MULTI-SOURCE FUSION USING BAYESIAN ONLINE CHANGE DETECTION: APPLICATION TO DEFORESTATION MONITORING USING SAR-OPTICAL TIME SERIES

Marta Bottani [1,2,3,4], *Laurent Ferro-Famil* [2,4], *Jean-Yves Tourneret* [1,5]

[1] TéSA, Toulouse [2] ISAE Supaero, Toulouse [3] CNES, Toulouse
[4] CESBIO, University of Toulouse [5] IRIT, INP Toulouse, France

ABSTRACT

An online Bayesian changepoint detection framework is proposed to identify structural changes in multiple time series. An existing approach is extended to support asynchronous, multi-source inputs via both deterministic and probabilistic fusion strategies. The resulting framework enables timely, interpretable, and sensor-agnostic detection of forest changes, addressing key limitations of traditional offline and single-sensor methods. Experiments are conducted using both synthetic data and real Sentinel-1 SAR and Sentinel-2 optical data over tropical forests affected by deforestation. Results highlight the benefits of multi-source fusion for accurate and timely disturbance detection.

***Index Terms*—** Online Change Detection, Bayesian, Data Fusion, Time Series, Deforestation.

1. INTRODUCTION

Timely and accurate detection of forest disturbances is critical for addressing global challenges such as biodiversity loss and climate change [1]. Remote sensing imagery has become indispensable for forest monitoring, yet the limitations of individual data types often restrict their standalone effectiveness. Optical imagery offers rich spectral information that reflects vegetation properties. However, its utility is frequently hampered by cloud cover and limited canopy penetration. SAR imagery offers a complement to optical data, as it is largely unaffected by atmospheric conditions and provides insight into vegetation structure. However, monitoring forest loss using SAR data also presents challenges due to the sensitivity of SAR reflectivity to other factors, such as reduced contrast in the presence of vegetation remnants on the ground, sensitivity to moisture variations, and a forest loss signature that may fade a few weeks after the deforestation event [2, 3]. The complementary nature of optical and SAR data has led to growing interest in multi-source fusion approaches, with efforts focused on land-use classification using probabilistic fusion methods that handle sensors acquiring data asynchronously [4, 5], as well as decision-level fusion strategies combining independent classifiers [6]. However, such offline methods are unsuitable for Near Real-Time (NRT) applications like forest disturbance alerting. More recent approaches have focused on online change detection in time series, using both frequentist and Bayesian methods [7]. Some studies have investigated Bayesian changepoint estimation using mixture models [8] and hidden Markov models [9]. In the context of forest monitoring, a temporal Markov model can integrate optical and SAR observations to generate forest probability maps [10], and a probabilistic framework has been used to combine data from Sentinel-1, Landsat, and ALOS-PALSAR-2 for deforestation monitoring [11].

This work extends the Bayesian Online Changepoint Detection (BOCD) framework [12] to handle both synchronous and asynchronous, unequally sampled time series with independent statistics. The resulting method provides a sensor-agnostic solution applicable to any changepoint detection problem involving multiple asynchronous time series, and is evaluated in this work for NRT deforestation detection.

The remainder of the paper is organized as follows. Section 2 summarizes the BOCD algorithm. Section 3 extends it to synchronous time series, relevant for dual-polarization Sentinel-1 data. Section 4 addresses asynchronous multi-source fusion using deterministic and Bayesian strategies, supporting SAR-optical integration. Section 5 presents experiments on synthetic and real data. Finally, Section 6 provides concluding remarks.

2. BOCD FRAMEWORK

Let $\mathbf{x}_{1:n} = (x_1, \ldots, x_n)^T \in \mathbb{R}^n$ denote a time-ordered sequence, with x_n being the most recent observation. This sequence can be partitioned into segments of consecutive independent and identically distributed (i.i.d.) observations. The objective of BOCD [13] is to estimate, in an online manner, the number and location of changepoints, i.e., the starting points of new segments, by modeling the run length r_n, which represents the number of time steps since the last changepoint and serves as a hidden variable in a Markov process [12]. At each step, r_n either increments by one, if no change occurs, or resets to zero in case of change, as illustrated in Fig.1. Changes are detected by evaluating $p(r_n|\mathbf{x}_{1:n})$, obtained from

the following joint distribution computed recursively:

$$\begin{aligned} & p(r_n, \mathbf{x}_{1:n}) \\ &= \sum_{r_{n-1}=0}^{n-1} p(r_n, x_n | r_{n-1}, \mathbf{x}_{1:n-1})\, p(r_{n-1}, \mathbf{x}_{1:n-1}), \\ &= \sum_{r_{n-1}=0}^{n-1} p(x_n | r_{n-1:n}, \mathbf{x}_{n-1}^{(r_n)})\, p(r_n | r_{n-1})\, p(r_{n-1}, \mathbf{x}_{1:n-1}), \end{aligned} \tag{1}$$

where $r_{n-1:n} = (r_{n-1}, r_n)^T$, $\mathbf{x}_{n-1}^{(r_n)} = \mathbf{x}_{n-1:n-r_n}$ is the data from the last r_n steps, $p(x_n | r_{n-1:n}, \mathbf{x}_{n-1}^{(r_n)})$ is the data posterior predictive distribution for the run length r_n, and $p(r_n | r_{n-1})$ is the transition probability. Within a segment, i.e., between changepoints, $p(x_n | r_{n-1:n}, \mathbf{x}_{n-1}^{(r_n)})$ is modeled based on the data likelihood $p_{\boldsymbol{\theta}}(x_n)$ and parameter prior distribution $\pi_{\boldsymbol{\eta}}(\boldsymbol{\theta})$ (depending on the parameter vector $\boldsymbol{\eta}$):

$$p(x_n | r_{n-1:n}, \mathbf{x}_{n-1}^{(r_n)}) = \int p_{\boldsymbol{\theta}}(x_n)\, \pi_{\boldsymbol{\eta}}(\boldsymbol{\theta} | r_{n-1}, \mathbf{x}_{n-1}^{(r_n)})\, d\boldsymbol{\theta}, \tag{2}$$

where the posterior distribution of the parameters, is given by:

$$\pi_{\boldsymbol{\eta}}(\boldsymbol{\theta} | r_{n-1}, \mathbf{x}_{n-1}^{(r_n)}) = \frac{p_{\boldsymbol{\theta}}(\mathbf{x}_{n-1}^{(r_n)}) \pi_{\boldsymbol{\eta}(r_{n-1})}(\boldsymbol{\theta})}{\int p_{\boldsymbol{\theta}}(\mathbf{x}_{n-1}^{(r_n)}) \pi_{\boldsymbol{\eta}(r_{n-1})}(\boldsymbol{\theta})}\, d\boldsymbol{\theta}. \tag{3}$$

Calculating the parameter posterior distribution may be computationally costly. Using conjugate priors simplifies these calculations to a simple parameter update [14], yielding at step n and run length r_n the posterior distribution parameters $\boldsymbol{\eta}(r_n)$. For normal data with unknown mean and variance, the prior distribution is a normal inverse-gamma density, and the posterior predictive is a t-distribution [15]. Changepoints can then be estimated by tracking the position of the most probable run length at step n, defined as $M_n = \arg\max_{r_n} p(r_n | \mathbf{x}_{1:n})$. In the absence of a change, the Maximum A Posteriori (MAP) run length value increases, i.e., $M_n = M_{n-1} + 1$. In practice, the detection of a change may be delayed by a few time instants as M_n may not drop to zero. Thus a change is detected when:

$$M_n < M_{n-1} - \Delta M, \tag{4}$$

where $\Delta M > 0$ is a threshold limiting false detections.

3. INDEPENDENT SYNCHRONOUS TIME SERIES

Consider N_s synchronous sources providing, at each time step, a set of independent observations $\mathbf{x}_n \in \mathbb{R}^{N_s}$, which may be concatenated with past multivariate data as:

$$\mathbf{X}_{1:n} = \left[\mathbf{x}_1, \ldots, \mathbf{x}_n\right] = \left[\mathbf{x}_{1,1:n}, \ldots, \mathbf{x}_{N_s,1:n}\right]^T \in \mathbb{R}^{N_s \times n}, \tag{5}$$

where $\mathbf{x}_{i,1:n} \in \mathbb{R}^n$ is the time series from the i^{th} sensor. Assuming source independence between two consecutive changepoints, the data likelihood factorizes as:

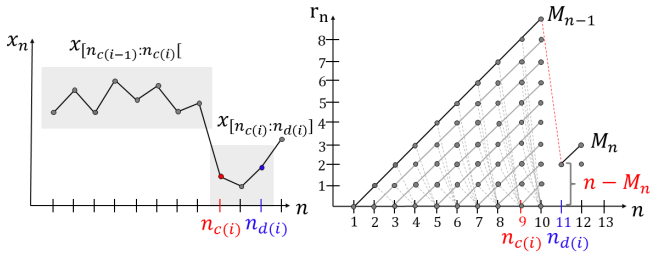

Fig. 1. Example of change detection. $n_{c(i)}$ is the i^{th} index of change and $n_{d(i)}$ is the corresponding index of detection.

$$p_{\boldsymbol{\Theta}}(\mathbf{x}_n) = \prod_{i=1}^{N_s} p_{\boldsymbol{\theta}_i}(x_{i,n}), \tag{6}$$

where $\boldsymbol{\Theta}$ is the set of parameters $\{\boldsymbol{\theta}_i\}_{i=1}^{N_s}$. For a given run length value, i.e., considering time steps lying between two consecutive changes, the multivariate data predictive posterior of N_s independently distributed source data is:

$$\begin{aligned} & p(\mathbf{x}_n | r_{n-1:n}, \mathbf{X}_{n-1}^{(r_n)}) \\ &= \prod_{i=1}^{N_s} \int p_{\boldsymbol{\theta}_i}(x_{i,n})\, \pi_{\boldsymbol{\eta}_i}(\boldsymbol{\theta_i} | r_{n-1}, \mathbf{x}_{i,n-1}^{(r_n)})\, d\boldsymbol{\theta}_i, \\ &= \prod_{i=1}^{N_s} f_i\left(\boldsymbol{\eta}_i(r_{n-1})\right), \end{aligned} \tag{7}$$

where $f_i(\boldsymbol{\eta}_i(r_{n-1}))$ is the posterior predictive of the i^{th} source and $\mathbf{X}_{n-1}^{(r_n)}$ is the multi-source data history $\{\mathbf{x}_{i,n-1}^{(r_n)}\}_{i=1}^{N_s}$. The BOCD formalism considers groups of acquisitions confined within single temporal segments which significantly reduces the complexity of multivariate posterior predictive processing to N_s times the cost of univariate approaches [15]. The BOCD for multiple independent simultaneous time series, following potentially different distributions, is implemented using (7) with the posterior predictive in (1). Changes are detected using (4), as in the univariate case.

4. INDEPENDENT ASYNCHRONOUS TIME SERIES

This section considers N_s sensors acquiring data asynchronously (i.e., at different acquisition dates). Each sensor produces a time series $\mathbf{x}_{i,1:n_i} = (x_{i,1}, \ldots, x_{i,n_i})^T \in \mathbb{R}^{n_i}$ of arbitrary length, with corresponding acquisition dates $\mathbf{t}_{i,1:n_i} = (t_{i,1}, \ldots, t_{i,n_i})^T \in \mathbb{R}^{n_i}$. A joint processing strategy is proposed for asynchronous multi-source datasets, where acquisition instants are fused into a single Markov chain modeling the run length over time. This online fusion extends the multi-source posterior predictive in (7) to asynchronous data by defining a common time axis $\mathbf{t}$, formed by

merging and sorting the acquisition dates of all sensors:

$$\mathbf{t} = \{t_{i,n}, i = 1, \ldots, N_s, n = 1, \ldots, n_i\}, \tag{8}$$

with $t_i < t_{i+1}, \forall i$. The distributions in (7) are adjusted based on the availability of data from each sensor:

$$f_i(\boldsymbol{\eta}_i(t_n)) = p(x_{i,n_i}|r_{n-1:n}, \mathbf{x}_{i,n_i-1}^{(r_n)}), \tag{9}$$

where the vector of past acquisitions by the i^{th} sensor for the global time scale run length r_n is given by:

$$\mathbf{x}_{i,n_i-1}^{(r_n)} = (x_{i,n_i-1}, \ldots, x_{i,n_i-r_{n_i}})^T, \tag{10}$$

with $r_{n_i} = \min(m)$ such that (s.t.) $t_{i,m} \geq t_{n-r_n}$. Hence, $t_{i,r_{n_i}}$ represents the time of the first acquisition from source i occurring after the hypothesized change date, t_{n-r_n}. The posterior predictive parameters are updated only when a new data sample is measured by the considered sensor for a given run length path, i.e., $\boldsymbol{\eta}_i(t_n) = \boldsymbol{\eta}_i(t_{n-1})$ if $t_{i_{n_i}} < t_{i_n}$.

4.1. Deterministic Source Fusion

Multi-source fusion can be performed by reformulating (7) with an arbitrary weight vector $\mathbf{w}(r_{n-1}) \in \mathbb{R}^{N_S}$ which allows older data to be down-weighted:

$$p_{\mathbf{w}(r_{n-1})}(\mathbf{x}_n|r_{n-1:n}, \mathbf{X}_{n-1}^{(r_n)}) = \prod_{i=1}^{N_s} f_i^{w_i(r_{n-1})}(\boldsymbol{\eta}_i(t_n)). \tag{11}$$

Sources having no available data history at step n, are either not considered in (11) or assigned $w_i = 0$. Setting $w_i = 1$ for source i and step n leads to fusion with full memory of the last acquisition from the i^{th} sensor. One option is to delay fusion and wait for the next available observation from each source [16]. An alternative avoiding latency is adaptive power weighting [17], which emulates a fading memory effect by gradually reducing the influence of outdated observations. The weight w_i evolves dynamically according to $dw_i = -\lambda w_i dt$, yielding the closed-form solution:

$$w_i(r_{n-1}) = e^{-\lambda \Delta t_{i,n}} \, \forall r_{n-1}, \quad \Delta t_{i,n} = t_n - t_{i,n_i} \geq 0, \tag{12}$$

with $\lambda \geq 0$, $t_{i,n_i} = \max_n(t_{i,n})$ s.t. $t_{i,n} \leq t_n$. Full memory is retained for source i when $\lambda = 0$, while setting $\lambda = +\infty$ results in no memory of past acquisitions. Fig. 2 illustrates an example of computation of $\Delta t_{i,n}$.

4.2. Bayesian Source Fusion

Bayesian source fusion assigns a prior distribution $p_{\boldsymbol{\eta}_w}(\mathbf{w})$ to the weights, allowing uncertainty about each source at every time step to be taken into account. The objective here is to determine optimal weight values that maximize the run length

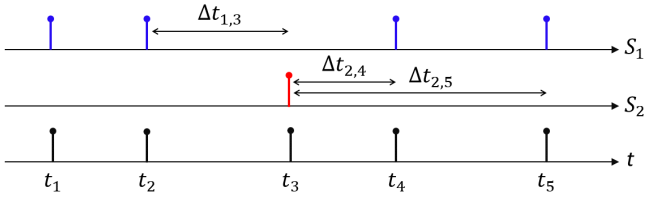

Fig. 2. Computation of $\Delta t_{i,n}$ in case of two sensors (S_1, S_2).

posterior for a given prior. The joint distribution of the run length and fusion weights is:

$$p(r_n, \mathbf{W}_n, \mathbf{X}_{1:n}) = \sum_{r_{n-1}=0}^{n-1} p(r_n|r_{n-1})\, p(r_{n-1}, \mathbf{x}_{1:n-1})$$
$$p_{\boldsymbol{\eta}_w}(\mathbf{w}_n(r_{n-1}))\, p_{\mathbf{w}(r_{n-1})}(\mathbf{x}_n|r_{n-1:n}, \mathbf{X}_{n-1}^{(r_n)}), \tag{13}$$

where $\mathbf{W}_n = (\mathbf{w}_n(0), \ldots, \mathbf{w}_n(n-1)) \in \mathbb{R}^{N_s \times n}$ denotes the set of weight vectors corresponding to all run length values. The MAP estimate of the run length is computed as $M_n = \arg\max_{r_n} p_{\mathbf{W}_{opt_n}}(r_n, \mathbf{X}_{1:n})$, whereas the weight optimization is conducted separately for each n [17, 18] and maximizes each component of the sum in (13). Substituting (11) into (13) and maximizing the resulting expression leads to the following problem:

$$\mathbf{w}^{\text{opt}} = \arg\max_{\mathbf{w}} \left\{ \log p_{\boldsymbol{\eta}_w}(\mathbf{w}) - \sum_{i=1}^{N_s} w_i l_i \right\}, \tag{14}$$

with $l_i = -\log f_i^{w_{i,n}}(\boldsymbol{\eta}_i(t_n))$. This study uses a beta prior, $w_i \sim \text{Beta}(\alpha_i, \beta_i)$ with $\alpha_i, \beta_i > 1$, which restricts the prior to concave beta distributions with a single maximum [18, 19]. The joint prior distribution of the weights is:

$$p_{\boldsymbol{\eta}_w}(\mathbf{w}) \propto \prod_{i=1}^{N_s} w_i^{\alpha_i - 1}(1 - w_i)^{\beta_i - 1}. \tag{15}$$

The optimal weight solution of (14) is given by the roots of a second-order polynomial equation [19], i.e.:

$$w_i^{\text{opt}} = \frac{\nu_i + l_i - 2 - \sqrt{(\nu_i + l_i - 2)^2 - 4(\mu_i \nu_i - 1) l_i}}{2 l_i}. \tag{16}$$

To fix the parameters α_i and β_i, the prior mean $\mu_i = \alpha_i / \nu_i$ is set to the deterministic expression in (12), while the concentration $\nu_i = \alpha_i + \beta_i$ is treated as a free parameter:

$$\mu_i = e^{-\lambda_i \Delta t_{i_n}}, \nu_i > f_{\nu_i} \nu_{m_i}, \tag{17}$$

with $\nu_{m_i} = \max\left(\frac{1}{\mu_i}, \frac{1}{1-\mu_i}\right)$ the minimal concentration value for which $\alpha_i, \beta_i > 1$, and $f_{\nu_i} \geq 1$ a factor controlling the weight variability around the prior mean. The weight variance is maximal for $f_{\nu_i} = 1$, whereas $f_{\nu_i} \to +\infty$ enforces the deterministic weight setting.

5. RESULTS

5.1. Application to Real Data

The proposed fusion strategy is evaluated on real Sentinel-1 (S1) SAR and Sentinel-2 (S2) optical time series capturing a deforestation event. Specifically, the normalized backscatter in cross-polarization is used from S1, and the Normalized Difference Vegetation Index (NDVI) is derived from S2 [20]. Four representative cases are selected and shown in Fig. 3.

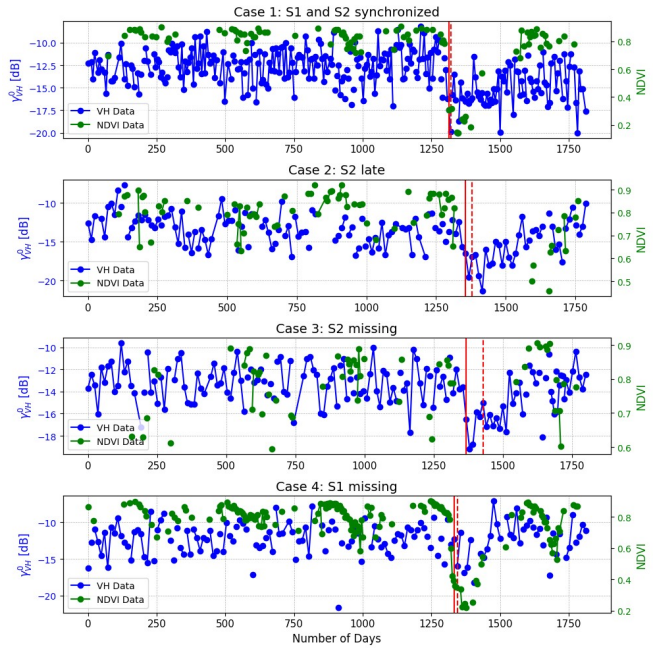

Fig. 3. Time series for four real cases. Solid red lines: true changepoints; dashed red lines: detected changepoints.

Fig.4 shows detection delays for the deterministic and Bayesian fusion methods applied to the cases in Fig.3, across varying $\tau = 1/\lambda$ values (λ in (12) and (17)). The results confirm that the weighted fusion performs well in the deterministic case and that the introduction of stochastic weights allows extending the effective range of the time constant parameter.

5.2. Assessment of Model Parameters

Synthetic data are used to test the fusion method for different model parameters. Reference trajectories are obtained by smoothing the original time series from Case 2, and Gaussian noise is added ($\sigma_{S2} = 0.04$, $\sigma_{S1} = 2.2$ dB) to simulate realistic variability. A dataset of 100 noisy time series is generated. For each time series, the fusion algorithm is run using 7 values of λ, and for each λ, 3 weighting factors f_{ν_i} are tested. The mean detection delay and detection rate from the simulations are shown in Fig. 5. The results demonstrate that the best performance is achieved with the Bayesian weighted fusion

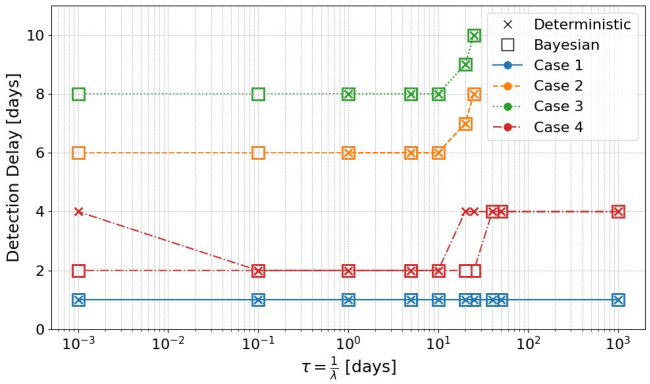

Fig. 4. Real data: detection delays for varying λ values.

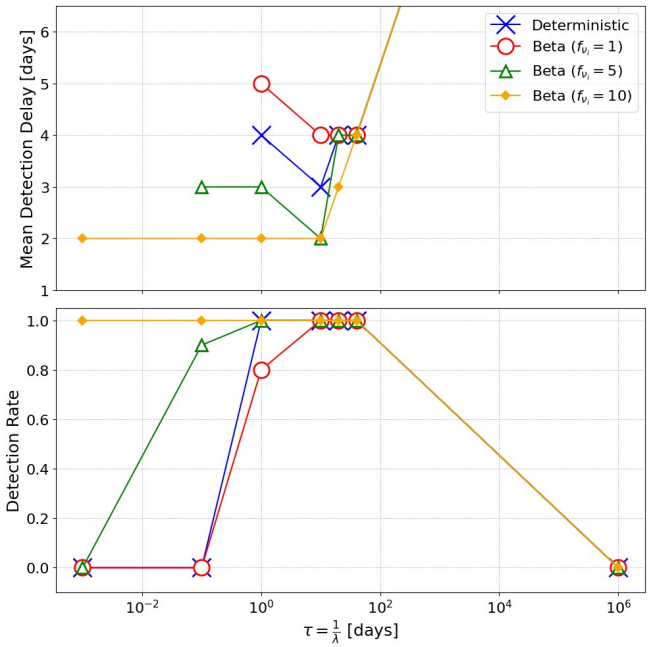

Fig. 5. Synthetic data: (**top**) mean detection delay and (**bottom**) detection rate. No data implies no detection.

strategy using $f_{\nu_i} = 10$, suggesting that optimal results occur at an intermediate setting between a purely deterministic approach ($f_{\nu_i} \to \infty$) and maximum randomness ($f_{\nu_i} = 1$).

6. CONCLUSION

This work studies a Bayesian online change detection framework for independent synchronous and asynchronous multi-source fusion, applied to NRT deforestation detection. Experiments on synthetic and real S1/S2 data show that deterministic weighted fusion performs well, while stochastic beta-weighted fusion broadens the method's operability. Future work will explore alternative Bayesian priors and the integration of additional sensor modalities.

7. REFERENCES

[1] M. Finer, S. Novoa, M. J. Weisse, R. Petersen, J. Mascaro, T. Souto, F. Stearns, and R. G. Martinez, "Combating deforestation: From satellite to intervention," *Science*, vol. 360, pp. 1303–1305, 2018.

[2] M. Bottani, L. Ferro-Famil, J. Doblas, S. Mermoz, A. Bouvet, and T. Koleck, "A statistical method for near real-time deforestation monitoring using time series of sentinel-1 images," in *Proc. Int. Geoscience and Remote Sensing Symp. (IGARSS)*, Athens, Greece, 2024, pp. 11457–11460.

[3] M. Bottani, L. Ferro-Famil, J. Doblas, S. Mermoz, A. Bouvet, and T. Koleck, "Advanced bayesian method for timely small-scale forest loss detection in the brazilian amazon and cerrado with sentinel-1 time-series," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLVIII-3-2024, pp. 43–49, 2024.

[4] A. H. S. Solberg, A. K. Jain, and T. Taxt, "Multisource classification of remotely sensed data: fusion of landsat tm and sar images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 32, no. 4, pp. 768–778, 1994.

[5] L. Bruzzone, D. F. Prieto, and S. B. Serpico, "A neural-statistical approach to multitemporal and multisource remote-sensing image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 3, pp. 1350–1359, 1999.

[6] B. Waske and J. A. Benediktsson, "Fusion of support vector machines for classification of multisensor data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 12, pp. 3858–3866, 2007.

[7] G. J. J. van den Burg and C. K. I. Williams, "An evaluation of change point detection algorithms," *arXiv*, 2022.

[8] T. Çelik, "Bayesian change detection based on spatial sampling and Gaussian mixture model," *Pattern Recognition Letters*, vol. 32, no. 12, pp. 1635–1642, 2011.

[9] E. Epaillard and N. Bouguila, "Proportional data modeling with hidden Markov models based on generalized Dirichlet and beta-Liouville mixtures applied to anomaly detection in public areas," *Pattern Recognition*, vol. 55, pp. 125–136, 2016.

[10] E. A. Lehmann, P. A. Caccetta, Z. S. Zhou, S. J. McNeill, X. Wu, and A. L. Mitchell, "Joint processing of Landsat and ALOS-PALSAR data for forest mapping and monitoring," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 1, pp. 55–67, 2012.

[11] J. Reiche, E. Hamunyela, J. Verbesselt, D. Hoekman, and M. Herold, "Improving near-real time deforestation monitoring in tropical dry forests by combining dense sentinel-1 time series with landsat and ALOS-2 PALSAR-2," *Remote Sensing of Environment*, vol. 204, no. 5, pp. 147–161, 2018.

[12] R. P. Adams and D. J. C. MacKay, "Bayesian online changepoint detection," *arXiv*, 2007.

[13] P. Fearnhead and Z. Liu, "On-line inference for multiple changepoint problems," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 69, no. 4, pp. 589–605, 2007.

[14] W. Feller, *An Introduction to Probability Theory and Its Applications: Volume I*, Wiley, 1968.

[15] K. P. Murphy, "Conjugate bayesian analysis of the gaussian distribution," Technical report, The Univeristy of British Columbia, 2007.

[16] F. Caron, M. Davy, E. Duflos, and P. Vanheeghe, "Particle filtering for multisensor data fusion with switching observation models: Application to land vehicle positioning," *IEEE Transactions on Signal Processing*, vol. 55, no. 6, pp. 2703–2719, 2007.

[17] D. McCarthy and S. T. Jensen, "Power-weighted densities for time series data," *The Annals of Applied Statistics*, vol. 10, no. 1, pp. 305–334, 2016.

[18] Y. Wang, A. Kucukelbir, and D. M. Blei, "Robust probabilistic modeling with Bayesian data reweighting," in *Proc. Int. Conf. Machine Learning (ICML)*, Sydney, Australia, 2017, vol. 70, pp. 3646–3655.

[19] L. Romero-Medrano and A. Artès-Rodrìguez, "Multisource change-point detection over local observation models," *Pattern Recognition*, vol. 134, pp. 109116, 2023.

[20] C. J. Tucker, "Red and photographic infrared linear combinations for monitoring vegetation," *Remote Sensing of the Environment*, vol. 8, pp. 127–150, 1979.