of two Gaussians: $\mathbf{p}(z_i) = \frac{r}{\sqrt{2\pi}\sigma_1}\exp(-\frac{(z_i-\mu_1)^2}{2\sigma_1^2}) + \frac{1-r}{\sqrt{2\pi}\sigma_2}\exp(-\frac{(z_i-\mu_2)^2}{2\sigma_2^2})$, with the 5 parameters $\mu_1, \sigma_1, \mu_2, \sigma_2$, drawn according to a uniform distribution in intervals defined by the type of forest studied (Table I), and $r \in [0,1]$ defining the proportion of the ground contribution. The covariance matrices $\mathbf{\Sigma} = \mathbf{A}\mathrm{diag}(\mathbf{p})\mathbf{A}^H$ were built from steering matrices $\mathbf{A}$ of an actual tomographic dataset, each steering matrix corresponding to a pair of azimuth and range values drawn uniformly. Circular complex Gaussian samples $\{\mathbf{y}_l\}_{l=1..L}$ were then drawn according to these covariance matrices (by multiplying a white speckle noise $\mathbf{w}_l$ by $\mathbf{A}\mathrm{diag}(\sqrt{\mathbf{p}})$). Sample correlation matrices $\widehat{\mathbf{R}}$ were computed from each set of L samples $\mathbf{y}_l \in \mathbb{C}^N$. The *beamforming* profiles were then reconstructed:

$$\tilde{p}_i^{(\mathrm{B})} = \frac{\mathbf{a}^H(z_i)\widehat{\mathbf{R}}\mathbf{a}(z_i)}{N^2} \tag{4}$$

where $\mathbf{a}$ is a steering column vector of the matrix $\mathbf{A}$. These *beamforming* profiles were then used as input by the neural network.

TABLE I
Range values for the uniform sampling of the 5 parameters of the Gaussian mixture model, in the case of boreal forests (L-band) and tropical forests (P-band).

Model parameter	μ_1 (m)	σ_1 (m)	μ_2 (m)	σ_2 (m)	r
Boreal forest ranges	$[-5, 5]$	$[0.1, 2]$	$[-2, 20]$	$[0.5, 4]$	$[0, 1]$
Tropical forest ranges	$[-10, 10]$	$[0.1, 2]$	$[0, 40]$	$[0.5, 4]$	$[0, 1]$

The trained network consists of a 4 linear layer encoder, with a decreasing number of neurons from $N_z = 512$ up to a latent space of size 5, and a symmetrical decoder with 4 linear layers. The depth of the network is voluntarily kept small and has been empirically validated, as discussed in section III-A2. Each layer is unbiased and is followed by a leaky ReLU activation function. Training was performed on the simulated *beamforming* profile dataset, divided into training (75%) and validation (25%) sets, in mini-batches of size 32 with an Adam optimizer and a learning rate of 10^{-3} for 200 epochs.

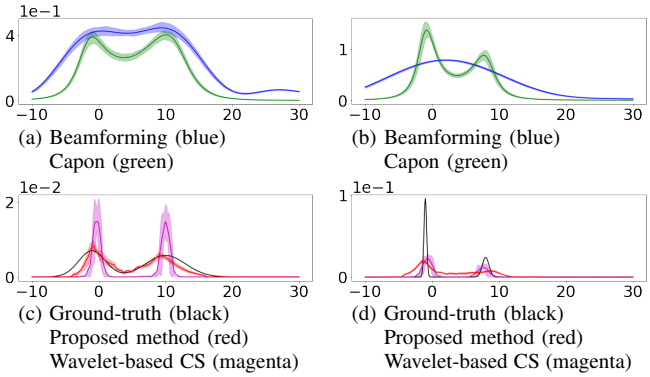
In the following subsections, the results presented were computed using the steering matrices of the BioSAR-2 campaign, more thoroughly presented in section III-B.

1) Analysis of the reconstructed profiles: Fig. 2 shows, for two different simulated profiles, the reconstructions obtained with the *beamforming* algorithm, *Capon*'s filter, *wavelet-based CS* and our method for $L = 100$ speckle realizations in the measurement simulation. As we are interested in comparing the shapes of the profiles reconstructed by each algorithm and since they do not compute the same physical quantity, the methods corresponding to the output power of a filter are plotted above and the reconstructed and reference reflectivity profiles below. The left profile composed of two large Gaussians favors the spectral estimation methods and shows that the *wavelet-based CS* fails to reconstruct large lobes, while the profile on the right simulates very narrow responses, more suitable for the latter but only roughly reconstructed by *Beamforming* for the level of noise considered ($L = 100$). The neural network produces profiles with a width that follows more closely the ground-truth profiles in each case.

2) Architecture choice: The chosen architecture features a latent space, i.e. an embedding of each profile in a space that brings similar profiles closer together. The use of a low-dimensional latent space reduces the risk of overfitting and increases the generalization capability of the network. Different sizes for this latent space (number of parameters at the output of the encoder) were tested to check whether the size of the latent space should be kept close to the number of parameters of the generative model. The number of looks L was set to 100 for this study. The reconstruction errors, compared to a normalized error between the input *beamforming* $\widehat{\mathbf{p}}^{(\mathrm{B})}$ and reference profile $\mathbf{p}$, $\|\widehat{\mathbf{p}}^{(\mathrm{B})} \frac{\langle \widehat{\mathbf{p}}^{(\mathrm{B})}, \mathbf{p}\rangle}{\langle \widehat{\mathbf{p}}^{(\mathrm{B})}, \widehat{\mathbf{p}}^{(\mathrm{B})}\rangle} - \mathbf{p}\|_2^2$, where $\langle \cdot\,, \cdot \rangle$ is the L_2 scalar product, are presented in Table II. These experiments show that a latent space size smaller than 5

TABLE II
MEAN AND STANDARD DEVIATION OF THE MEAN SQUARED ERROR (MSE) VALUES AFTER 20 TRAININGS DURING 200 EPOCHS FOR EACH DIFFERENT LATENT SPACE SIZES, COMPARED TO A NORMALIZED ERROR BETWEEN THE INPUT BEAMFORMING AND REFERENCE PROFILE.

Latent space size	**3**	**4**	**5**	**6**	**8**	**10**	**15**	**20**
MSE ($\times 10^{-1}$)	5.12 ± 0.10	4.74 ± 0.12	4.65 ± 0.15	4.68 ± 0.19	4.64 ± 0.16	4.68 ± 0.18	4.67 ± 0.15	4.68 ± 0.13



(a) Beamforming (blue) Capon (green)
(b) Beamforming (blue) Capon (green)
(c) Ground-truth (black) Proposed method (red) Wavelet-based CS (magenta)
(d) Ground-truth (black) Proposed method (red) Wavelet-based CS (magenta)

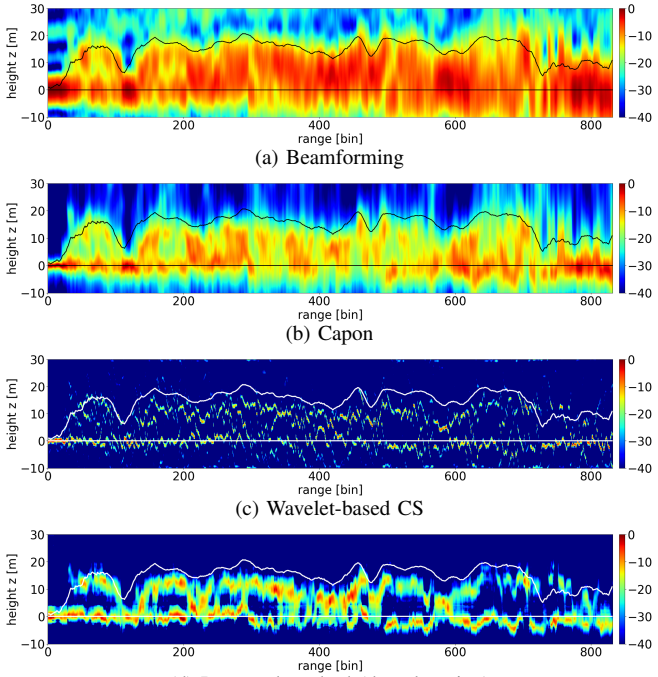
Fig. 2. Reconstruction of a profile consisting of wide (left) and narrow (right) Gaussian reflectivity functions with Beamforming, Capon, Wavelet-based CS and our method. Average profiles and their interquartile ranges are computed using 100 simulated measurements, each of which uses $L = 100$ speckle realizations. Reference profiles are shown in black.

(number of parameters of our generative model) increases the reconstruction error, whereas a larger latent space brings only marginal improvements.

B. Boreal forest at L band

Testing was first performed on a tomographic stack of 6 airborne L-band SAR images of a boreal forest in northern Sweden, acquired during the BioSAR-2 campaign led by the DLR in 2008 [14], with a vertical resolution varying from 6 m in near range to 25 m in far range. A local window of around 60 looks is used to compute the covariance matrix used by the *beamforming* algorithm to compute the tomograms. Fig. 3 shows an example of a tomogram in HH polarization reconstructed with the proposed method, after compensation of the topography. It is compared to the results obtained with *beamforming*, *Capon* and the *wavelet-based CS* method developed in [8], showing that the trained network effectively improves the reconstructed profiles over spectral estimation methods, while maintaining a representative volume of the tree crown. This is not the case with the *wavelet-based CS*, which reconstructs narrow peaks and can therefore predict several peaks when this volume is large, depending on the choice of the regularization parameter value.

The orders of magnitude of the computational time required to compute these tomograms are given in Table III (all computations are done on one CPU, even the network training). The *wavelet-based CS* approach was computed using the optimisation library CVX and the default solver SDPT3 [15]. These statistics highlight one of the major advantages of using deep learning to reconstruct forest reflectivity profiles, which is the gain in time of this method compared to classical optimization algorithms. This feature will be even more important when scaling up and applying tomographic reconstruction on an entire SAR image.



(a) Beamforming
(b) Capon
(c) Wavelet-based CS
(d) Proposed method (deep learning)

Fig. 3. Tomographic profile estimated over a boreal forest at L band using: (a) Beamforming; (b) Capon; (c) Wavelet-based CS; (d) the proposed neural network-based approach. LiDAR estimates of (compensated) ground and tree top heights are shown as continuous lines.

TABLE III
ORDER OF MAGNITUDE OF THE COMPUTATION TIME FOR THE RECONSTRUCTION OF A TOMOGRAM OF SIZE 1.4 km × 1.6 m ON A CPU.

Method		**Computation time (s)**
Beamforming		2
Capon		3
Wavelet-based CS (CVX)		1500
Proposed method	**Training**	200
	Inference	3

C. Tropical forest at P band

Other tests were performed on airborne tomographic SAR data acquired at P band by the ONERA over the test site of Paracou in French Guiana, during the TropiSAR campaign in 2009 [16], also comprising 6 tracks, with a vertical resolution of around 15 m. The correlation matrix used to compare the different methods is derived from 56 looks. Some results and comparisons in HH polarization are shown in Fig. 4.

IV. DISCUSSION

During training, the network is fed with *beamforming* profiles computed from samples generated according to a model of covariance matrix. This model, defined in equation (2), neglects several phenomena such as the integration of backscattered complex amplitudes produced by scatterers located

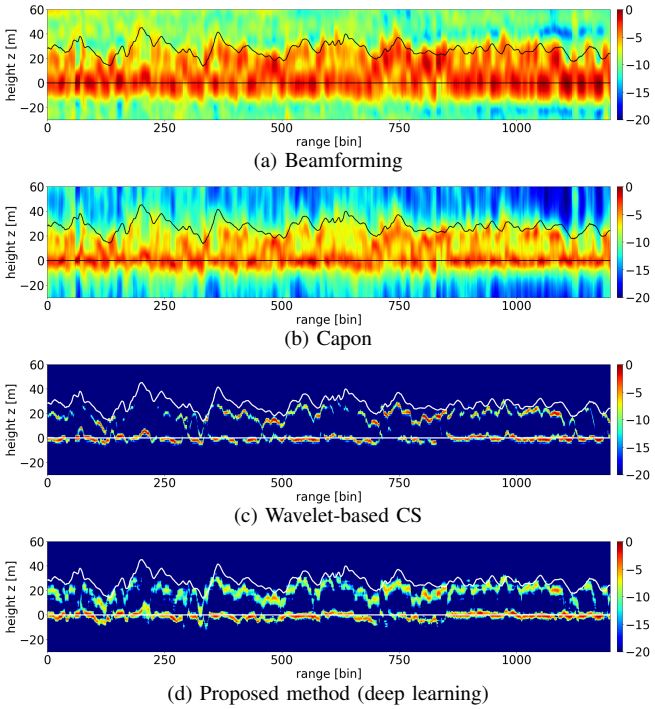


Fig. 4. Tomographic profile estimated over a tropical forest at P band using: (a) Beamforming; (b) Capon; (c) Wavelet-based CS; (d) the proposed neural network-based approach. LiDAR estimates of (compensated) ground and tree top heights are shown as continuous lines.

in the neighborhood according to the SAR impulse response, or the temporal decorrelation between acquisitions. It could be refined, but our results on real data already show a good generalization ability of the network trained with this simple covariance model.

For the training, we used the steering matrix of a specific campaign to simulate measurements and *beamforming* profiles that match at best the actual SAR data considered at test time. The network is thus trained for a given geometric configuration and the efficiency of the network may be affected when it differs, in which case a retraining would be required. However, this step is relatively inexpensive in terms of computational time (a few minutes), and tomographic data acquisition configurations in a given spectral band vary only slightly due to hardware constraints and the possibility of ambiguities.

In our experiments, the network was trained to reconstruct profiles composed of two Gaussians. It cannot therefore perfectly recover a profile with a single or an extra peak, even with input profiles composed of three very distinct Gaussians. The use of a refined simulation model could further improve the reconstruction quality, which encourages current research on a model representative of different forest types [9].

V. Conclusion

In this study, we explored the possibility of using deep learning approaches to improve the performance of the tomographic inversion task on forests. Tests on real data show promising results, both in terms of quality and in terms of the computation time needed to reconstruct a large image, with an overhead for the feedforward pass negligible compared to the computation of the *beamforming* profile and an acceleration by several orders of magnitude with respect to iterative regularized inversion algorithms. This will be essential for the systematic processing of future data from the ESA BIOMASS [17] mission. It confirms the strong potential of the application of deep learning in this field to super-resolve existing reconstructions.

It could also be interesting to consider directly using the measured intensities and acquisition geometry as input to the neural network, to avoid the need to train a different network for each tomographic configuration.

References

[1] L. Ferro-Famil, Y. Huang, and E. Pottier, "Principles and applications of polarimetric SAR tomography for the characterization of complex environments," *International Association of Geodesy Symposia. F. Sanso Ed.., Springer-Verlag*, vol. 142, no. 1-13, pp. 243–255, 2016.

[2] F. Gini and F. Lombardini, "Multibaseline cross-track SAR interferometry: a signal processing perspective," *IEEE Aerospace and Electronic Systems Magazine*, vol. 20, no. 8, pp. 71–93, 2005.

[3] Y. Huang, J. Lévy-Vehel, L. Ferro-Famil, and A. Reigber, "Three-dimensional imaging of objects concealed below a forest canopy using SAR tomography at L-band and wavelet-based sparse estimation," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 9, pp. 1454–1458, Sep. 2017.

[4] A. Budillon, A. Evangelista, and G. Schirinzi, "Three-dimensional SAR focusing from multipass signals using compressive sampling," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 1, pp. 488–499, 2011.

[5] C. Rambour, A. Budillon, A. C. Johnsy, L. Denis, F. Tupin, and G. Schirinzi, "From interferometric to tomographic SAR: A review of synthetic aperture radar tomography-processing techniques for scatterer unmixing in urban areas," *IEEE Geosci. Remote Sens. Mag.*, vol. 8, no. 2, pp. 6–29, 2020.

[6] H. Aghababaee, G. Ferraioli, L. Ferro-Famil, Y. Huang, M. Mariotti D'Alessandro, V. Pascazio, G. Schirinzi, and S. Tebaldini, "Forest SAR tomography: Principles and applications," *IEEE Geosci. Remote Sens. Mag.*, vol. 8, no. 2, pp. 30–45, Feb. 2020.

[7] P. Stoica and R. L. Moses, *Spectral analysis of signals*. Upper Saddle River, NJ: Prentice Hall, 2005.

[8] E. Aguilera, M. Nannini, and A. Reigber, "Wavelet-based compressed sensing for SAR tomography of forested areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 12, pp. 5283–5295, 2013.

[9] L. Ferro-Famil, Y. Huang, and N. Ge, "Estimation of the vertical structure of a tropical forest using basis functions and parametric SAR tomography," in *IGARSS*, 2022.

[10] K. Qian, Y. Wang, Y. Shi, and X. X. Zhu, "Super-resolving SAR tomography using deep learning," in *IGARSS*, 2021, pp. 4810–4813.

[11] Y. Hu, X. Zhang, S. Wei, Y. Ren, N. Wang, and J. Shi, "Adatomo-Net: a novel deep learning approach for SAR tomography imaging and autofocusing," in *IGARSS*, 2022.

[12] W. Yang, S. Vitale, H. Aghababaei, G. Ferraioli, V. Pascazio, and G. Schirinzi, "A deep learning solution for height estimation on a forested area based on Pol-TomoSAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, pp. 1–14, 2023.

[13] Y. Zérah, S. Valero, and J. Inglada, "Physics-guided interpretable probabilistic representation learning for high resolution image time series," *IEEE Trans. Geosci. Remote Sens.*, 2022.

[14] I. Hajnsek, R. Scheiber, M. Keller, R. Horn, S. Lee, L. Ulander, A. Gustavsson, G. Sandberg, T. Toan, S. Tebaldini, A. Guarnier, and F. Rocca, *BIOSAR 2008 technical assistance for the development of airborne SAR and geophysical measurements during the BioSAR 2008 experiment, final report*, 2009.

[15] M. Grant and B. SP, "CVX: MATLAB software for disciplined convex programming," 2014. [Online]. Available: http://cvxr.com/cvx

[16] P. Dubois-Fernandez, T. L. Toan, S. Daniel, H. M. Oriot, J. Chave, L. Blanc, L. Villard, M. Davidson, and M. Petit, "The TropiSAR airborne campaign in French Guiana: objectives, description, and observed temporal behavior of the backscatter signal," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, pp. 3228–3241, 2012.

[17] K. Fletcher and H. Rider, "Report for mission selection: an earth explorer to observe atmospheric composition. BIOMASS." ser. ESA SP-1324/1 (3 volume series), European Space Agency, Noordwijk, The Netherlands, 5 2012.