

Approche bayésienne pour la détection de la déforestation à partir de séries temporelles d'images SAR polarimétriques Sentinel-1

Marta BOTTANI^{1,2,3,4} Laurent FERRO-FAMIL^{2,4}

¹Laboratoire TéSA, 7 boulevard de la gare, 31500, Toulouse, France

²ISAE Supaero, université de Toulouse, 1 avenue Marc Pélegrin, 31400, Toulouse, France

³Centre National d'Etudes Spatiales (CNES), 18 avenue Edouard Belin, 31400, Toulouse, France

⁴CESBIO, université de Toulouse, 18 avenue Edouard Belin, 31400, Toulouse, France

Résumé – Les séries temporelles d'images SAR polarimétriques Sentinel-1 (S1) sont cruciales pour analyser les changements environnementaux. La détection de la déforestation, associée à un point de changement dans une série temporelle, doit surmonter les défis liés à la saisonnalité et au caractère aléatoire de la réflectivité des environnements forestiers. Cet article propose une méthode bayésienne appliquée à des séries temporelles de données SAR, exploitant les propriétés polarimétriques partielles des mesures S1 pour une meilleure détection. Les résultats réels montrent que le modèle surpasse les modèles bivariés standards, les méthodes à polarisation unique et une méthode existante basée sur l'imagerie optique.

Abstract – Sentinel-1 (S1) time series of polarimetric SAR images are crucial for analyzing environmental changes. Accurately detecting forest loss events, associated to change points in a time series, must overcome challenges related to seasonality and the random nature of reflectivity in forest environments. This article proposes a Bayesian method applied to SAR data time series, leveraging the partial polarimetric characteristics of S1 measurements for improved detection. Real-world results show that the model outperforms standard bivariate models, single-polarization methods, and an existing method based on optical imagery.

1 Introduction

La télédétection permet de surveiller la déforestation à l'échelle globale en quasi temps réel, facilitant ainsi la mise en place d'actions de gestion et de préservation. L'accessibilité croissante à un grand nombre de données multi-temporelles a permis de développer de nombreuses approches de suivi de la déforestation. Les méthodes traditionnelles s'appuient sur des données optiques [1], mesurées par les capteurs Landsat ou Sentinel-2, mais leur efficacité est fortement limitée par la couverture nuageuse dans les régions tropicales. Pour surmonter cette contrainte, des approches basées sur l'imagerie SAR, notamment S1 en bande C, ont été développées [2, 3, 4, 5, 6]. Cependant, si ces méthodes sont généralement opérationnelles dans les forêts humides denses, où la rétrodiffusion est typiquement constante par morceaux, elles peinent dans les forêts tropicales caractérisées par de fortes variations saisonnières. De plus, ces méthodes reposent soit sur un seul canal de polarisation [2, 3, 4], soit sur une approche où les canaux VV et VH de S1 sont traités indépendamment, puis combinés par intersection [5] ou par union [6], ce qui peut conduire à de nombreuses omissions dans le premier cas et à une augmentation des fausses alarmes dans le second.

Cet article propose une approche de Détection Bayésienne de Points de Changement (DBPC) en ligne pour des séries temporelles bivariées. La méthode, inspirée du cadre bayésien modulaire pour l'estimation en ligne des changements dans les paramètres génératifs de données séquentielles [7], a été adaptée pour traiter les séries temporelles en double polarisation de S1, afin de détecter les pertes forestières dans des zones caractérisées par une végétation hétérogène (forêt, savane) et affectées par une forte saisonnalité. En considérant les données

comme des observations provenant d'un processus génératif, l'accent est mis sur la détection des instants où les paramètres de ce processus changent.

La section 2 formalise le problème de détection de changement et présente l'approche proposée. La section 3 propose différentes méthodes de type DBPC et discute du réglage de leurs paramètres. La section 4 présente les données d'entrée et de référence, ainsi que les résultats obtenus. Enfin, la section 5 propose des conclusions.

2 Analyse bayésienne conjuguée

2.1 Énoncé du problème

Soit $\mathbf{x}_{1:t} = [x_1, \dots, x_t]$ un vecteur de données. On suppose que les t éléments $\mathbf{x}_{1:t}$ peuvent être partitionnés de manière à ce que les données dans chaque partition soient des échantillons i.i.d. provenant d'une distribution spécifique. La détection bayésienne en ligne de points de changement repose sur la modélisation du temps écoulé depuis le dernier point de rupture, appelé la longueur de partition [7]. Cette longueur à l'instant t est notée r_t (figure 1). Elle peut prendre l'une des deux valeurs suivantes :

$$r_t = \begin{cases} 0 & \text{si un changement survient à l'instant } t \\ r_{t-1} + 1 & \text{sinon.} \end{cases} \quad (1)$$

L'objectif est d'estimer la distribution *a posteriori* de la longueur de partition, qui peut être exprimée comme suit :

$$p(r_t | \mathbf{x}_{1:t}) = \frac{p(r_t, \mathbf{x}_{1:t})}{\sum_{r_t=0}^t p(r_t, \mathbf{x}_{1:t})}. \quad (2)$$

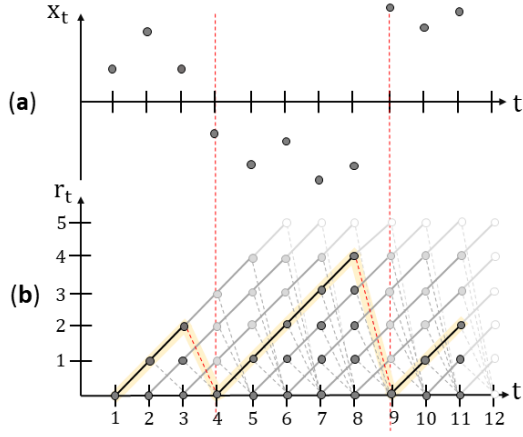


FIGURE 1 : Diagrammes conceptuels de (a) données synthétiques avec deux points de rupture, et (b) longueur de partition r_t en fonction du temps t , réinitialisée à chaque changement.

Selon les propriétés des modèles de Markov cachés [7], la distribution conjointe peut être exprimée de manière récursive comme suit :

$$p(r_t, \mathbf{x}_{1:t}) = \sum_{r_{t-1}=0}^{t-1} p(\mathbf{x}_t | \mathbf{x}_{t-1}^{(r_t)}) p(r_t | r_{t-1}) p(r_{t-1}, \mathbf{x}_{1:t-1}), \quad (3)$$

où $p(r_{t-1}, \mathbf{x}_{1:t-1})$ représente l'information transmise depuis la date précédente, $p(r_t | r_{t-1})$ est la probabilité *a priori* d'un changement qui modélise la probabilité de transition, et $p(\mathbf{x}_t | \mathbf{x}_{t-1}^{(r_t)})$ est la distribution prédictive *a posteriori* qui prédit la prochaine observation, à partir de toutes les données observées précédemment.

Cet article propose une extension du problème d'estimation au cas des données SAR bivariées.

2.2 Distribution a posteriori bivariée

Soit $\mathbf{X}_{1:t} = [\mathbf{x}_1, \dots, \mathbf{x}_t] \in \mathbb{R}^{2 \times t}$, où chaque $\mathbf{x}_i \in \mathbb{R}^2$ est un vecteur contenant l'information de réflectivité, mesurée pour les polarisations VH et VV au temps i . Dans un segment de données entre deux changements consécutifs, $p(\mathbf{x}_t | \mathbf{X}_{t-1}^{(r_t)})$ est modélisée par une approche bayésienne, supposant des échantillons i.i.d. suivant une fonction de vraisemblance $p_\theta(\mathbf{x}_t)$. Les paramètres θ sont régis par une distribution *a priori* $\pi_\eta(\theta)$, requise pour calculer la distribution *a posteriori* de (2). La distribution prédictive *a posteriori* peut alors être exprimée comme suit :

$$p(\mathbf{x}_t | \mathbf{X}_{t-1}^{(r_t)}) = \int p_\theta(\mathbf{x}_t) \pi_\eta(\theta | \mathbf{X}_{t-1}^{(r_t)}) d\theta, \quad (4)$$

où la distribution *a posteriori* des paramètres est donnée par :

$$\pi_\eta(\theta | \mathbf{X}) = \frac{p_\theta(\mathbf{X}) \pi_\eta(\theta)}{p(\mathbf{X})} = \frac{p_\theta(\mathbf{X}) \pi_\eta(\theta)}{\int p_\theta(\mathbf{X}) \pi_\eta(\theta) d\theta}. \quad (5)$$

Le calcul de la distribution *a posteriori* des paramètres peut s'avérer coûteux en raison de la sommation dans (5). Les lois *a priori* conjuguées offrent une solution en ramenant le calcul à une simple mise à jour des paramètres [8]. Une distribution *a priori* est conjuguée à une vraisemblance $p_\theta(\mathbf{x}_t)$ si la distribution *a posteriori* $\pi_\eta(\theta | \mathbf{X})$ appartient à la même famille que $\pi_\eta(\theta)$.

Comme démontré dans une étude précédente [9], les données d'intensité S1 utilisées dans cette étude suivent une distribution log-normale, et l'emploi d'intensité en échelle logarithmique conduit à : $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, avec une moyenne et une matrice de covariance inconnues. La loi *a priori* conjuguée est une distribution normale-inverse Wishart, i.e. $\theta = [\boldsymbol{\mu}, \boldsymbol{\Sigma}] \sim \mathcal{NIW}(\theta; \eta_0)$, et la distribution *a posteriori* des paramètres suit la même loi :

$$p(\boldsymbol{\mu}, \boldsymbol{\Sigma} | \mathbf{X}) = \mathcal{N}(\boldsymbol{\mu} | \boldsymbol{\mu}_n, \boldsymbol{\Sigma} / \kappa_n) \mathcal{IW}_{\nu_n}(\boldsymbol{\Sigma}, \mathbf{T}_n), \quad (6)$$

dont les paramètres sont donnés par :

$$\begin{aligned} \boldsymbol{\mu}_n &= \frac{\kappa_0 \boldsymbol{\mu}_0 + n \bar{\mathbf{x}}}{\kappa_0 + n} \\ \nu_n &= \nu_0 + n \\ \kappa_n &= \kappa_0 + n \\ \mathbf{T}_n &= \mathbf{T}_0 + \mathbf{S} + \frac{\kappa_0 n}{\kappa_0 + n} (\bar{\mathbf{x}} - \boldsymbol{\mu}_0)(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)^T, \end{aligned} \quad (7)$$

avec $\mathbf{S} = \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T$, et $\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i$. La distribution prédictive *a posteriori* possède une expression analytique et correspond à une distribution *t*-bivariée [10].

2.3 Estimation de la matrice de covariance

Sur des environnements naturels tels que les forêts, qui présentent des propriétés de symétrie polarimétrique [11], les canaux en polarisation directe, VV, et croisée, VH, des données S1 ne sont pas corrélés. Les estimées empiriques de la matrice de covariance, comme celle utilisée dans (7), ne convergent que très lentement [12] vers une forme diagonale, i.e. pour une valeur importante du nombre de vues n . La figure 2 présente une comparaison entre le module du coefficient de corrélation normalisé théorique et celui estimé pour différents nombres de vues, sur trois sites illustratifs de 300×300 pixels situés en zones forestières intactes. Cette lente convergence du biais d'estimation vers une valeur nulle peut nuire à la détection des changements en réduisant artificiellement le rang de la covariance estimée et en perturbant fortement sa structure propre. Une solution ad hoc consiste à imposer la décorrélation to-

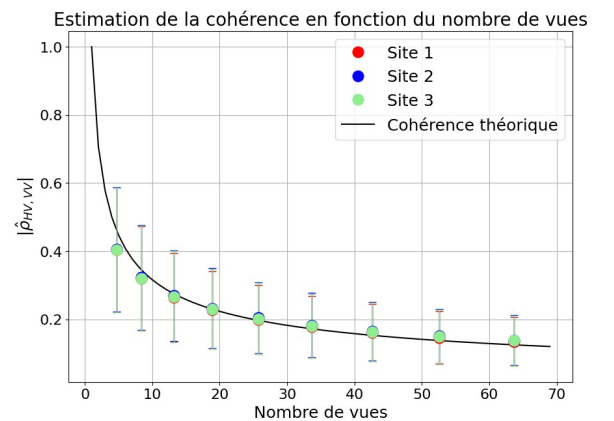


FIGURE 2 : Module du coefficient de corrélation normalisée théorique vs. estimée pour différents nombres de vues, sur trois sites de 300×300 pixels situés en zones forestières intactes.

tale en ne conservant que la partie diagonale de la matrice de covariance estimée, comme :

$$\mathbf{T}_n \rightarrow \tilde{\mathbf{T}}_n = \mathbf{T}_n \odot \mathbf{I}_M, \quad (8)$$

où $\mathbf{I}_M$ est la matrice identité de taille $M \times M$ et $\odot$ représente le produit de Hadamard élément par élément.

2.4 Alternative à la distribution t -bivariée

Dans le cas d'une vraisemblance Gaussienne et d'une distribution *a priori* normale-inverse Wishart, la distribution bivariée prédictive *a posteriori* s'écrit :

$$p(\mathbf{x}_t | \mathbf{X}_{t-1}^{(r_t)}) = \int \mathcal{N}(\mathbf{x}_t | \boldsymbol{\mu}, \boldsymbol{\Sigma}) \mathcal{N}(\boldsymbol{\mu} | \boldsymbol{\mu}_{r_t}, \boldsymbol{\Sigma} / \kappa_{r_t}) \mathcal{IW}(\boldsymbol{\Sigma} | \mathbf{T}_{r_t}, \nu_{r_t}) d\boldsymbol{\mu} d\boldsymbol{\Sigma}. \quad (9)$$

La sommation dans (9) sur l'espace des matrices positives définies fait intervenir des matrices de covariance $\boldsymbol{\Sigma}$ quelconques, même dans le cas de matrices $\mathbf{T}_{r_t}$ diagonales. Ainsi, la vraisemblance normale est évaluée avec des matrices non-diagonales, ce qui est contraire aux hypothèses physiques que l'on sait vérifiées sur les milieux mesurés. Une des conséquences de cette intégration sur un domaine inapproprié est que les statistiques *a posteriori* des différents canaux ne peuvent être factorisées, même dans le cas d'une séparabilité de leur vraisemblances, cet aspect peut être illustré à partir du processus génératif de données suivant une distribution t bivariée [13], qui s'écrit :

$$\mathbf{x}_t = \frac{\sqrt{\nu}}{s_t} \mathbf{y}_t + \boldsymbol{\mu}, \quad (10)$$

où $\mathbf{y}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$ et $s_t^2 \sim \chi_\nu^2$. Cette définition montre que toutes les composantes de $\mathbf{y}_t$ sont affectées de manière similaire par le facteur s_t , ce qui induit une dépendance mutuelle entre les dimensions de $\mathbf{x}_t$ même pour $\mathbf{C}$ diagonale.

Comme montré dans [13], diverses extensions de la distribution t standard peuvent être obtenues en ajustant le processus de génération sous-jacent pour mieux capturer les propriétés des données traitées. Dans cette étude, les données polarimétriques de S1 sont mieux représentées par des facteurs de modulation indépendants, de la manière suivante :

$$\mathbf{x}_t = \begin{bmatrix} \frac{\sqrt{\nu}}{s_{1t}} & 0 \\ 0 & \frac{\sqrt{\nu}}{s_{2t}} \end{bmatrix} \mathbf{y}_t + \boldsymbol{\mu}, \quad (11)$$

ce qui revient à réaliser la sommation de (9) sur l'espace des matrices de covariance diagonales. Dans ce cas, la distribution prédictive *a posteriori* est donnée par le produit de deux distributions t -univariées de mêmes paramètres ν_n et κ_n :

$$p(\mathbf{x} | \mathbf{X}) = t_{2\nu_n}(x_1 | \boldsymbol{\eta}_{1,n}) t_{2\nu_n}(x_2 | \boldsymbol{\eta}_{2,n}) \quad (12)$$

3 Détection et réglage des paramètres

3.1 Stratégie de détection

À l'instant t , la longueur de partition la plus probable, M_t , est donnée par :

$$M_t = \arg \max_{r_t} p(r_t | \mathbf{X}_{1:t}) \quad (13)$$

En l'absence de changement, M_t suit la relation :

$$M_t = M_{t-1} + 1, \quad (14)$$

alors qu'un changement se traduit par une chute de l'indice, $M_t < M_{t-1}$, pour atteindre une valeur qui dépend du retard de détection. En effet, le changement s'est produit au temps $t_c = t - M_t$ et n'a été détecté que M_t acquisitions plus tard. En pratique on utilise la règle suivante :

$$M_t < M_{t-1} - \Delta M, \quad (15)$$

où $\Delta M > 0$ permet d'éviter les fausses alarmes dues aux fluctuations aléatoires de la série temporelle.

3.2 Réglage des paramètres

Pour ajuster la sensibilité de détection de points de rupture, il est possible d'agir sur le seuil ΔM comme mentionné précédemment, mais aussi sur les paramètres *a priori*, ν_0 , le nombre de degrés de liberté de la distribution t , et κ_0 , sa dispersion, qui jouent un rôle important. Une valeur élevée de ν_0 ou κ_0 rend l'algorithme plus conservateur, tandis qu'une valeur plus faible augmente la réactivité aux changements. Enfin, lorsque le nombre d'observations augmente, l'influence des paramètres *a priori* initiaux diminue.

4 Résultats

4.1 Données de référence

La méthode est validée à l'aide du jeu de données MapBiomass Alerta (MBA) de l'année 2020 [14], qui regroupe des alertes de déforestation provenant de divers systèmes, validées par l'inspection d'images à haute résolution (PlanetScope, 3,7 m). Cette étude se concentre sur les déforestations de taille inférieure ou égale à un hectare dans le Cerrado, un biome tropical Brésilien à forte saisonnalité et caractérisé par une végétation hétérogène, pour lequel l'utilisation de données à diversité de polarisation peut représenter un avantage.

Les pertes forestières sont évaluées à l'échelle des polygones. Les polygones MBA de 2020 (629) identifient les vrais positifs (VP), un polygone étant perturbé si une portion significative de sa surface (T_{poly}) est affectée. Les VP sont les polygones considérés comme pertes forestières par DBPC et MBA en 2020. Pour corriger les biais dus aux zones non détectées dans MBA, les faux positifs (FP) sont estimés par une approche temporelle utilisant les polygones MBA de 2021 (196) pour identifier les FP. Un FP correspond à une perte détectée par DBPC en 2020, mais enregistrée seulement en 2021 dans MBA. Un polygone est classé perte forestière si :

$$\frac{A_{\text{détectée}}}{A_{\text{polygone}}} \geq T_{\text{poly}} \quad (16)$$

où $A_{\text{détectée}}$ est la surface détectée, A_{polygone} est la surface totale du polygone, et T_{poly} est le seuil de détection.

4.2 Données d'entrée et configuration

Cette étude utilise les images S1 A/B en bande C (IW RTC) en double polarisation (VV, VH) fournies par l'ESA, après calibration radiométrique, correction topographique et orthorectification. Aucun filtrage spatial supplémentaire n'est appliqué afin de conserver la résolution originale de 20 mètres en azimut et de 22 mètres en distance au sol. Ce choix permet de maintenir la plus fine résolution possible tout en permettant la détection de changements à petite échelle.

Pour la validation, la configuration des méthodes DBPC a été fixée comme suit : $\Delta M = 10$, $\mu_0 = [0, 0]$, $\nu_0 = 1$, $\kappa_0 = 0,01$, et $T_0 = \text{diag}(1, 1)$.

4.3 Résultats de détection de la déforestation

Les différentes versions de la méthode DBCP, univariées DBCP- vh ou DBCP- vv , bivariées sans formatage de la matrice de covariance, DBCP- $2t$, avec diagonalisation DBCP- $2t_0$ et avec factorisation DBCP- pol , sont appliquées sur des données réelles. Les résultats sont montrés dans le tableau 1 pour différentes valeurs de T_{poly} , et sont comparés aux alertes GLAD annuelles ($aGLAD$) [1], dérivées des images optiques Landsat. La figure 3 montre un exemple de détection de changement avec DBCP- pol sur une série temporelle de S1, marquée par des motifs saisonniers.

TABLE 1 : Matrice de confusion comparant les différentes méthodes pour plusieurs valeurs de T_{poly} .

Vrais Positifs (VP) [%]						
T_{poly}	DBCP- pol	DBCP- $2t_0$	DBCP- $2t$	DBCP- vh	DBCP- vv	$aGLAD$
75%	47,4	41,3	39,9	38,2	25,0	13,4
50%	76,5	70,6	70,4	67,7	55,2	31,6
30%	88,6	87,9	81,9	83,9	76,6	50,4
10%	97,3	96,7	92,8	92,7	92,7	61,7
Faux Positifs (FP) [%]						
75%	0	0	0	0	0	0
50%	0	0	0	0	0	0,5
30%	0,5	0,5	0,5	0,5	1,5	1,0
10%	3,6	3,6	3,6	3,6	6,6	5,6

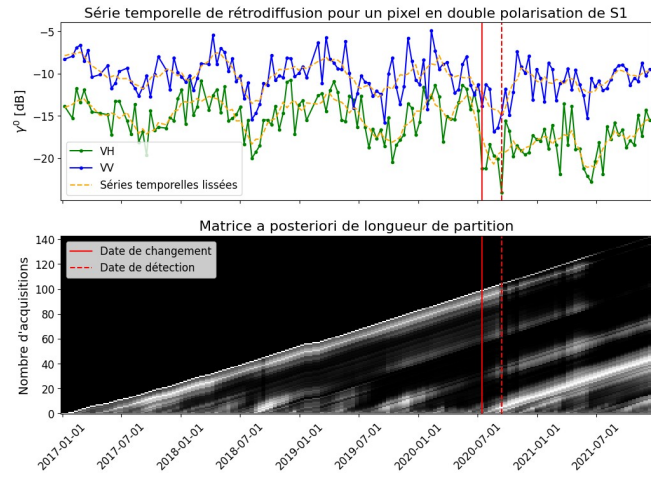


FIGURE 3 : Exemple de détection de point de rupture pour un pixel avec DBCP- pol , illustré par (**haut**) des séries temporelles de réflectivité, et (**bas**) la matrice des probabilités de longueur de partition pour chaque date.

Les résultats mettent en évidence la supériorité de DBCP- pol , qui offre le meilleur compromis entre sensibilité et précision. L'utilisation des données polarimétriques améliore la détection (+9 à 22 % de VP par rapport aux méthodes en simple polarisation), tout en préservant une robustesse élevée aux fausses alarmes. La modélisation bayésienne surpasse également $aGLAD$ (+35 à 55 % de VP), tirant parti des données SAR pour mieux gérer la saisonnalité (figure 3). La comparaison avec DBCP- $2t$ et DBCP- $2t_0$ illustre la progression vers le modèle final, confirmant l'avantage d'une approche bayésienne factorisée par rapport au modèle standard.

5 Conclusions

Cet article a proposé une méthode de détection de la déforestation basée sur une approche bayésienne appliqué aux séries temporelles en double polarisation de S1. Cette approche permet une estimation performante des points de rupture en tenant compte de la saisonnalité et en exploitant la non-corrélation entre les canaux de polarisation, offrant ainsi de meilleures performances que les modèles bivariés standards, les méthodes à polarisation unique et une méthode optique existante. Les résultats obtenus sur des données réelles valident l'efficacité du modèle. Il est prévu d'étendre cette approche aux séries temporelles multi-sources pour enrichir l'analyse des changements environnementaux.

Références

- [1] M. C. Hansen *et al.*, "Humid tropical forest disturbance alerts using landsat data," *Environmental Research Letters*, vol. 11, no. 3, p. 034008, 2016.
- [2] J. Reiche *et al.*, "Characterizing tropical forest cover loss using dense sentinel-1 data and active fire alerts," *Remote Sensing*, vol. 10, p. 777, May 2018.
- [3] A. Bouvet *et al.*, "Use of the sar shadowing effect for deforestation detection with sentinel-1 time series," *Remote Sensing*, vol. 10, no. 8, 2018.
- [4] J. Doblas *et al.*, "Optimizing near real-time detection of deforestation on tropical rainforests using sentinel-1 data," *Remote Sensing*, vol. 12, no. 23, 2020.
- [5] D. Hoekman *et al.*, "Wide-area near-real-time monitoring of tropical forest degradation and deforestation using sentinel-1," *Remote Sensing*, vol. 12, p. 3263, Oct. 2020.
- [6] B. Ygorra *et al.*, "Monitoring loss of tropical forest cover from sentinel-1 time-series : A cusum-based approach," *International Journal of Applied Earth Observation and Geoinformation*, vol. 103, p. 102532, 09 2021.
- [7] R. P. Adams and D. J. C. MacKay, "Bayesian online change-point detection," *arXiv*, 2007.
- [8] W. Feller, *An Introduction to Probability Theory and Its Applications : Volume I*. Wiley, 1968.
- [9] M. Bottani *et al.*, "Advanced bayesian method for timely small-scale forest loss detection in the brazilian amazon and cerrado with sentinel-1 time-series," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. XLVIII-3-2024, pp. 43–49, 2024.
- [10] K. P. Murphy, "Conjugate bayesian analysis of the gaussian distribution," tech. rep., Univeristy of British Columbia, 2007.
- [11] S. V. Nghiem *et al.*, "Symmetry properties in polarimetric remote sensing," *Radio Science*, vol. 27, no. 05, pp. 693–711, 1992.
- [12] I. R. Joughin, D. P. Winebrenner, and D. B. Percival, "Probability density functions for multilook polarimetric signatures," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 32, pp. 562–574, May 1994.
- [13] S. Kotz and S. Nadarajah, *Multivariate T-Distributions and Their Applications*, vol. 101. 01 2006.
- [14] RAD2023, "Radd2023 :annual deforestation report of brazil 2023," technical report, São Paulo, Brazil, 2024.