

# Apprentissage autosupervisé pour le despeckling d'images SAR avec MERLIN : application aux images Sentinel-1 Stripmap

Emanuele DALSSASSO<sup>1</sup>, Loïc DENIS<sup>2</sup>, Florence TUPIN<sup>1</sup>

<sup>1</sup>LTCI, Télécom Paris, Institut Polytechnique de Paris, Palaiseau, France

<sup>2</sup>Univ Lyon, UJM-Saint-Etienne, CNRS, Institut d'Optique Graduate School,  
Laboratoire Hubert Curien UMR 5516, F-42023, SAINT-ETIENNE, France

emanuele.dalsasso@telecom-paris.fr

loic.denis@univ-st-etienne.fr, florence.tupin@telecom-paris.fr

**Résumé** – Les fluctuations dues au phénomène de chatoiement sont un frein à l'interprétation des images acquises par un Radar à Synthèse d'Ouverture (RSO, aussi appelé SAR en anglais). Pour faciliter l'utilisation de ces images pour l'observation de la Terre, il est important de réduire le chatoiement. Nous décrivons ici MERLIN, une nouvelle méthode permettant d'entraîner de manière auto-supervisée un réseau neuronal convolutif pour la réduction de chatoiement exploitant la décomposition d'une image SAR complexe en partie réelle et imaginaire. Nous présentons l'adaptation de ce cadre aux images Sentinel-1 acquises en mode Stripmap.

**Abstract** – Speckle fluctuations limit the interpretation of Synthetic Aperture Radar (SAR) images. To ease their exploitation for Earth Observation purposes, speckle must be suppressed. In this article we describe MERLIN, a new self-supervised training strategy for speckle reduction relying on the decomposition of a single-look complex (SLC) SAR image into real and imaginary part. We present the adaptation of this framework to Sentinel-1 images.

## 1 Introduction

Le chatoiement est un phénomène physique lié à la présence de plusieurs rétro-diffuseurs élémentaires dans la même case radar. La somme cohérente de ces contributions donne lieu à de très fortes fluctuations dans le signal mesuré, ce qui dégrade la qualité des images SAR. La réduction de chatoiement a fait l'objet de nombreux travaux de recherche [1].

Le chatoiement peut être réduit en moyennant des échantillons (à l'intérieur d'une fenêtre, le long de l'axe temporel ou en combinant des patchs au sein de l'image selon un critère de similarité [2]). Les méthodes variationnelles et bayésiennes quant à elles cherchent à filtrer une image SAR en minimisant un terme d'attache aux données et un terme de régularisation [3]. Plus récemment, les avancées des approches d'apprentissage profond dans le domaine de la vision par ordinateur ont suscité un vif intérêt dans la communauté radar [4]. Suivant la stratégie d'entraînement, on peut distinguer trois familles de méthodes: (i) un réseau neuronal convolutif peut être entraîné de manière supervisée si l'on dispose de paires d'images sans chatoiement (dites de vérité terrain) et avec chatoiement; (ii) un entraînement auto-supervisé peut s'appuyer sur des paires d'images acquises sur la même zone à des dates différentes (présentant un chatoiement différent); (iii) le but ultime est de parvenir à entraîner un réseau avec une stratégie auto-supervisée ne nécessitant qu'une seule date (méthodes auto-supervisée mono-image).

Une des limites des méthodes supervisées est la génération d'images de vérité terrain associées aux images avec chatoiement. La plupart des techniques appartenant à cette catégorie repose sur la génération d'un chatoiement synthétique, appliqué soit sur des images optiques, soit sur des super-images (c.à.d. la version obtenue en moyennant le chatoiement sur une longue série temporelle d'images SAR). Cependant, cette simulation de chatoiement néglige généralement le phénomène de la corrélation spatiale introduite suite à l'application de la fonction du transfert du système SAR. Par conséquent, l'application de ces méthodes aux images SAR nécessite une étape de pré-traitement réduisant la corrélation spatiale du chatoiement (par inversion de la fonction de transfert SAR ou sous-échantillonnage).

En ce qui concerne les méthodes auto-supervisées multi-images, le principe est le suivant: lorsqu'un signal contient une composante déterministe et une composante aléatoire, un réseau entraîné à prédire une nouvelle réalisation de ce signal à partir d'une première réalisation indépendante ne pourra prédire que la composante déterministe, c.à.d. la scène sous-jacente, apprenant ainsi à réduire le chatoiement. Les approches auto-supervisées multi-images exploitent des images acquises à des dates différentes (suffisamment espacées pour que le chatoiement soit temporellement décorrélé). Ainsi l'on obtient plusieurs échantillons représentant la même scène sous-jacente, à condition de pouvoir compenser les changements [5].

Dans les approches auto-supervisées mono-image, on utilise

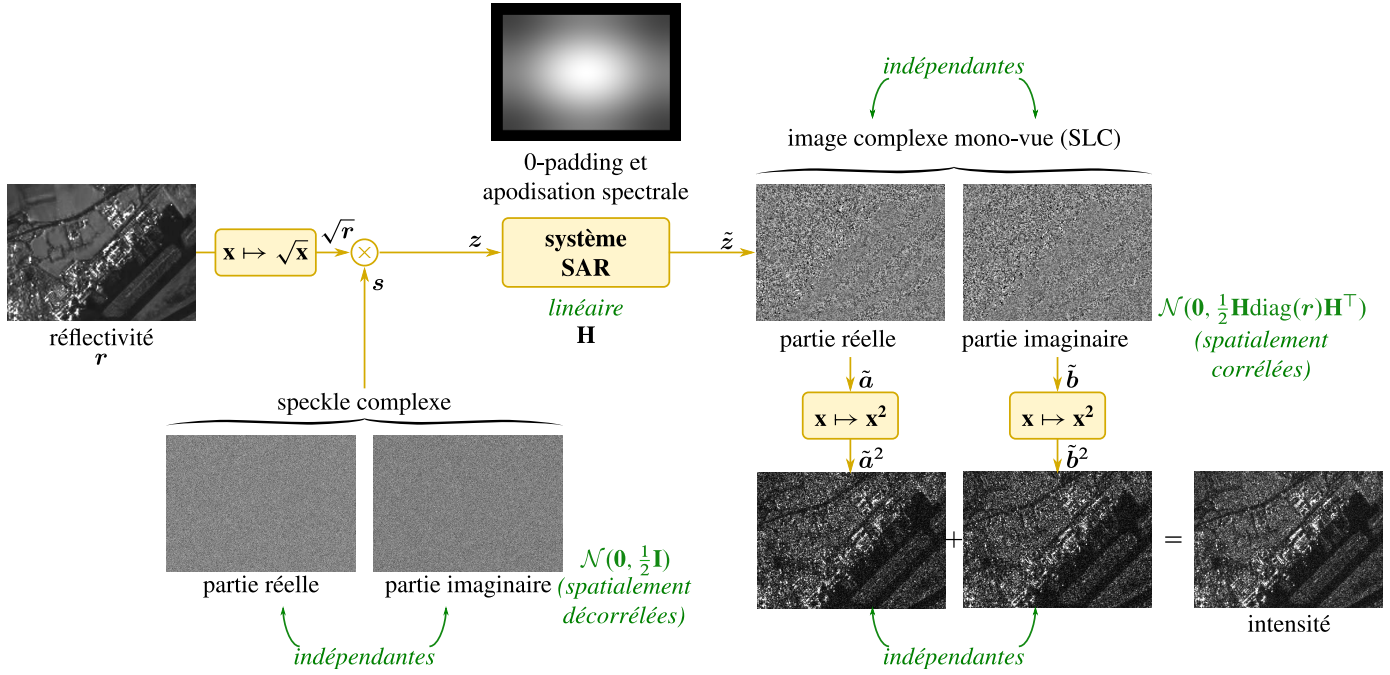


Figure 1: Le modèle statistique du chatoiement dans une image SAR.

des architectures spécifiques permettant d'exclure le pixel central du champ réceptif du réseau (on parle alors de *blind-spot*), ceci étant utilisé pour superviser l'entraînement du réseau. La particularité du champ réceptif de ces réseaux contraint leur architecture et limite ses performances. De plus, le chatoiement doit être spatialement décorrélé pour vérifier la condition d'indépendance requise dans ces stratégies auto-supervisées.

Dans cet article, nous proposons d'adapter la méthode MERLIN (coMplex sElf-supeRvised despeckLING) que nous avons proposée très récemment [6] afin de réduire le chatoiement d'un autre capteur: le satellite européen Sentinel-1 dans une modalité Stripmap. Cette méthode permet d'entraîner un réseau ayant une architecture quelconque avec des dates uniques d'images SAR complexes (SLC). Elle permet donc de relâcher les contraintes posées par les méthodes auto-supervisées discutées précédemment : aucune hypothèse sur la cohérence temporelle d'une paire ou série d'images ni sur l'indépendance spatiale du chatoiement n'est faite. Avec MERLIN, on montre que la phase d'une image SAR complexe, souvent considérée comme inutile lorsque l'on ne s'intéresse qu'à l'intensité radar (c.à.d. à l'exception de l'interférométrie SAR), devient étonnamment utile pour extraire d'une image SLC deux images dont le chatoiement est indépendant et identiquement distribué. Ceci crée le cadre idéal pour la mise en œuvre d'une stratégie auto-supervisée mono-image de type noise2noise [7].

## 2 Modèle statistique d'une image SAR

Selon le modèle de Goodman du chatoiement pleinement développé [8], l'amplitude complexe  $z = \rho \exp(j\phi)$  en un

pixel donné suit une distribution normale circulaire complexe:

$$p_Z(z) = \frac{1}{\pi r} \exp(-|z|^2/r), \quad (1)$$

avec  $r$  la réflectivité de ce pixel. La décomposition de  $z$  en partie réelle et imaginaire conduit à:

$$\begin{aligned} p_Z(z) &= p_Z(a + jb) = \frac{1}{\pi r} \exp(-(a^2 + b^2)/r) \\ &= \underbrace{\frac{1}{\sqrt{2\pi} \sqrt{r/2}} \exp(-a^2/r)}_{\mathcal{N}(0, r/2)} \underbrace{\frac{1}{\sqrt{2\pi} \sqrt{r/2}} \exp(-b^2/r)}_{\mathcal{N}(0, r/2)}. \end{aligned} \quad (2)$$

Ceci montre que la partie réelle et imaginaires d'une amplitude complexe sont indépendantes et identiquement distribuées selon une loi normale de variance  $r/2$ . Étant donné la nature multiplicative du phénomène de chatoiement  $s$ , avec  $z = s\sqrt{r}$ , ceci équivaut à dire que la partie réelle et imaginaire du chatoiement complexe  $s$  sont i.i.d. selon une loi normale de variance égale à  $1/2$ .

Le modèle multiplicatif est illustré dans la partie gauche de la figure 1. D'un pixel à l'autre, la réalisation de chatoiement est différente et le champ aléatoire  $s \in \mathbb{C}^K$  d'une image ayant  $K$ -pixels est un champ gaussien blanc.

Suivant le mode d'acquisition, l'échantillonnage et la fonction d'apodisation appliquée pour réduire les lobes secondaires autour des points brillants, la fonction de transfert SAR associée à ces opérations transforme  $z$  en un champ spatialement corrélé  $\tilde{z}$ :

$$\tilde{z} = \mathbf{H}z, \quad (3)$$

avec  $\mathbf{H}$  l'opérateur spatial associé à la fonction de transfert du système SAR. Si l'on suppose  $\mathbf{H}$  réel (pour un système invari-

ant par décalage, ceci correspond à une réponse fréquentielle à symétrie hermitienne), par linéarité on obtient:

$$\tilde{z} = \tilde{a} + j\tilde{b} \quad \text{avec} \quad \tilde{a} = \mathbf{H}\mathbf{a} \quad \text{et} \quad \tilde{b} = \mathbf{H}\mathbf{b}, \quad (4)$$

c.à.d. que  $\tilde{a}$  et  $\tilde{b}$  sont des champs spatialement corrélés mais *mutuellement indépendants*, distribués selon  $\mathcal{N}(\mathbf{0}, \frac{1}{2}\mathbf{H}\text{diag}(\mathbf{r})\mathbf{H}^\top)$ , avec  $\text{diag}(\mathbf{r})$  la matrice diagonale de taille  $K \times K$  dont la diagonale est égale au vecteur des réflectivités  $\mathbf{r} \in \mathbb{R}_{+*}^K$ . Pour conclure, l'intensité est obtenue en additionnant les parties réelles et imaginaires.

### 3 L'approche MERLIN: réduction de chatolement auto-supervisée

Dans la section 2 nous avons présenté le modèle statistique d'une image SAR complexe. Cette image peut ainsi être décomposée en deux composantes i.i.d., chacune contenant la moitié de l'information d'une image en intensité. Il en découle que l'on peut entraîner un réseau de manière auto-supervisée comme suit: le réseau  $f_\theta(\cdot)$  prend en entrée une composante (p.ex la partie réelle  $\tilde{a}$ ) et on évalue la qualité de l'image restaurée ( $\tilde{r} = f_\theta(\tilde{a})$ ) par rapport à l'autre composante (p.ex la partie imaginaire  $\tilde{b}$ ). Une expression naturelle pour la fonction de coût d'entraînement est la vraisemblance de  $\tilde{b}$  par rapport à l'estimation du réseau  $\tilde{r}$ :

$$\mathcal{L}(\tilde{r}, \tilde{b}) = \sum_k -\log p(\tilde{b}_k | \tilde{r}_k) = \sum_k \frac{1}{2} \log(\tilde{r}_k) + \frac{\tilde{b}_k^2}{\tilde{r}_k}, \quad (5)$$

où l'indice  $k$  désigne le  $k$ -ème pixel de l'image des réflectivités estimées par le réseau  $\tilde{r}$  ou la partie imaginaire  $\tilde{b}$  de l'image observée. En pratique, lors de l'entraînement du réseau, les parties réelles et imaginaires sont permutées, du fait de leur indépendance.

Une fois le réseau entraîné, il peut être appliqué pour réduire le chatolement d'une image SAR complexe. Pour cela, les parties réelles et imaginaires sont traitées indépendamment par le même réseau. Les deux estimations  $f_\theta(\tilde{a})$  et  $f_\theta(\tilde{b})$  sont enfin moyennées pour obtenir l'estimation finale (figure 2).

#### 3.1 Condition d'indépendance des parties réelle et imaginaire d'une image SAR complexe

Comme évoqué dans la section 2, pour vérifier les conditions d'indépendance entre la partie réelle et imaginaire d'une image SAR complexe, l'opérateur spatial  $\mathbf{H}$  doit être réel. Cela implique que, pour un système invariant par décalage, la fonction de transfert doit vérifier la symétrie hermitienne. Du fait d'un support spectral non centré ou d'un dépointage continu de l'antenne au fil de l'acquisition (p.ex mode SpotLight, mode TOPS), la réponse du système SAR peut être asymétrique, corrélant les parties réelles et imaginaires.

Afin de corriger une asymétrie due à un support spectral non centré autour de la fréquence 0-Doppler, il est nécessaire d'effectuer une étape de recentrage du spectre. Pour cela, les

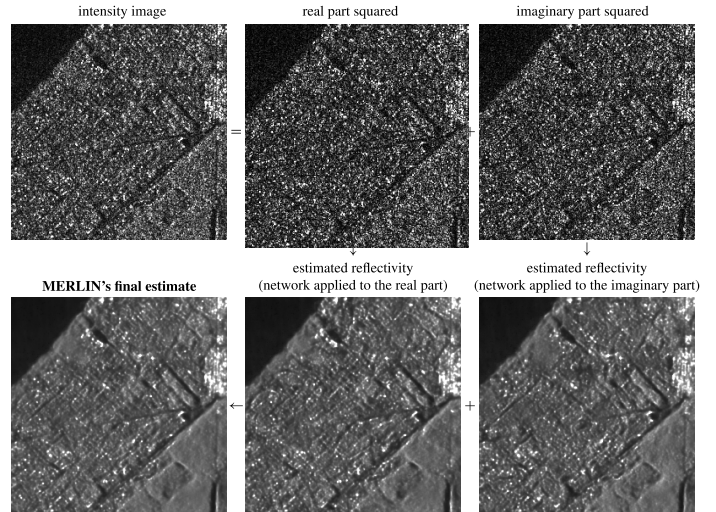


Figure 2: Résultat sur une image Sentinel-1 stripmap au Brésil

profils spectraux dans les directions azimut et distance sont calculés. On note  $\mathbf{p}$  un de ces profils. On cherche alors à trouver le paramètre  $\delta$  maximisant la corrélation entre le profil translaté de  $\delta$  pixels  $T_\delta\{\mathbf{p}\}$  et son symétrique  $\mathcal{S}\{T_\delta\{\mathbf{p}\}\}$ :  $\hat{\delta} = \arg \max_\delta T_\delta\{\mathbf{p}\}^\top \mathcal{S}\{T_\delta\{\mathbf{p}\}\}$ . Ainsi, pour le décalage  $\delta$ , le profil translaté est recentré et se superpose parfaitement à son symétrique, vérifiant la condition  $\mathbf{H}$  réel qui garantit l'indépendance de la partie réelle et imaginaire (figure 3).

#### 3.2 Cas des images Sentinel-1 en modalité Stripmap

Dans le mode d'acquisition Sentinel-1 Stripmap, le système de transfert SAR est invariant par décalage. Il est donc possible d'effectuer un recentrage globale du spectre sur toute l'image. Ceci n'est pas le cas des images Sentinel-1 acquises en modalité TOPS : en effet, un dépointage continu de l'antenne est opéré, permettant des fauchées plus grandes qu'en mode Stripmap (au détriment de la résolution). Il est alors nécessaire d'effectuer d'abord une opération dite de *deramping* correspondant à une correction quadratique de la phase [9].

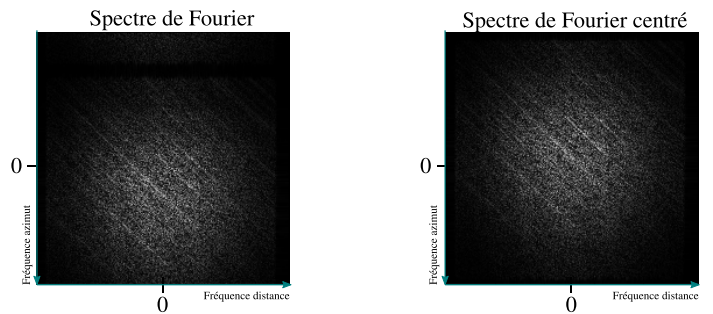
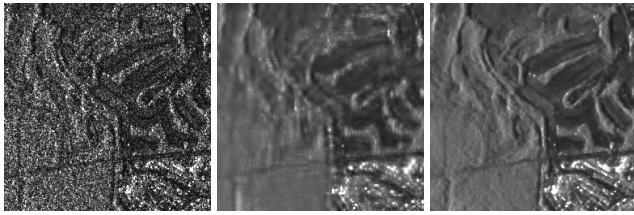


Figure 3: Le spectre de Fourier d'une image SAR avant et après l'étape de recentrage nécessaire pour garantir l'indépendance des parties réelles et imaginaires.

Avant de pouvoir entraîner MERLIN sur des données



(a) Image SAR (b) Résultat de SAR-BM3D [10] (c) Résultat de MERLIN (cet article)

Figure 4: Résultat du filtrage d’une image SAR acquise par Sentinel-1 en mode Stripmap au Brésil.

Sentinel-1 Stripmap, le spectre doit être recentré par translation, comme illustré dans la section 3.1. Une fois les parties réelles et imaginaires décorréées, un réseau de type U-Net a été entraîné sur une image de taille  $17453 \times 47890$  pixels acquise au Brésil en polarisation VV. L’image a été découpée en 50367 patches de taille  $256 \times 256$  pixels avec un stride de 128 pixels, regroupés en 4197 groupes de 12 patches chacun. Le réseau a été entraîné sur 20 époques, avec un *learning rate* initial de 0.001, divisé par 10 après 2 époques.

Dans la figure 4, le résultat produit par MERLIN est comparé à celui obtenu par la méthode SAR-BM3D [10]. Cette dernière suppose que le chatoiement soit blanc. Pour éviter l’apparition d’artefacts dus à la corrélation spatiale du chatoiement, l’image a été préalablement sous-échantillonnée d’un facteur 2. Par conséquent, la résolution a été dégradée et certaines structures fines n’ont pas pu être restaurées (comme la route qui coupe l’image de droite à gauche). En revanche, MERLIN a été entraîné spécifiquement pour le mode d’acquisition Sentinel-1 Stripmap et peut donc être directement appliqué sur l’image, sans ré-échantillonnage. Le filtrage effectué avec MERLIN montre la robustesse du réseau à la corrélation spatiale du chatoiement : les fluctuations ont été fortement réduites dans les zones homogènes et aucun artefact visible n’a été créé. De plus, les détails sont très bien préservés. Ceci montre l’intérêt d’utiliser la stratégie d’entraînement auto-supervisé MERLIN pour réduire le chatoiement dans les images SAR et ainsi faciliter leur interprétation.

## 4 Conclusion

Les images SAR complexes offrent un cadre idéal pour l’apprentissage auto-supervisé en s’appuyant sur la décomposition de l’intensité en partie réelle et imaginaire (après une étape simple de pré-traitement permettant de rendre ces deux composantes mutuellement indépendantes). L’approche illustrée, MERLIN, est générique : un réseau de neurones profond d’architecture quelconque peut être entraîné pour chaque mode d’acquisition spécifique. Une seule image SAR est suffisante pour mener à bien un entraînement. De plus, l’utilisation directe des images radar acquises par le satellite permet de rendre le réseau robuste à la corrélation spatiale du chatoiement : la résolution spatiale des images est ainsi préservée.

Libérés de la préoccupation de créer une base d’entraînement contenant des vérités terrain, les travaux futurs pourront se concentrer sur le développement d’architectures plus adaptées à la réduction de chatoiement dans les images SAR. De plus, elles pourront facilement être entraînées sur les immenses bases d’images disponibles auprès des agences spatiales pour chaque mode d’acquisition, ce qui pourrait conduire à des performances inédites.

Le code de MERLIN est disponible en accès libre à l’adresse suivante : [gitlab.telecom-paris.fr/RING/MERLIN](https://gitlab.telecom-paris.fr/RING/MERLIN).

## References

- [1] F. Argenti, A. Lapini, T. Bianchi, and L. Alparone, “A tutorial on speckle reduction in synthetic aperture radar images,” *IEEE GRSM*, vol. 1, no. 3, pp. 6–35, 2013.
- [2] C.-A. Deledalle, L. Denis, G. Poggi, F. Tupin, and L. Verdoliva, “Exploiting patch similarity for SAR image processing: The nonlocal paradigm,” *IEEE SPM*, vol. 31, no. 4, pp. 69–78, 2014.
- [3] C.-A. Deledalle, L. Denis, S. Tabti, and F. Tupin, “Mulog, or how to apply gaussian denoisers to multi-channel sar speckle reduction?” *IEEE TIP*, vol. 26, no. 9, pp. 4389–4403, 2017.
- [4] G. Fracastoro, E. Magli, G. Poggi, G. Scarpa, D. Valsesia, and L. Verdoliva, “Deep Learning Methods For Synthetic Aperture Radar Image Despeckling: An Overview Of Trends And Perspectives,” *IEEE GRSM*, vol. 9, no. 2, pp. 29–51, 2021.
- [5] E. Dalsasso, L. Denis, and F. Tupin, “SAR2SAR: A semi-supervised despeckling algorithm for SAR images,” *IEEE JSTARS*, vol. 14, pp. 4321–4329, 2021.
- [6] —, “As If by Magic: Self-Supervised Training of Deep Despeckling Networks With MERLIN,” *IEEE TGRS*, vol. 60, pp. 1–13, 2022.
- [7] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila, “Noise2noise: Learning image restoration without clean data,” in *ICML*, 2018.
- [8] J. W. Goodman, “Some fundamental properties of speckle,” *JOSA*, vol. 66, no. 11, pp. 1145–1150, 1976.
- [9] R. Abergel, L. Denis, S. Ladjal, and F. Tupin, “Subpixellic methods for sidelobes suppression and strong targets extraction in single look complex SAR images,” *IEEE JS-TARS*, vol. 11, no. 3, pp. 759–776, 2018.
- [10] S. Parrilli, M. Poderico, C. V. Angelino, and L. Verdoliva, “A nonlocal SAR image denoising algorithm based on LLMMSE wavelet shrinkage,” *IEEE TGRS*, vol. 50, no. 2, pp. 606–616, 2011.