



HAL
open science

Recursive classification of satellite imaging time-series: An application to land cover mapping

Helena Calatrava, Bhavya Duvvuri, Haoqing Li, Ricardo Augusto Borsoi,
Edward Beighley, Deniz Erdoğan, Pau Closas, Tales Imbiriba

► **To cite this version:**

Helena Calatrava, Bhavya Duvvuri, Haoqing Li, Ricardo Augusto Borsoi, Edward Beighley, et al. Recursive classification of satellite imaging time-series: An application to land cover mapping. ISPRS Journal of Photogrammetry and Remote Sensing, 2024, 218, pp.447-465. 10.1016/j.isprsjprs.2024.09.003 . hal-04728259

HAL Id: hal-04728259

<https://hal.science/hal-04728259v1>

Submitted on 9 Oct 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

Recursive classification of satellite imaging time-series: An application to land cover mapping

Helena Calatrava^{a,1,*}, Bhavya Duvvuri^{a,1}, Haoqing Li^a, Ricardo Borsoi^b, Edward Beighley^a, Deniz Erdoğan^a, Pau Closas^a, Tales Imbiriba^a

^aNortheastern University, Boston, 02215, MA, USA

^bCRAN, University of Lorraine, CNRS, Vandoeuvre-les-Nancy, F-54000, France

Abstract

Despite the extensive body of literature focused on remote sensing applications for land cover mapping and the availability of high-resolution satellite imagery, methods for continuously updating classification maps in real-time remain limited, especially when training data is scarce. This paper introduces the recursive Bayesian classifier (RBC), which converts any instantaneous classifier into a robust online method through a probabilistic framework that is resilient to non-informative image variations. Three experiments are conducted using Sentinel-2 data: water mapping of the Oroville Dam in California and the Charles River basin in Massachusetts, and deforestation detection in the Amazon. RBC is applied to a Gaussian mixture model (GMM), logistic regression (LR), and our proposed spectral index classifier (SIC). Results show that RBC significantly enhances classifier robustness in multitemporal settings under challenging conditions, such as cloud cover and cyanobacterial blooms. Specifically, balanced classification accuracy improves by up to 26.95% for SIC, 12.4% for GMM, and 13.81% for LR in water mapping, and by 15.25%, 14.17%, and 14.7% in deforestation detection. Moreover, without additional training data, RBC improves the performance of the state-of-the-art DeepWaterMap and WatNet algorithms by up to 9.62% and 11.03%. These benefits are provided by RBC while requiring minimal supervision and maintaining a low computational cost that remains constant for each time step regardless of the time-series length.

Keywords: Bayesian Inference, Water Mapping, Land Cover Mapping, Deforestation Detection, Spectral Indices, Time-Series Analysis

PACS: 0000, 1111

2000 MSC: 0000, 1111

1. Introduction

1.1. Background

For the purpose of allaying increasing concerns on global environmental changes and sustainability, and thanks to the vast amount of high resolution remotely sensed data available today, there exists a considerable body of work focused on remote sensing applications involving land cover mapping and change detection (Mashala et al., 2023; Wang et al., 2023). Some examples are studies on land conservation, sustainable development, and the management of resources such as water. Changes in water dynamics can be studied by surface water mapping to monitor floods (Yaseen, 2024), describe water quality (Wasehun et al., 2024), and for coastline extraction (Sun et al., 2023). Land cover mapping also plays a crucial role in identifying the distribution of different crop types (Zhang et al., 2024) and understanding the dynamic evolution of land use in urban environments (Yu and Fang, 2023). Furthermore, many studies focus on deforestation detection in areas such as the Amazon rainforest, a region with unparalleled biodiversity and a crucial role in global climate regulation, yet facing alarming deforestation rates. Numerous works address the environmental impact that this ongoing crisis has on climate change and public health (Lapola et al., 2023; Ellwanger et al., 2020). Of particular concern is the 129% increase in deforestation within Indigenous territories since 2013 (Silva-Junior et al., 2023). Deforestation detection conducted by human experts through visual inspection is time-consuming and costly due to the vast geographic areas involved, making automated detection methods necessary (Karaman et al., 2023; Martinez et al., 2024).

*Corresponding author.

Email address: calatrava.h@northeastern.edu

¹Indicates shared first authorship.

Several sources of remotely sensed data are currently available, presenting different characteristics when it comes to spatial, spectral, radiometric and temporal resolution (Chuvieco, 2020). Spatial resolution varies from centimeters, as with high-resolution sensors on GeoEye and QuickBird-2 satellites, to a few meters, as with Landsat 9 and Sentinel-2 A/B sensors. These satellites can acquire weekly images of the same scene. Conversely, satellites with MODIS and VIIRS sensors offer daily image acquisitions but at a lower spatial resolution of hundreds of meters. Modern commercial satellite systems push the boundaries of Earth observation (Miura et al., 2023). An example is the Pléiades Neo by Airbus, with a twice-daily revisit capacity and a ground sampling distance of up to 30 cm (Chouteau et al., 2022). Notably, hundreds of CubeSats with subweekly or even daily temporal resolution and medium to high spatial resolution (Zhu et al., 2022), such as the ones providing Planetscope data (Huang and Roy, 2021). Also, very high resolution optical imagery is available with the constellations of SkySat, BlackSky and Nu-Sat micro-satellites. The SkySat satellites can provide ground sampling distances of up to 50 cm and a sub-daily revisit time (6-7 times per day) when considering the whole constellation (Jacobsen, 2022). Additionally, multimodal image fusion techniques are used to generate high spatio-temporal image sequences, contributing to generating a wealth of remotely sensed data (Li et al., 2022; Karmakar et al., 2023).

Spectral indices are one of the main land cover mapping tools given their simplicity and required low computational cost (Tran et al., 2022). They compute scalar-valued features as a function of specific spectral bands, whose value can be used to distinguish between different land cover classes contained in a pixel. Although they show a limited performance when compared to other techniques such as deep learning methods, they are widely used in remote sensing applications given their unsupervised nature (Khalid et al., 2021). They can be considered to be unsupervised because their output depends on the ratio between a combination of spectral bands, which does not require any training. The decision threshold, however, must be selected and this can be challenging when no reference data is available. Another advantage is that spectral index values are easy to interpret because they minimize the effect of illumination in satellite imagery while enhancing different spectral features present in the scene under study. For instance, the normalized difference vegetation index (NDVI) enhances the presence of trees, bushes, and others (Huang et al., 2021; Xu et al., 2022). This is due to the reflectance given by the spectral response of vegetation decreasing in the red and increasing in the infrared wavelengths. On the other hand, water indices are used for water extraction at pixel level, given the difference in spectral reflectance of land and water in the near and middle infrared wavelengths (Liu et al., 2023). The most widely used water indices are the normalized difference water index (NDWI), the modified NDWI (MNDWI) and the automated water extraction index (AWEI) (Acharya et al., 2018). It has been shown that challenging weather conditions may disrupt the extraction of water bodies with these indices (Yang et al., 2022). This can be solved by using modified methods like the one proposed by Gao et al. (2016), which effectively mitigates mountain shadows and allows the extraction of particularly challenging small water bodies. Khalid et al. (2021) also suggest that the land surface temperature based water extraction index (LBWEI) provides high accuracy under a wide variety of weather conditions.

Aside from spectral index methods, there is a wide choice of land cover classification approaches based on machine learning available in the literature, whose main advantage is an increased flexibility (Wang et al., 2022). The taxonomy of image classification techniques in remote sensing proposed by Satir and Berberoglu (2012) groups them into supervised/unsupervised, parametric/non-parametric and hard/soft classifiers, among others. Some of the explored machine learning methods used in remote sensing applications include maximum likelihood classifiers (Zeb et al., 2019), support vector machines (Kok et al., 2021), logistic regression (LR) (Li et al., 2024), random forests (Pelletier et al., 2017; Tariq et al., 2023), naive Bayes (Bai et al., 2023) and clustering methods like the widely used K-means algorithm (Ghezlbash et al., 2023). Likó et al. (2024) showed an improvement in maximum likelihood classification maps of Himalayan regions through post-classification correction measures, including ancillary data, digital elevation models, and spectral vegetation indices. Various studies assess the need for a comprehensive comparison between widely used machine learning algorithms for land cover classification (Adugna et al., 2022; Zhang et al., 2023). Ko Oo et al. (2022) propose random forest as the best classification model in mining districts when compared to maximum likelihood, support vector machines, and classification and regression trees. Furthermore, the authors in Qiu et al. (2019) suggest that deep learning methods like artificial neural networks provide high accuracy results in land cover classification, even when compared to other machine learning classifiers such as support vector machines.

Despite their widespread use in land cover mapping, the previously mentioned techniques suffer from several limitations. First, they are highly sensitive to illumination and atmospheric interferences (e.g., different aerosol concentrations or viewing angles), which can significantly impact the spectra of pixels

from a given material class (Borsoi et al., 2021b). The lack of robustness to such non-informative spectral variations is a significant limitation of, for instance, spectral indices such as the MNDWI (Yang et al., 2018). Moreover, due to the high sensor-to-target distances involved in remote sensing applications, many image pixels do not belong to a single class, but are instead composed of a mixture of different material classes (Cavalli, 2023). Although this can be addressed by spectral mixture analysis techniques (Cai et al., 2024) or by assigning a pixel to more than one class with sub-pixel mapping (Wang et al., 2020; Hu et al., 2021), it poses a significant challenge to traditional classification algorithms. As a consequence, an apparent need for robustness to outliers and spurious artifacts in remotely sensed data arises.

1.2. Multitemporal time-series classification

The analysis of multitemporal or time-series data is of increasing interest for remote sensing applications (Miller et al., 2024). Exploiting multitemporal data makes it possible to improve the performance of tasks such as classification (Deng et al., 2019; Wei et al., 2023) and spectral mixture analysis (Borsoi et al., 2021a) given the temporal correlation, while at the same time supplying the end-user with a more complete product that shows the spatial as well as the temporal distribution of land classes or their proportions. The simplest approach to perform multitemporal land cover mapping is to apply an instantaneous classifier to each image in the sequence, being spectral indices such as the NDVI a popular choice (Sun et al., 2018). However, this does not exploit the temporal information available in the data. Significant effort has been dedicated to developing techniques specifically suited to process multitemporal image sequences. For instance, Kenduywo et al. (2017) and Shi et al. (2022) proposed classification methods based on conditional random field models, which represent the interactions between class labels in both time and space. Transfer and active learning have been combined to adapt a pre-trained classifier to new images acquired at other time instants (Cao et al., 2020). Time-series classification accounting for missing pixels using Gaussian process regression was proposed by Constantin et al. (2022), while other works considered 1D temporal convolutional neural networks (CNNs) (Pelletier et al., 2019), and 3D spatio-temporal CNNs (Ji et al., 2018).

These techniques, known as *batch* or *offline* time-series classification methods, require a complete image time-series to generate classification maps. However, with satellites like Sentinel-2 and Landsat 9 continuously acquiring images, reprocessing the entire series each time a new image is acquired becomes computationally expensive. *Recursive* methods, also known as *online* methods, iteratively update multitemporal classification maps with new images by leveraging previously computed results, making them ideal for studies involving ongoing data collection. Bayesian recursion is widely used in target tracking applications, where an efficient filter (Imbiriba and Closas, 2020) must be designed to recursively obtain target state estimates from a state-space model (Ji et al., 2022). Filtering methods can also be applied to online parameter learning (Wu et al., 2019; Borsoi et al., 2020; Demirkaya et al., 2021; Imbiriba et al., 2022), which is of special interest in machine learning tasks involving the processing of time-series data (Campbell et al., 2021), such as video prediction (Franceschi et al., 2020) and speech enhancement (Martín-Doñas et al., 2020). Bayesian recursion has also been applied in pattern recognition for bioengineering applications, such as tracking retinal vasculature by estimating vessel geometry parameters (Uslu and Bharath, 2019).

The earliest recursive remote sensing classification techniques were based on Bayesian filtering ideas, by recursively updating the probabilities of each class given the measurements after each datum is acquired (Swain, 1978; Strahler, 1980). These techniques are based on a statistical model that represents the pixel spectra given its class, called a *generative model*, which is non-trivial to obtain. More recent Bayesian approaches have proposed classification strategies that are recursive both in time and across multiple spatial scales (multiresolution) (Hedhli et al., 2021), using computationally expensive algorithms such as expectation maximization to learn the model parameters and a generative model for the pixels. Other recent works have leveraged deep learning strategies, in particular different instantiations of recurrent neural networks, such as long short-term memory (LSTM) networks (Zhou et al., 2023; Wang et al., 2024). These have been applied to predict flood susceptibility and for crop identification, among others (Rußwurm and Korner, 2017; Fang et al., 2021). These methods enable learning intrinsic spatial and temporal dependencies of remotely sensed data, along with spectral patterns in specific classes over time, all with minimal supervision. The main disadvantage of other recursive algorithms is that they require large amounts of training data and long training times compared to the framework proposed in this paper. This is the case for the algorithm in Sharma et al. (2018), which uses patch-based recurrent neural networks, and the one in Pastorino et al. (2024), which combines fully convolutional networks with hierarchical probabilistic graphical models and decision tree ensembles. Mountrakis and Heydari (2023) show that deep learning methods provide substantial classification improvements over the commonly im-

plemented random forest. Nevertheless, demonstrative numbers on simulation times suggest that these methods require up to four times larger running times.

The lack of open-source labeled Sentinel-2 data for time-series analysis (i.e., containing the dynamical evolution of the true class labels) poses an important challenge to the evaluation of the methods. Although there are available multitemporal satellite imagery datasets, most of them contain pixel-wise annotations that are unique for the whole time-series. That is, no changes over time exist or are properly mapped to labels, and therefore, it is not possible to evaluate the algorithm at different time instants. One example of this is the benchmark dataset for multi-temporal and multi-modal land use land cover mapping MultiSenGE (Wenger et al., 2022). To address this, we generated our own ground truth data, as detailed in the next subsection. For the Amazon deforestation experiment, ground truth labels were available in the MultiEarth challenge dataset (Cha et al., 2023).

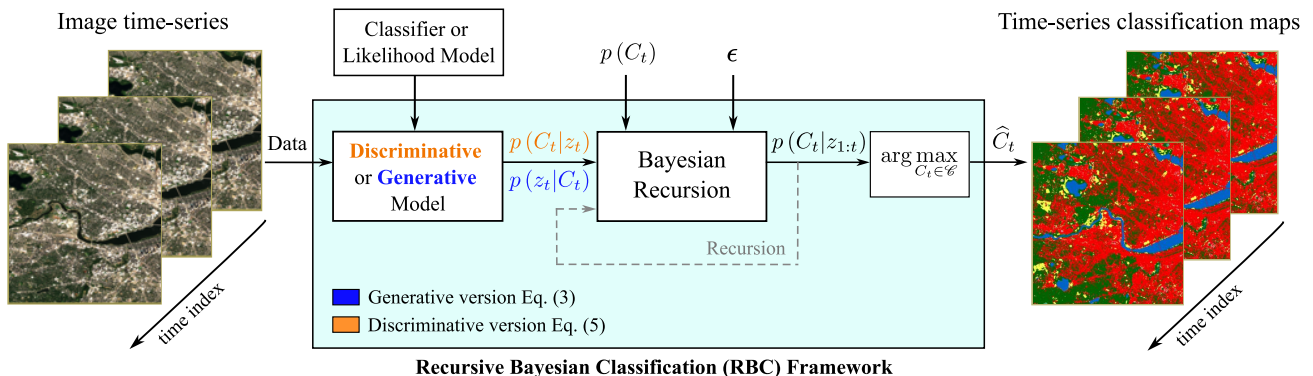


Figure 1: Overview of the proposed recursive Bayesian classification framework. Our method allows to convert an instantaneous generative model (i.e., modeling the observation of a pixel given its class label with the likelihood function $p(z_t|C_t)$) or a discriminative model (i.e., modeling the observation of a class label given a corresponding pixel with $p(C_t|z_t)$) into a recursive Bayesian classifier that exploits the temporal relationship of time-series data. Knowledge about the class prior probabilities $p(C_t)$ and the transition probability hyperparameter ϵ is assumed. The hat operator in \hat{C}_t denotes the decision from the classifier. \mathcal{C} is an experiment-dependent set containing K different labels.

1.3. Contributions

This paper aims to develop a recursive classification framework that improves the decision-making process in multitemporal and multispectral land cover classification algorithms by leveraging previous classification results. An overview of the proposed framework may be found in Fig. 1. The key contributions of this paper are as follows:

- We propose the recursive Bayesian classification (RBC) framework, which converts any instantaneous classifier into a robust online method using a probabilistic framework that enables the integration of temporal information. RBC addresses the trade-off between adaptability to natural changes in the scene and robustness to outliers caused by illumination variations and spurious artifacts in remotely sensed data. This is achieved without the need for additional training data, as opposed to more complex deep learning models.
- In Section 3.2, we introduce the spectral index classification (SIC) algorithm, which uses standard broadband spectral indices to generate predictive probabilities. This approach allows for the incorporation of class prediction uncertainty into the RBC framework.
- To overcome the lack of available ground truth data for time-series analysis, we created our own ground truth data with manually generated labels. This enabled a rigorous quantitative assessment of the proposed framework (see Section 3.3.3). The ground truth dataset, along with the pre-processed satellite imagery, has been made publicly available at (Calatrava et al., 2024).

The performance of the proposed approach is demonstrated using Sentinel-2 images in three different experiments: water mapping of a reservoir and downstream river in Oroville, California, USA; water mapping of the Charles river basin in Boston, Massachusetts, USA; and deforestation detection in the Amazon rainforest. RBC may be applied atop a generative model (see Section 3.1.1) or atop discriminative models (see Section 3.1.2). We evaluate the proposed RBC framework against three instantaneous classifiers: a Gaussian mixture model (GMM), an LR classifier, and the SIC algorithm, along with their

recursive counterparts. For the water mapping experiments, we include two state-of-the-art deep learning classifiers: the DeepWaterMap (DWM) (Isikdogan et al., 2017) and WatNet (WN) (Luo et al., 2021) algorithms.

The remainder of this paper is structured as follows. Section 2 describes the satellite data and the areas under study. Section 3 introduces the proposed RBC framework and the SIC algorithm, followed by details on the experimental setup. Section 4 presents the results, while their implications and impact are discussed in Section 5. Finally, Section 6 provides the concluding remarks. The successful extension of the RBC framework to a three-class classification task for the Charles River area is detailed in Appendix A.

2. Area of study and satellite data

Our research focuses on two areas of study located in the US and one area located in the Amazon rainforest, as we believe it is essential to test our methodology across varied geographical locations to achieve a more comprehensive performance assessment. In this section, we describe the three selected areas of study and the challenges they pose to land cover mapping algorithms.

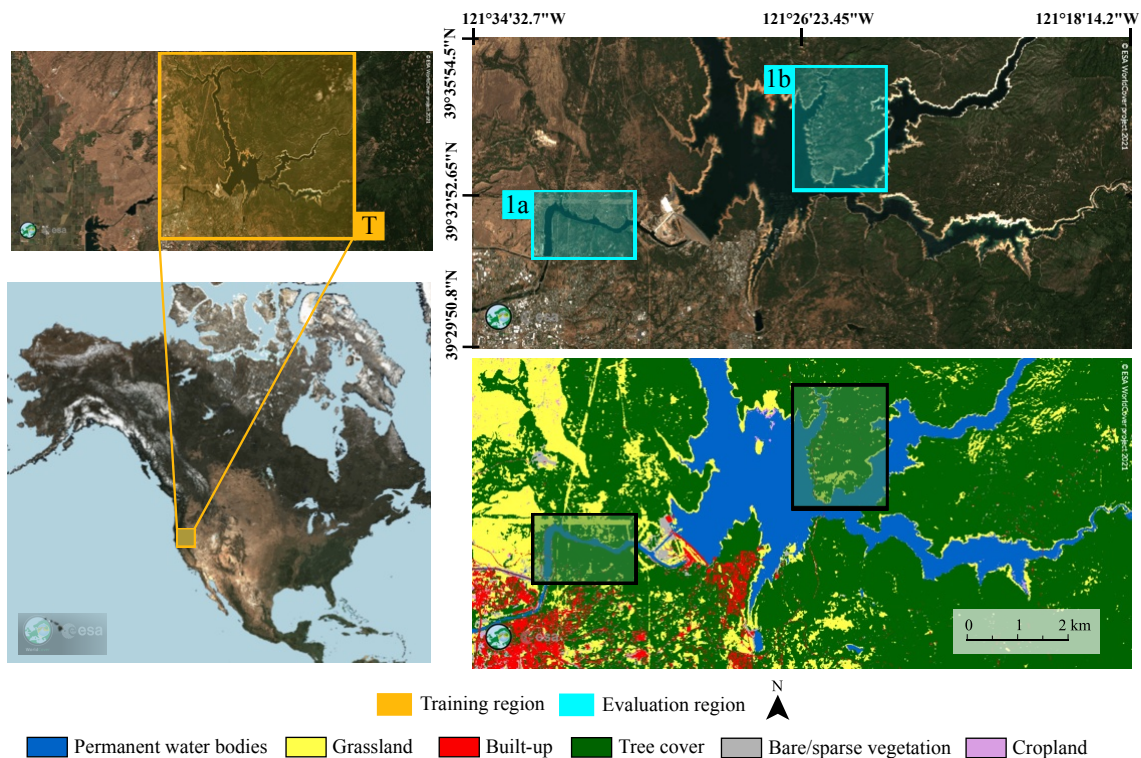


Figure 2: Sentinel-2 RGB composite image with LAT/LON information and classification map from the ESA WorldCover Map tool (Zanaga et al., 2021) of test sites 1a and 1b (Oroville Dam, California, USA).

We consider images at the blue, green, red, near-infrared (NIR), narrow NIR and shortwave infrared (SWIR) bands, with resolution and central wavelength listed in Table 1. Band 4 (red) is useful for identifying soil, water and many urban features, band 3 (green) gives excellent contrast between clear and turbid waters, and band 2 (blue) is useful for identifying vegetation and also human-made features (Maciej Huk, 2020). The SWIR bands are useful for measuring vegetation, water and soil moisture.

Table 1: Resolution and central wavelength of the processed Sentinel-2 spectral bands.

Band	Description	Resolution (m)	Central wavelength (nm)
2	Blue	10	490
3	Green	10	560
4	Red	10	665
8	NIR	10	842
8A	Narrow NIR Edge	20	865
11	SWIR 1	20	1610

2.1. Test site 1: Oroville Dam

The first test site is located in Oroville Dam, an embankment dam on the east side of the city of Oroville, in the state of California (see Fig. 2). Being 235 meters high, it is the tallest dam in the US. The area of study has geographic center coordinates of LAT/LON: 39.61, -121.43. Test sites 1a and 1b belong to the dam downstream and upstream, respectively. The water mapping of areas with geographic features like this reservoir is imperative to study, as they are essential for flood control, management and sustainability of water resources.

Illumination variations in images from test site 1a can be observed in Fig. 4. These may be caused by fluctuations in the solar incidence angle and also by differences in image acquisition times. Ripples and other artifacts in test site 1a are caused by the high flows in the river stream, making the classification of water pixels challenging and increasing the probability of false negatives. As illustrated in Fig. 4 for test site 1b, and based on reservoir storage data obtained from the NWIS USGS website ², the water level in the reservoir changes abruptly with the season. In October 2020, the recorded water storage was of 200,485.8 hc-m, decreasing to 160,783.2 hc-m in December 2020. Subsequent changes resulted in a recorded water storage of 183,223.8 hc-m and 97,542.6 hc-m in May and September of 2021, respectively. These phenomena demand the flexibility of the proposed recursive classification framework, which ensures its ability to adapt to changes in the scene. However, the very high flexibility of the method can put its robustness at risk.

Sentinel-2 Level-2A images are downloaded using the Google Earth Engine platform from the COPERNICUS/S2.SR collection, with dates between 2020-09-01 and 2021-09-26. This platform atmospherically corrects the images using the standard SEN2COR software package and indicates a cloud cover percentage. Only images with at most 10% cloud cover are downloaded. The downloaded images have dimensions of 2229 × 3341 pixels. For evaluation purposes, images were cropped to sizes of 200 × 500 pixels for test site 1a and 150 × 110 pixels for test site 1b. To ensure consistency, images from bands 8A and 11, initially at a resolution of 20 meters, were resampled to 10 meters using nearest-neighbor interpolation. A total of 45 images remained available for further processing.

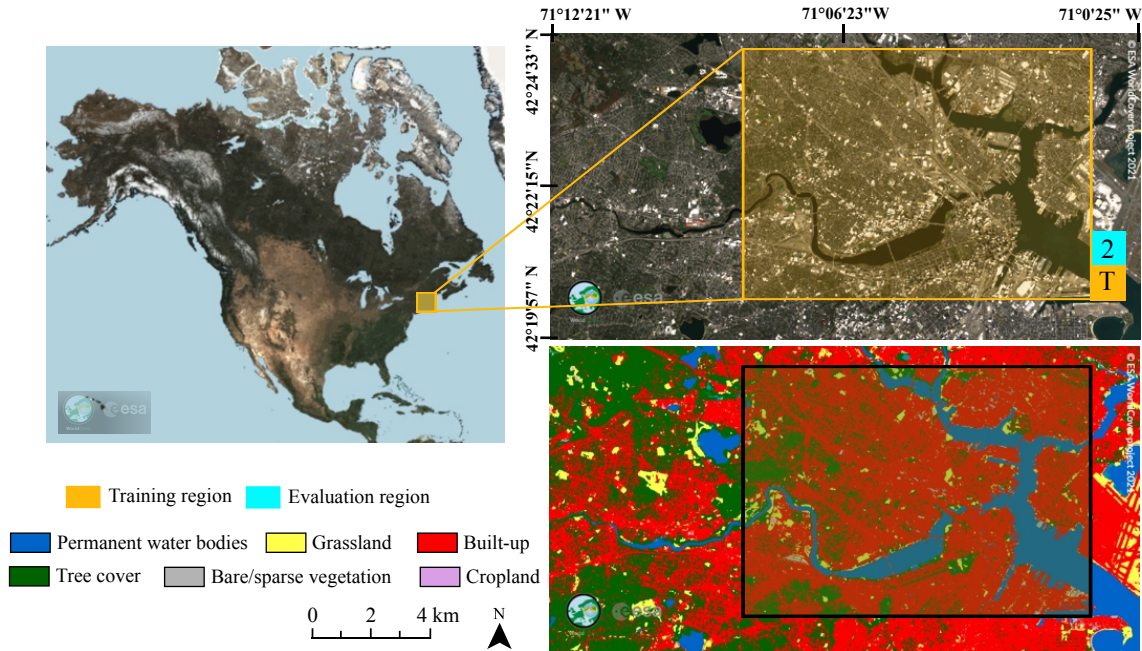


Figure 3: Sentinel-2 RGB composite images with LAT/LON information and classification map from the ESA WorldCover Map tool (Zanaga et al., 2021) of test site 2 (Charles river, Boston, USA).

2.2. Test site 2: Charles river basin

The area of study for this experiment covers the Charles river, the Mystic river and the Boston harbor in Massachusetts, with geographic center coordinates of LAT/LON: 42.36, -71.12 (see Fig. 3). This location includes a big permanent water body, urban vegetation, and a built-up area, which makes

²<https://waterdata.usgs.gov/nwis/>

it a site of interest for land cover classification. Tracking land cover changes in urban environments is of great help for urban and agriculture planning, and also when trying to identify correlations between social activities and land changes. Challenges in test site 2 include illumination variations (see Fig. 4) and the presence of reflective surfaces from buildings and seasonal cyanobacterial blooms in the Charles river and in the Boston harbor waters. Algal blooms mostly occur during summer (Rome et al., 2021), as in the image captured on 2021-07-31 shown in Fig. 4.

Sentinel-2 Level-2A images are downloaded using the Google Earth Engine platform from the COPERNICUS/S2_SR collection, with dates between 2020-09-04 and 2021-09-26. Only images with at most 10% cloud cover are downloaded. The downloaded images have dimensions of 927×2041 pixels. For evaluation purposes, images were cropped to sizes of 700×1241 pixels. Images from bands 8A and 11 are resampled to 10 meters by nearest-neighbor interpolation. After visual inspection, 15 images depicting snow-covered land or exhibiting significant disparities are excluded from the dataset, resulting in a total of 28 images for further processing.

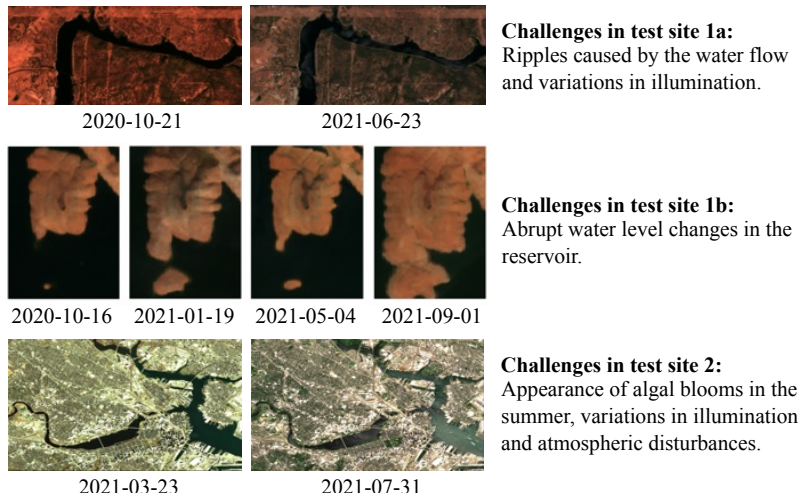


Figure 4: Sentinel-2 RGB composite images showing the challenges posed by test sites 1a, 1b, and 2, within the context of water mapping and land cover classification.

2.3. Test site 3: Amazon rainforest

The third experiment focuses on the geographic area of the Amazon rainforest in Brazil, at the geographic center coordinates LAT/LON: -4.05, -54.6 (see Fig. 5). Sentinel-2 images of the Amazon rainforest downloaded using the Google Earth Engine platform are obtained from the MultiEarth challenge dataset (Cha et al., 2023). For some dates, this dataset includes manually generated deforestation labels using mosaic satellite images from the Planet APIs³. We initially considered a total of 225 images with dates between 2018-12-03 and 2021-12-27. These have the same dimensions as the MultiEarth dataset images after being tiled and segmented, i.e., 256×256 pixels. The study area is selected to be relatively small in order to ease the analysis of results since the evolution in time of output classification maps can be more easily interpreted for smaller regions. Also, the overall computational cost decreases when lowering the number of pixels to be evaluated.

The percentages of cloud and cloud shadow are calculated with the CI_2 and CSI indices proposed by Zhai et al. (2018). From the initially 225 selected images, the ones with a cloud/shadow cover above 20% are filtered out. We also discard 12 images by visual inspection. This results in a total of 31 images for further processing. Pixels in the images from the MultiEarth dataset are shifted accordingly such that surface reflectance values are between 0 and 1, showing a considerable bias under the presence of clouds or illumination factors. To solve this, a time-varying bias is fitted to each image, for which an area where the statistics are expected to be time-invariant (i.e., no clouds or disturbances are observed inside that area for all evaluation dates) is selected. Taking the first evaluation image as a reference, the mean \bar{x}_{ref} of pixel values inside that area is calculated. For each subsequent image, a bias is computed as $b(t) = \bar{x}_{\text{ref}} - \bar{x}_{\text{test}}(t)$ and applied in the pre-processing stage so that the pixel mean in the selected

³<https://api.planet.com/>

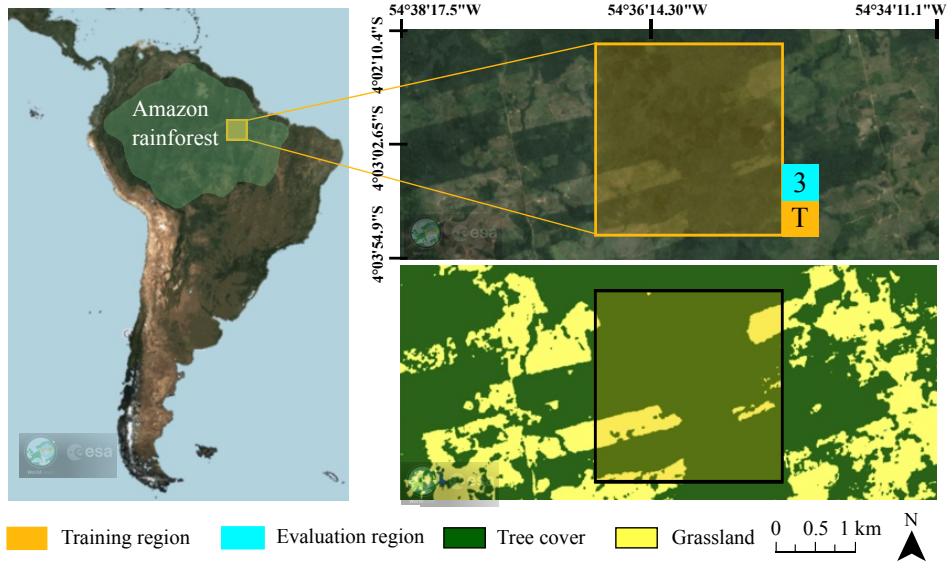


Figure 5: Sentinel-2 RGB composite images with LAT/LON information and classification maps from the ESA WorldCover Map tool (Zanaga et al., 2021) of test site 3, located in the endangered Amazon rainforest.

region is the same for all time instants. With this procedure, pixel values corresponding to the surface reflectance are not altered by the presence of clouds or other atmospheric effects.

Even after the detection and filtering of images with a relatively high cloud percentage, scenes in the Amazon rainforest often show a large number of small clouds, which may disrupt instantaneous classification. On the one hand, this demands the robustness of the proposed RBC framework. On the other hand, the temporal variability of the spectra of both vegetated and, especially, deforested areas, requires flexibility in the algorithm. This variability can be observed in the Sentinel-2 RGB composite images from dates 2019-08-10 and 2020-06-10 in Fig. 13. Overall, the proposed framework must address a trade-off between adaptability and robustness to surpass these challenges.

3. Methodology

The main contribution of this manuscript is the framework for recursive Bayesian classification using multispectral and multitemporal data, which is introduced in Section 3.1. We also propose, in Section 3.2, a classifier that uses spectral indices to generate predictive class probabilities. The experimental setup is described in Section 3.3.

3.1. Algorithm: recursive Bayesian classification (RBC)

The Bayesian philosophy involves updating beliefs based on evidence. It begins with a prior probability, i.e., the initial belief about a class label, which is updated with new data to generate a posterior probability, i.e., the refined belief after considering the new observation (Särkkä and Svensson, 2023). Let us denote by $\mathbf{Z}_t \in \mathbb{R}^{B \times N}$ an image with B bands and N pixels observed at time instant $t \in \{1, \dots, T\}$. The images at the different time instants are supposed to be coregistered, that is, they constitute observations of the same geographic scene. For each pixel $\mathbf{z}_{t,n} \in \mathbb{R}^B$, being $n \in \{1, \dots, N\}$, we associate a label $C_{t,n} \in \mathcal{C}$, where \mathcal{C} is an experiment-dependent set containing the possible K labels. Bayesian recursion allows the refinement of class probabilities at each time step as new data is observed. For a set of images \mathbf{Z}_t over time, the most likely label $C_{t,n}$ for each pixel $\mathbf{z}_{t,n}$ (i.e., the n -th column of \mathbf{Z}_t) can be estimated based on all the previous imagery $\{\mathbf{Z}_t, \mathbf{Z}_{t-1}, \dots, \mathbf{Z}_1\}$ by maximizing the posterior probability $p(C_{t,n} | \mathbf{Z}_t, \mathbf{Z}_{t-1}, \dots, \mathbf{Z}_1)$ as

$$\hat{C}_{t,n} = \arg \max_{C_{t,n} \in \mathcal{C}} p(C_{t,n} | \mathbf{Z}_t, \mathbf{Z}_{t-1}, \dots, \mathbf{Z}_1), \quad (1)$$

where $\hat{C}_{t,n}$ denotes the decision from the classifier. The expression in Eq. (1) is powerful, as it considers both temporal and spatial information. However, learning the posterior PMF in Eq. (1) can be hard, especially with high dimensional images. A spatial independence assumption can be applied to reduce the

computational cost when calculating the conditional PMF. We propose to treat the label of every pixel as independent of the data from other pixels, meaning that $C_{t,n}$ only depends on $\mathbf{z}_{t,n}, \mathbf{z}_{t-1,n}, \dots, \mathbf{z}_{1,n}$, or, equivalently, on $\mathbf{z}_{1:t,n} \triangleq \{\mathbf{z}_{t,n}, \mathbf{z}_{t-1,n}, \dots, \mathbf{z}_{1,n}\}$. This is without loss of generality, as the proposed approach can be directly extended to consider spatial information (i.e., from multiple pixels). Thus, the posterior in Eq. (1) becomes $p(C_{t,n}|\mathbf{z}_{1:t,n})$, disregarding spatial information, and leading to

$$\hat{C}_t = \arg \max_{C_t \in \mathcal{C}} p(C_t|\mathbf{z}_{1:t}), \quad (2)$$

where the pixel index n is omitted for simplicity. The classifier proposed in Eq. (2) still considers a temporal dependence on previous data, meaning that the labels and images at previous time instants influence the results of the current time t . Specifically, we assume first-order Markovity, with the dependencies between class labels (states) and pixel observations (measurements) depicted in Fig. 6a. This Markov property implies that C_t , (and the whole future C_{t+1}, C_{t+2}, \dots) given C_{t-1} is independent of anything that occurred prior to time step $t-1$.

The probability of a class label C_{t-1} given all the data up to time $t-1$ corresponds to the posterior distribution at time $t-1$, denoted as $p(C_{t-1}|\mathbf{z}_{1:t-1})$. From this distribution, the transition probabilities allow us to build the prior or predictive distribution at time step t as $p(C_t|\mathbf{z}_{1:t-1})$. Finally, with Bayes' rule, we can compute the posterior distribution with the newly observed pixel z_t as $p(C_t|\mathbf{z}_{1:t})$. This process allows the probability of class labels to be refined at each time step and is illustrated in Fig. 6b. With RBC, recursion can be applied to generative models (see Section 3.1.1), and to discriminative models (see Section 3.1.2).

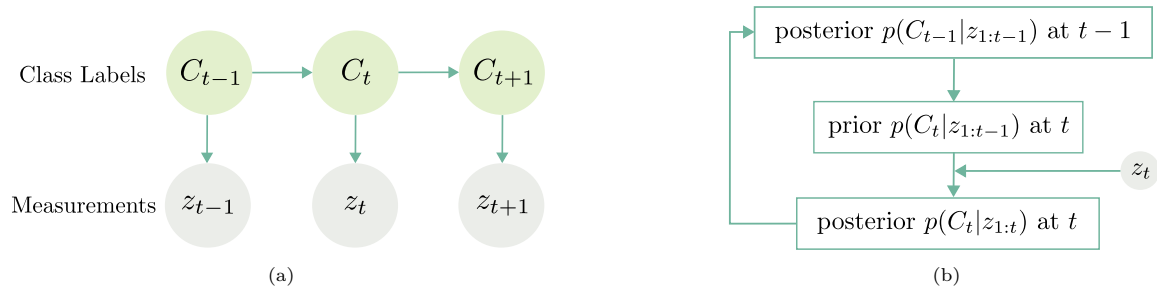


Figure 6: Diagrams illustrating the Bayesian recursion process. (a) Dependencies between class labels and pixel observations based on the first-order Markov assumption. (b) Bayesian recursion flow, updating the prior distribution to the posterior distribution using new pixel observations.

3.1.1. Classification based on a generative model (RBGM)

The posterior PMF in Eq. (2) can be computed recursively using Bayes theorem under conditional independence assumptions and assuming knowledge about the class prior probabilities $p(C_0)$, which reflect the initial understanding about the scene, the state transition $p(C_t|C_{t-1})$ and the likelihood distribution $p(\mathbf{z}_t|C_t)$ given by a generative model. Thus, the posterior PMF can be computed as

$$\begin{aligned} p(C_t|\mathbf{z}_{1:t}) &\stackrel{(a)}{=} \frac{p(\mathbf{z}_t|C_t)p(C_t|\mathbf{z}_{1:t-1})}{p(\mathbf{z}_t|\mathbf{z}_{1:t-1})} \\ &\stackrel{(b)}{=} p(\mathbf{z}_t|C_t) \frac{\sum_{C_{t-1} \in \mathcal{C}} p(C_t|C_{t-1})p(C_{t-1}|\mathbf{z}_{1:t-1})}{\sum_{C'_t \in \mathcal{C}} p(\mathbf{z}_t|C'_t)p(C'_t|\mathbf{z}_{1:t-1})} \\ &= p(\mathbf{z}_t|C_t) \frac{\sum_{C_{t-1} \in \mathcal{C}} p(C_t|C_{t-1})p(C_{t-1}|\mathbf{z}_{1:t-1})}{\sum_{C'_t \in \mathcal{C}} p(\mathbf{z}_t|C'_t) \sum_{C'_{t-1} \in \mathcal{C}} p(C'_t|C'_{t-1})p(C'_{t-1}|\mathbf{z}_{1:t-1})}, \end{aligned} \quad (3)$$

where in equality (a) we assumed the conditional independence of measurements, that is, given the class label C_t , \mathbf{z}_t is independent of the previous $\mathbf{z}_{1:t-1}$; in equality (b) we assumed a first-order Markov model, considering that given the previous class label C_{t-1} , the current class label is independent of past measurements. We refer to the method in Eq. (3) as recursive Bayesian classification based on a generative model (RBGM) due to its dependence on the likelihood function $p(\mathbf{z}_t|C_t)$. The term $p(C_{t-1}|\mathbf{z}_{1:t-1})$ denotes the posterior PMF of the previous time step. When $t=1$, $p(C_{t-1}|\mathbf{z}_{1:t-1}) = p(C_0)$ becomes equivalent to the marginal class probabilities at $t=0$, which we assume to be uniform in the absence of prior information about the scene, i.e., $p(C_0) = \frac{1}{K} \forall C_t$. The transition PMF $p(C_t|C_{t-1})$ is described

later in this section.

3.1.2. Classification based on a discriminative model (RBDM)

The posterior probability can also be calculated as a function of the probability of the labels given the pixel values, which allows existing classification algorithms to be used in the RBC framework. We refer to this as recursive Bayesian classification based on a discriminative model (RBDM). Applying the Bayes' rule to the likelihood $p(\mathbf{z}_t|C_t)$ we obtain

$$p(\mathbf{z}_t|C_t) = \frac{p(C_t|\mathbf{z}_t)p(\mathbf{z}_t)}{p(C_t)}, \quad (4)$$

where $p(C_t|\mathbf{z}_t)$ is the prediction of the classifier to which the RBC framework is applied, i.e., the benchmark classifier. As the RBC framework is agnostic to the classifier that is used, the prediction can be the result of any type of classifier, including deep learning methods as well. The Bayes theorem can be used to extend RBGM to RBDM by applying the Bayes theorem to Eq. (3) as

$$\begin{aligned} p(C_t|\mathbf{z}_{1:t}) &= \frac{p(C_t|\mathbf{z}_t)p(\mathbf{z}_t)}{p(C_t)} \frac{\sum_{C_{t-1} \in \mathcal{C}} p(C_t|C_{t-1})p(C_{t-1}|\mathbf{z}_{1:t-1})}{\sum_{C'_t \in \mathcal{C}} \frac{p(C'_t|\mathbf{z}_t)p(\mathbf{z}_t)}{p(C'_t)} \sum_{C'_{t-1} \in \mathcal{C}} p(C'_t|C'_{t-1})p(C'_{t-1}|\mathbf{z}_{1:t-1})} \\ &= \frac{p(C_t|\mathbf{z}_t)}{p(C_t)} \frac{\sum_{C_{t-1} \in \mathcal{C}} p(C_t|C_{t-1})p(C_{t-1}|\mathbf{z}_{1:t-1})}{\sum_{C'_t \in \mathcal{C}} \frac{p(C'_t|\mathbf{z}_t)}{p(C'_t)} \sum_{C'_{t-1} \in \mathcal{C}} p(C'_t|C'_{t-1})p(C'_{t-1}|\mathbf{z}_{1:t-1})}, \end{aligned} \quad (5)$$

where $p(C_t)$ denotes the marginal class probability. In the widely used naive Bayes classifier, the marginal class probabilities are also used (Barber, 2011). In the absence of labeled training data and prior information about the scene, we set their value to $p(C_t) = \frac{1}{K} \forall C_t$.

Note that the proposed recursive classification solution is in closed-form and consists of a summation of probability distributions over the different classes. Eqs. (3) and (5) are straightforward to compute given the likelihood of the pixels or their posterior probability, respectively, which correspond to the result given by the instantaneous classifier. Considering this, it can be stated that recursion does not add a significant computational overhead to the classification problem.

3.1.3. Class transition probabilities

In this work, we assume that the class transition probability $p(C_t|C_{t-1})$, which represents the likelihood of transitioning from class C_{t-1} to class C_t , is time-invariant. Although strong, this assumption copes with the lack of knowledge we assume regarding the studied scene. Moreover, we highlight that this is without loss of generality since prior knowledge about, e.g., seasonality, can be easily incorporated in a time-dependent transition PMF. The stationary or time-invariant case requires selecting K^2 parameters, but this is simplified by assuming $p_{ij} = p(C_t = j|C_{t-1} = i) = \epsilon$ for all $i \neq j$ when $K = 2$, reducing the problem to selecting a single parameter. This parameter, known as the transition probability hyperparameter ϵ , corresponds to the probability of a pixel transitioning from one label to another in the two-class case. A study on the model sensitivity to this hyperparameter is presented in Section 4.3.

3.1.4. Implicit regularization of the posterior

Note that the proposed RBC framework relies in probabilistic classifiers or generative models. Although many deep learning classifiers are currently trained based on the cross-entropy loss, which leads to a maximum likelihood estimation of the class labels (Barber, 2011), very flexible models, such as deep neural networks, can lead to overconfident classification results, i.e., there being some j such that $p(C_t = j|\mathbf{z}_t) \approx 1$. This can be damaging when such models are integrated into the proposed RBC framework since such overconfidence diminishes the relevance of the prior information obtained in previous time instants through the recursion. To remedy this issue, we propose to empirically reduce the confidence in the predictions of deep learning models before integrating them into the proposed framework, with this simple relation

$$p(C_t|\mathbf{z}_t) = \frac{p'(C_t|\mathbf{z}_t) + \lambda}{\sum_{C'_t \in \mathcal{C}} (p'(C'_t|\mathbf{z}_t) + \lambda)}, \quad (6)$$

where $p'(C_t|\mathbf{z}_t)$ is the distribution of the overconfident discriminative model, being it the probability of the labels $C_t \in \mathcal{C}$ given the pixel value at time instant t , and $\lambda \in \mathbb{R}_+$ is a positive constant used to slightly push the predicted class probabilities towards $\frac{1}{K}$ (i.e., towards a discrete uniform distribution). The same idea can be applied to an overconfident generative model $p'(\mathbf{z}_t|C_t)$ (Barber, 2011, Chapter 20.3).

The proposed approach is motivated, at a high level, from the *maximum entropy principle* (Jaynes, 1957). This principle states that among all available solutions that fit some measurements, the most suitable solution is the one with the highest entropy. In this paper, an estimation problem has not been defined for this matter because the PMFs (i.e., the likelihood or class posterior) are directly obtained. Consequently, the PMF entropy is increased in an *ad-hoc* fashion by following Eq. (6). This regularization of the posterior, or likelihood, is of great importance within the context of recursive classification of time-series data since overconfident classifiers can mask the prior information from previous time instants, jeopardizing the algorithm performance.

3.1.5. Computational Overhead of the Recursion

Neural networks or more complex algorithms that classify entire batches of data often experience a substantial increase in complexity with sequence length (Hoberg et al., 2015; Kenduywo et al., 2017). In contrast, and given the nature of Bayesian recursion described at the beginning of Section 3.1, the computational cost of RBC is not affected by the baseline classifier it is built upon and remains constant for each time step regardless of the length of the image time-series. To quantify this, we analytically evaluate the overhead introduced by recursion in terms of operations.

Assuming that each operation consists of a sum and product or a sum and division, and given that $p(z_t|C_t)$, $p(C_{t-1}|z_{1:t-1})$ and $p(C_t|z_t)$ are already computed (hence, their evaluation cost is the same as for the instantaneous classifier), we can present the following calculations. For the RBGM, the overhead cost is $K \times (K^2 + K + 2) = K^3 + K^2 + 2K$ operations. This breaks down into K operations in the summation from the numerator, K operations in the inner summation for the denominator multiplied by K operations in the outer summation also in the denominator, and 2 operations from the division and multiplication by the instantaneous classifier result. This result is then multiplied by K because the expression in Eq. (3) needs to be computed for each class. By applying the same logic, the overhead cost for the RBDM can be expressed as $K \times (K \times (K + 1) + K + 2) = K^3 + 2K^2 + 2K$. We can argue that this is not substantial in the context of most classifiers since K is generally small. For detailed quantitative results on the time required by recursion and instantaneous classifiers across all test sites and algorithms tested in this study, please refer to Section 4.4.

3.2. Spectral Index Classification (SIC)

We introduce the SIC algorithm, which uses broadband spectral indices to generate the predictive probability of occurrence of land classes, such as water or soil. Spectral indices are of interest for classification algorithms due to their clear interpretability and lack of supervision, as explained in Section 1.1. An overview of this classifier can be found in Fig. 7. The class probability $p(C_t|z_t)$ is defined as

$$p(C_t|z_t) = \frac{f_{C_t}(y(z_t))}{\sum_{C'_t \in \mathcal{C}} f_{C'_t}(y(z_t))}, \quad (7)$$

where $y(z_t)$ corresponds to the spectral index value, which is computed as a function of the pixel z_t . This is a similar but not equivalent idea to applying a softmax function. To compute the probability value, we use a Gaussian function as $f_{C_t} = \mathcal{N}(\mu_{C_t}, \sigma_{C_t})$. A different value of mean and standard deviation can be assigned for each class as $\boldsymbol{\mu} = \text{col}\{\mu_{C_t}\}$ and $\boldsymbol{\sigma} = \text{col}\{\sigma_{C_t}\}$, being $\text{col}\{\cdot\}$ the operator returning a vector whose elements are μ_{C_t} and σ_{C_t} for $C_t \in \mathcal{C}$, respectively. The function f_{C_t} can be expressed as

$$f_{C_t}(y(z_t)) = \frac{1}{\sigma_{C_t} \sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{y(z_t) - \mu_{C_t}}{\sigma_{C_t}}\right)^2\right), \quad (8)$$

thus giving a measure of how close the spectral index $y(z_t)$ is to the mean value of each class C_t , denoted as μ_{C_t} . This is used as an indication of the likelihood of z_t being of class C_t . The standard deviation σ_{C_t} accounts for the length of the spectral index interval defining class C_t . For ease of exposition, let us consider for the remainder of this section that $C_t \in \mathcal{C} = \{1, \dots, K\}$, and also that the class indices are ordered in the same way as the threshold intervals, i.e., class i corresponds to the i -th spectral index interval.

The length of the intervals defining each class can be highly non-homogeneous and depends on the spectral index class thresholds τ_i , where $i \in \{0, \dots, K\}$. These thresholds define a hard classification result based on the spectral index value, with pixel z_t being assigned to the i -th class if and only if $y(z_t) \in (\tau_{i-1}, \tau_i]$. Their length can be calculated as $L_j = \tau_j - \tau_{j-1}$, where $j \in \{1, \dots, K\}$. The values of $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ are calculated as $\mu_j = L_j/2 + \tau_{j-1}$ and $\sigma_j = L_j/2$, so that the probability of a pixel belonging

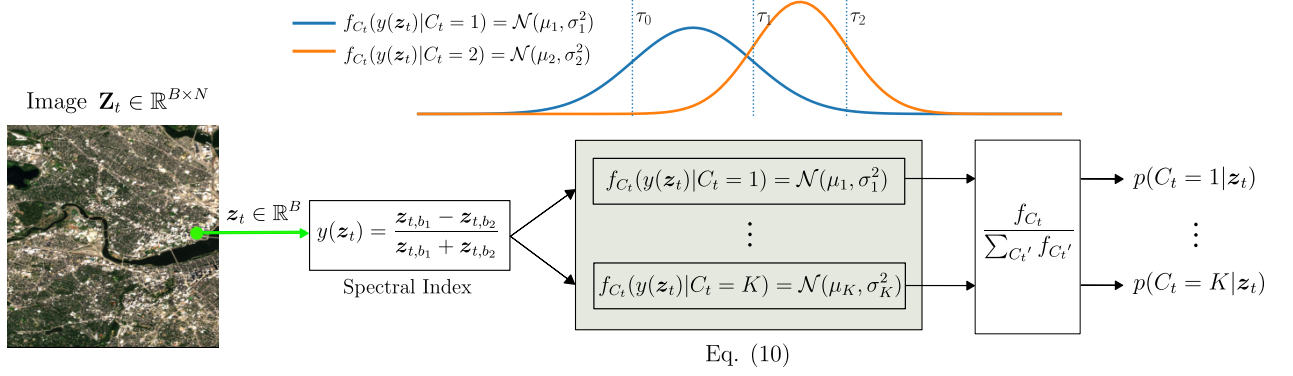


Figure 7: Overview of the proposed Spectral Index Classifier (SIC), where the values of standard broadband spectral indices such as the NDWI are converted into probabilities $p(C_t|\mathbf{z}_t)$ as in Eq. (7). The proposed RBC framework can be applied on top of the SIC algorithm, resulting in the recursive SIC (RSIC) model.

to a given class decreases smoothly as $y(\mathbf{z}_t)$ moves away from the center of the interval and approaches one of the thresholds. The threshold values are determined empirically and are experiment-dependent, as discussed in Section 3.3.3.

3.3. Experimental setup

A classifier based on a GMM, an LR classifier, and the SIC algorithm introduced in Section 3.2, are compared to their recursive counterparts, namely the RGMM, RLR and RSIC algorithms. When working with data from test sites 1a, 1b, and 2, two additional pre-trained deep learning models are used as a benchmark in the context of water mapping: the DeepWaterMap (Isikdogan et al., 2017) and the WatNet (Luo et al., 2021) algorithms. All models under consideration are listed in Table 3. To ensure consistency when evaluating the RBC framework in different areas of study, we maintain the same number of classes across the three test sites.

As discussed in the introduction, the instantaneous likelihood or class posterior to which the RBC framework is applied may be either semi-supervised, supervised or unsupervised. When the data model is unsupervised, the entire procedure may be viewed as unsupervised. The converse is also true for supervised methods. When we describe the methodology followed in the *training stage* in Section 3.3.2, we are referring to the training stage of the GMM and LR models, which are the models that need supervision.

3.3.1. Data splitting

A proportionally scaled timeline with the dates of images used for training and evaluation can be found in Fig. 8. Each downloaded image belongs to a different date and they are mostly spaced 5 days apart, i.e., the temporal resolution of Sentinel-2 satellites. However, temporal spacing between images may vary as a consequence of filtering images with high cloud/shadow cover and other discrepancies. On the one hand, large temporal spacings between training images translate into training data diversity. On the other hand, large temporal spacings between evaluation images can pose a challenge, because a change in land that occurs gradually can be interpreted as a sudden artifact to be discarded by the recursive algorithm. This matter is further discussed in Section 5.

3.3.2. Training the GMM and LR models

The LR and GMM models are trained in a weakly supervised approach. Training images correspond to those acquired on the dates indicated with green markers in Fig. 8. To generate surrogate ground truth class labels, or pseudo-labels, the pixels from the training images are classified based on their spectral index value (MNDWI for test sites 1a, 1b and 2, and NDWI for test site 3) and considering the class thresholds τ in Table 2. To obtain the generative model $p(\mathbf{z}_t|C_t)$ used in the RBGM from Eq. (3), one GMM is trained for each class label, i.e., $p(\mathbf{z}_t|C_t)$ is a GMM for each choice of C_t . To adequately represent the training pixels without overfitting, we select the smallest number of components for each GMM such that the histograms of the training data distribution and the one generated by the respective GMM are visually close.

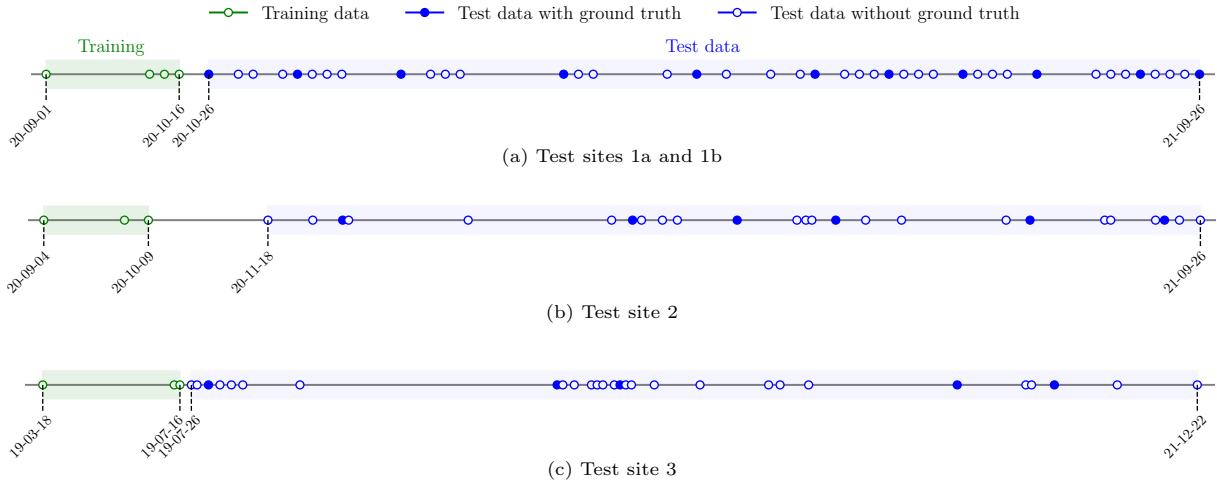


Figure 8: Proportionally scaled timeline showing Sentinel-2 image dates (yy-mm-dd) used to train the GMM and LR models (green markers), and to test the GMM, LR, SIC, DWM and WN models, and their recursive counterparts (blue markers). The filled blue markers correspond to test images for which manually generated ground truth labels are available. This allows to evaluate the proposed framework quantitatively.

3.3.3. Evaluation

Test images correspond to those acquired on the dates indicated with blue markers in Fig. 8. Across all test sites, the dataset is imbalanced, with the majority of pixels being attributed to the *land* class in the two water mapping experiments and the *forest* class in the deforestation detection experiment. To prevent biases and ensure equal contribution from each class when benchmarking between classification models, balanced classification accuracy is used as a metric for the comparative analysis between instantaneous classifiers and their recursive counterparts in Section 4.1 and Section 4.2. In Section 4.3, the same metric is used to assess the model sensitivity to the transition probability hyperparameter.

Ground truth: The lack of openly accessible labeled Sentinel-2 data for time-series analysis presents a significant challenge to the assessment of our framework. Consequently, water labels were manually generated for the dates employed in the quantitative analysis (filled blue markers in Fig. 8), and shared by the authors at (Calatrava et al., 2024). This procedure was facilitated by the LabelStudio tool⁴. In the case of test site 3, the MultiEarth challenge dataset contains deforestation labels, which removes the need to manually generate ground truth labels for the third experiment. The dates for which deforestation labels are available do not necessarily match any date from the Sentinel-2 images in the MultiEarth dataset. Taking this into account, for each of the provided five labels, error classification maps are computed between the label and the classification result that is closest in time after the label date.

Parameter settings: The parameter values used for the conducted experiments are presented in Table 2. For the recursive algorithms, ϵ is chosen to maximize the average balanced accuracy across test images with available ground truth, with optimization done independently for each algorithm and test site. Optimal ϵ values are determined through sensitivity analysis as described in Section 4.3. To prevent the adverse effects of overconfident predictions of the instantaneous classifiers, we empirically selected $\lambda = 0.8$ for all methods. To convert a standard broadband spectral index into a probability measure as in Eq. (7), it is necessary to define the thresholds τ . These are tuned accordingly so that the generated classification maps obtained with the training images are visually close to the reference maps in Figs. 2, 3 and 5. As explained in Section 3.2, the thresholds are used to calculate μ and σ (see Fig. 9).

4. Results

Complete results can be reproduced following the instructions in <https://github.com/neu-spiral/RBC-SatImg>. Overall, the RBC framework significantly increases the robustness of existing classification algorithms in multitemporal settings, while preserving the adaptability to changes in the land. Substantial improvements are shown when compared to pre-trained state-of-the-art deep learning-based classifiers, without the need for additional training data.

⁴<https://labelstud.io/>

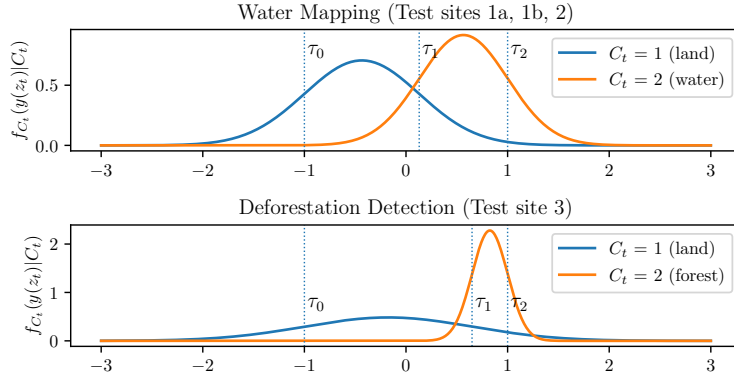


Figure 9: Function $f_{C_t}(y(z_t)|C_t)$ introduced in Section 3.2. The corresponding values of τ , μ , σ , and the spectral index $y(z_t)$ are the ones in Table 2 for the water mapping experiments (top subplot) and the deforestation detection experiment (bottom subplot).

4.1. Error classification maps

The following figures include the classification maps, error classification maps, and balanced classification accuracy results obtained using the test images with available ground truth labels. RGB composite images of the studied areas are shown as a reference because they highlight changes in the scene. In the case of test site 1b, and due to space limitations, only one every two test images with ground truth data are included in the analysis. The interested reader may find the classification map results for all test images, including those without ground truth data, in the supplemental material.

4.1.1. Test site 1

Water mapping results for test sites 1a and 1b are presented in Figs. 10 and 11, respectively. Fig. 10 shows noticeable differences among benchmark and recursive algorithms. For instance, the SIC algorithm classifies a large portion of the stream as land for dates 2021-05-19, 2021-06-13, 2021-09-06 and 2021-09-26. The same is observed with the LR and DWM classifiers for dates 2021-05-19 and 2021-09-06. Nevertheless, their recursive counterparts, i.e., the RSIC, RLR and RDWM algorithms, can adequately classify most of the stream pixels as water. This translates into an increase in balanced classification accuracy of more than 20% provided by the use of recursion. For instance, the RSIC algorithm provides an improvement of 26.95% and 20.31% for dates 2021-05-19 and 2021-06-13. The WN algorithm classifies some portions of land as water on dates between 2020-11-25 and 2021-07-08. While the RWN algorithm misclassifies part of the stream, it shows considerably more accurate classification maps. This can be especially observed for the dates 2021-04-09 and 2021-05-19, where the RBC framework offers an improvement in balanced classification accuracy of 8.85% and 4.9%, respectively. Overall, results suggest that the proposed RBC framework improves the performance of modern deep learning-based mapping algorithms. Moreover, it has been observed that the DWM and WN algorithms provide overconfident classification results, which can be an issue in recursive multitemporal classification. This is solved with the strategy proposed in Eq. (6) to regularize the predictive class posterior.

Results in Fig. 11 show that both the instantaneous classifiers and their recursive counterparts are able to adequately capture the decrease in water levels over time in the dam upstream in test site 1b. In the case of date 2020-12-30, the SIC, LR, DWM and WN algorithms misclassify a considerable amount of water as land, while their recursive counterparts output more correct classification maps. This results in an increase of balanced classification accuracy of 14.82%, 4.27%, 13.81%, 9.62%, and 11.03% for the RSIC, RGMM, RLR, RDWM and RWN algorithms respectively. Between these, the improvement provided by the RGMM classifier is more modest due to its non-recursive counterpart already performing well. A similar effect is observed on dates such as 2021-06-13, 2021-08-02, and 2021-09-26, for which the non-recursive methods provide adequate classification results and consequently the improvement introduced by recursion is moderate (under 10%). The trade-off between adaptability and robustness presents a challenge to the recursive framework for the upstream subscene of Oroville Dam. Prioritizing robustness makes the recursive framework less flexible, potentially leading to delayed detection of abrupt scene changes. For example, the increase in water level starting in February is better identified by the instantaneous classifiers, as they do not rely on information from previous images showing lower water levels. Consequently, when recursion is employed on the date 2021-04-09, there is a decrease in balanced

Table 2: Parameter settings for the three experiments conducted in this research. The transition probability ϵ is defined in Section 3.1.3 and the regularization constant λ is introduced in Section 3.1.4, Eq. (6). The thresholds τ_W and τ_D are used to generate surrogate ground truth labels to train the GMM and LR algorithms, as explained in Section 3.3.2, and to generate predictive probabilities of occurrence with the SIC algorithm (see Section 3.2).

Test sites 1a, 1b	$C_t \in \mathcal{C}_W = \{\text{land, water}\}$	
	SIC	$\tau_W = [-1, 0.13, 1];$ $\mu_W = [-0.435, 0.565];$ $\sigma_W = [0.565, 0.435];$ $y_{\text{MNDWI}}(\mathbf{z}_t) = \frac{z_{t,\text{green}} - z_{t,\text{SWIR}}}{z_{t,\text{green}} + z_{t,\text{SWIR}}}$
	RSIC	$\epsilon = 0.001$ (1a), 0.02 (1b); $\lambda = 0.8$
	RGMM	$\epsilon = 0.2, 0.09$; $\lambda = 0.8$
	RLR	$\epsilon = 0.001, 0.02$; $\lambda = 0.8$
	RDWM	$\epsilon = 0.001, 0.095$; $\lambda = 0.8$
	RWN	$\epsilon = 0.005, 0.085$; $\lambda = 0.8$
	Test site 2	$C_t \in \mathcal{C}_W$
SIC		$\tau_W; \mu_W; \sigma_W$ $y_{\text{MNDWI}}(\mathbf{z}_t)$
RSIC		$\epsilon = 0.1$; $\lambda = 0.8$
RGMM		$\epsilon = 0.001$; $\lambda = 0.8$
RLR		$\epsilon = 0.005$; $\lambda = 0.8$
RDWM		$\epsilon = 0.001$; $\lambda = 0.8$
RWN		$\epsilon = 0.001$; $\lambda = 0.8$
Test site 3	$C_t \in \mathcal{C}_D = \{\text{land, forest}\}$	
	SIC	$\tau_D = [-1, 0.65, 1];$ $\mu_D = [-0.175, 0.825];$ $\sigma_D = [0.825, 0.175]$ $y_{\text{NDWI}}(\mathbf{z}_t) = \frac{z_{t,\text{green}} - z_{t,\text{NIR}}}{z_{t,\text{green}} + z_{t,\text{NIR}}}$
	RSIC	$\epsilon = 0.03$; $\lambda = 0.8$
	RGMM	$\epsilon = 0.04$; $\lambda = 0.8$
	RLR	$\epsilon = 0.04$; $\lambda = 0.8$

classification accuracy of up to 6.79%, as indicated by the RSIC classifier. However, this setback is gradually resolved on subsequent dates, as indicated by improved classification results on 2021-06-13.

4.1.2. Test site 2

Water mapping results for test site 2 are presented in Fig. 12. The interested reader may refer to Appendix A, where classification map results are shown for a three-class classification experiment with the same data, thus demonstrating the scalability of the framework to handle more complex classification tasks. Test site 2 extends over an area covering the Boston harbor, the Charles river lower, mid and some upper basins, and the Mystic river lower basin. Since most of these correspond to urban and suburban areas, we can find many reflective surfaces from, e.g., building terraces and metal sheds, which lead to pixels with high spectral reflectance. Such pixels may be easily misclassified as water since their MNDWI values are close to zero. This can be observed in the classification maps from 2020-12-13, where the recursive algorithms result in fewer misclassifications of reflective surface pixels as water. However, this reduction in misclassifications does not necessarily translate to a performance increase in terms of balanced classification accuracy. In general, results for dates between 2020-12-13 and 2021-04-24 are already good enough for the instantaneous classifiers. Consequently, the improvement provided by their recursive counterparts is either non-existent or very small. On 2021-05-27, given the appearance of

Table 3: Full name and abbreviation of the benchmark classification models, i.e., SIC (introduced in Section 3.2), GMM, LR, DWM (Isikdogan et al., 2017) and WN (Luo et al., 2021), and their recursive counterparts, i.e., RSIC, RGMM, RLR, RDWM and RWN, which derive from applying the proposed recursive Bayesian classification framework.

Full Name	Abbreviation
Spectral Index Classifier (Section 3.2)	SIC
Gaussian Mixture Model	GMM
Logistic Regression	LR
DeepWaterMap (Isikdogan et al., 2017)	DWM
WatNet (Luo et al., 2021)	WN
Recursive Spectral Index Classifier	RSIC
Recursive Gaussian Mixture Model	RGMM
Recursive Logistic Regression	RLR
Recursive DeepWaterMap	RDWM
Recursive WatNet	RWN

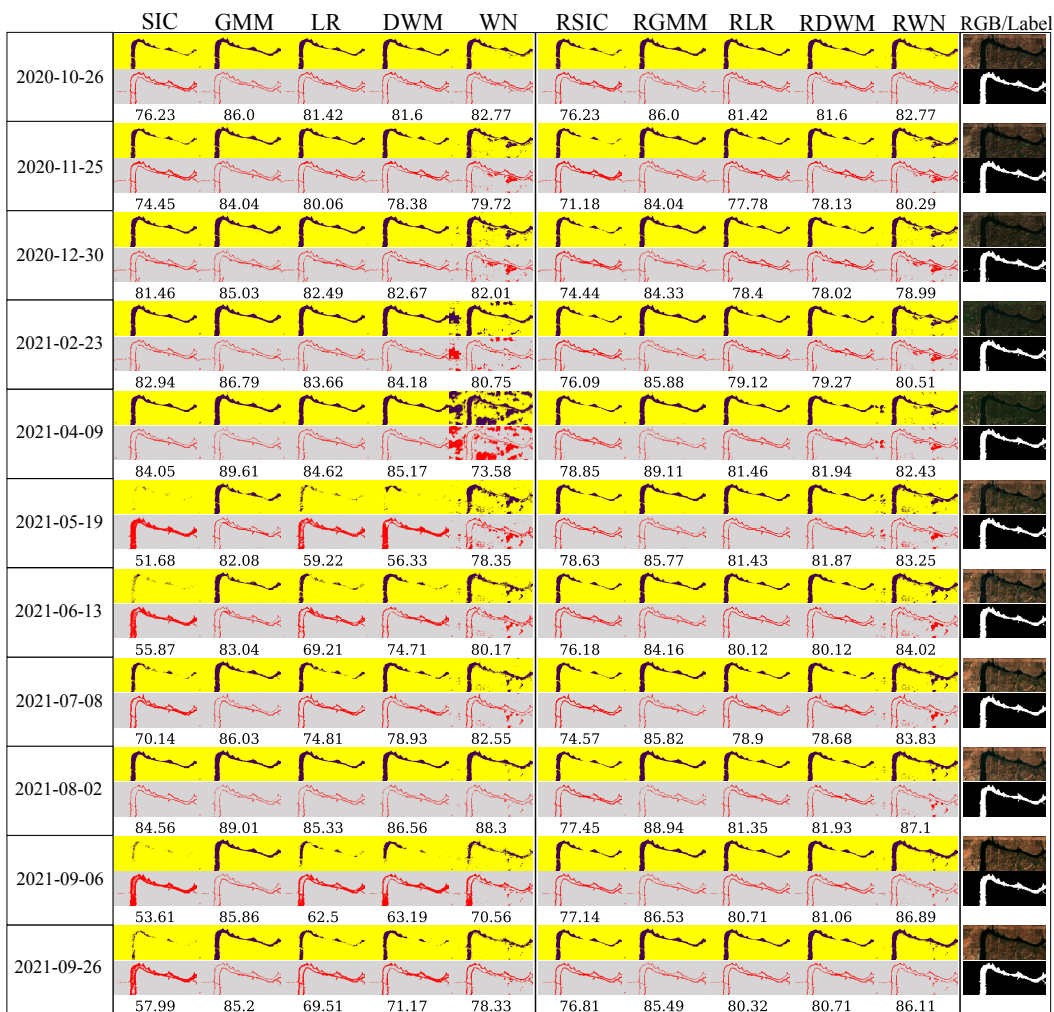


Figure 10: Water mapping results obtained for the Oroville dam downstream (test site 1a). Classification maps, error classification maps, and balanced classification accuracy results are presented. Purple and yellow represent water and land, respectively.

cyanobacterial blooms, due to which the water becomes diluted with chlorophyll pigments, an important portion of water pixels are classified as land by the instantaneous SIC, GMM, LR and DWM classifiers, whereas their recursive versions provide more accurate classification maps. This leads to an improvement in balanced classification accuracy of 12.77%, 12.4%, 8.59%, and 9.11% for the RSIC, RGMM, RLR and RDWM models respectively. The WN algorithm exhibits poor performance, particularly since 2021-

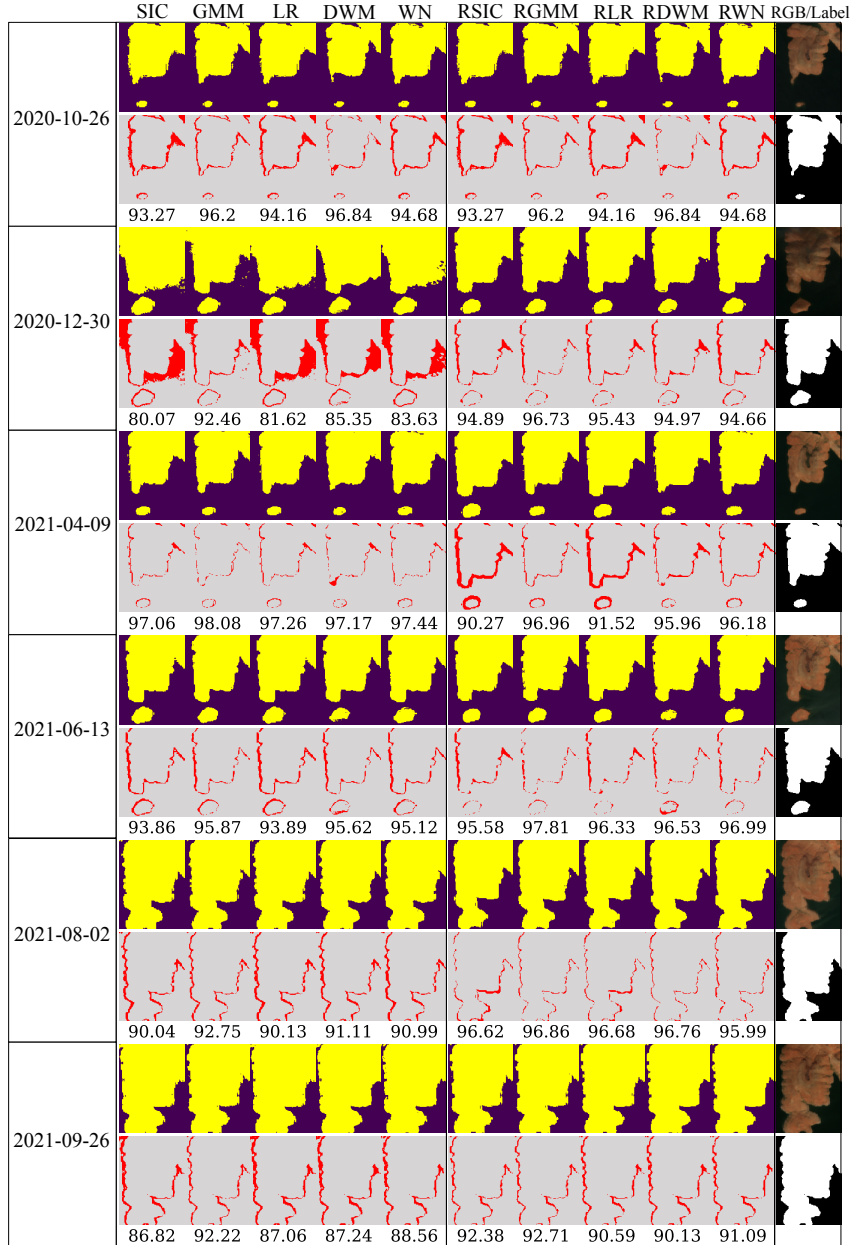


Figure 11: Water mapping results obtained for the Oroville dam upstream (test site 1b). Classification maps, error classification maps, and balanced classification accuracy results are presented. Purple and yellow represent water and land, respectively.

05-27, misclassifying a significant portion of land pixels as water. We observe an increase in balanced classification accuracy provided by its recursive counterpart of 10.58% on 2021-05-27, 6.02% on 2021-07-31, and 0.95% on 2021-09-14. The reason for this decrease in improvement over time is due to the repeated failures in the task of the WN instantaneous classifier. Overall, the recursive algorithms provide significantly more robust performance than their non-recursive counterparts. The latter are less sensitive to atmospheric interference and illumination factors. For some dates, the recursion introduces a smoothing effect, which makes it more difficult to adapt to class changes. This can be understood by comparing the classification maps obtained with the SIC and RSIC algorithms on 2021-03-20.

4.1.3. Test site 3

Deforestation detection results for test site 3 are illustrated in Fig. 13. On dates 2020-06-10, 2020-08-04, and 2021-05-26, cloud presence disrupts the performance of the instantaneous classifiers, while their recursive counterparts demonstrate adequate classification map results. This led to a rise in balanced classification accuracy of 7.06%, 14.17%, and 8.37% for the RSIC, RGMM, and RLR algorithms on 2020-

	SIC	GMM	LR	DWM	WN	RSIC	RGMM	RLR	RDWM	RWN	RGB/Label
2020-11-18											
	88.47	89.48	89.76	90.77	83.62	88.47	89.48	89.76	90.77	83.62	
2020-12-13											
	88.74	86.91	89.11	92.03	84.61	87.58	89.22	89.25	91.16	84.23	
2021-03-20											
	86.71	88.22	87.92	89.31	82.91	84.37	89.28	88.26	89.48	83.02	
2021-04-24											
	86.94	89.37	89.04	88.82	81.84	86.43	89.52	88.6	89.45	83.66	
2021-05-27											
	74.39	77.21	80.27	80.21	72.67	87.16	89.61	88.86	89.32	83.25	
2021-07-31											
	81.71	80.58	84.84	86.5	71.93	86.44	89.48	88.58	88.96	77.95	
2021-09-14											
	83.58	88.08	86.64	86.21	75.12	87.86	89.62	89.01	89.0	76.07	

Figure 12: Water mapping results obtained for the Charles river basin (test site 2). Classification maps, error classification maps, and balanced classification accuracy results are presented. Purple and yellow represent water and land, respectively.

06-10, and 15.25%, 10.58%, and 14.7% on 2021-05-26. On 2020-08-04, despite improved classification maps, there is a loss in balanced classification accuracy of 1.6% and 1.88% for the RGMM and RLR algorithms respectively. We attribute this to the class imbalance in the dataset, where the forested area substantially exceeds the deforested area. The RSIC algorithm, however, provides a slight improvement of 0.97% for that date. The adaptability of the framework is evident in the classification map results from 2021-05-26, which reveal a newly deforested area detected by the three recursive algorithms but missed by their non-recursive counterparts. On the subsequent date (2021-08-19), only the RSIC algorithm can detect this same deforested area, which remains undetected for the other algorithms. This failure of the RGMM and RLR algorithms is due to repeated failures in the task by the instantaneous classifiers, which disrupts the performance of their recursive counterparts. Supplemental results show that the recursive classifiers need two iterations where the instantaneous classifiers detect the deforested area to acknowledge this change, as a consequence of the trade-off between the robustness and adaptability of the framework.

4.2. Classification accuracy visualization

The boxplot in Fig. 14 shows the distribution of the balanced classification accuracy results presented in the previous subsection. The introduction of recursion mitigates the negative outliers from the non-recursive models. Additionally, negative outliers from the recursive models fall within the interquartile range of the non-recursive models, or are slightly below their lower quartile in the case of the RGMM for test site 1b and the RLR for test site 3. For instance, when analyzing the RSIC model with data from test site 1a, there is a significant reduction in result variability, with one negative outlier falling within the interquartile range of the SIC model.

A noticeable trend is the reduced variability in performance among the recursive algorithms, evident from the narrower spread of balanced accuracy values. This is specially apparent for the RGMM model. While the upper quartiles remain consistent and do not show a significant increase, the lower range of results for the recursive models is notably higher than the ones offered by their recursive counterparts. This suggests that although peak accuracies do not exhibit a significant rise, there is a marked enhancement

	SIC	GMM	LR	RSIC	RGMM	RLR	RGB/Label
2019-08-10							
	72.57	75.17	72.36	72.73	75.32	72.25	
2020-06-10							
	79.39	73.26	75.63	86.45	87.43	84.0	
2020-08-04							
	83.13	84.79	82.44	84.1	83.19	80.56	
2021-05-26							
	69.28	71.85	66.41	84.53	82.43	81.11	
2021-08-19							
	82.92	82.44	80.47	88.92	83.68	80.39	

Figure 13: Deforestation detection results obtained for the Amazon rainforest area (test site 3). Classification maps, error classification maps, and balanced classification accuracy results are presented. Purple and yellow represent forest and deforested area, respectively. For this experiment, Sentinel-2 images and deforestation labels have been obtained from the MultiEarth challenge dataset (Cha et al., 2023).

in the lower-end performance, implying a more consistent and improved overall performance across the tested algorithms. This results from the robustness provided by the RBC framework. Upper quartiles for the recursive classifiers for test sites 1b and 3 indicate a noticeable enhancement. Finally, in the case of the RSIC, RLR, and RDWM algorithms when tested with data from test site 1a, we observe a decrease in the upper quartile values that may have originated from the higher variance in the accuracies provided by their instantaneous classifiers.

4.3. Sensitivity analysis

The class transition probabilities are governed by the hyperparameter ϵ introduced in Section 3.1.3. This hyperparameter is site-specific, emphasizing the need for careful selection of its value. An analysis of the RBC model sensitivity to ϵ has been conducted using the data with available ground truth for each test site. This provided the optimal values for the hyperparameter, which have been used in the three conducted experiments and can be found in Table 2. The results in Fig. 15a show the balanced classification accuracy concerning values of ϵ between 0.001 and 0.8 for the water mapping experiment with data from test site 1a. Fig. 15b shows results for data from test site 3 in the context of deforestation detection.

Given that we consider uniform marginal class probabilities, results for $\epsilon = 0.5$ match the non-recursive benchmark given by the instantaneous classifiers. The improvement with respect to the non-recursive algorithms is notable for $\epsilon < 0.5$. For the water mapping task, the RBC framework provides an increase

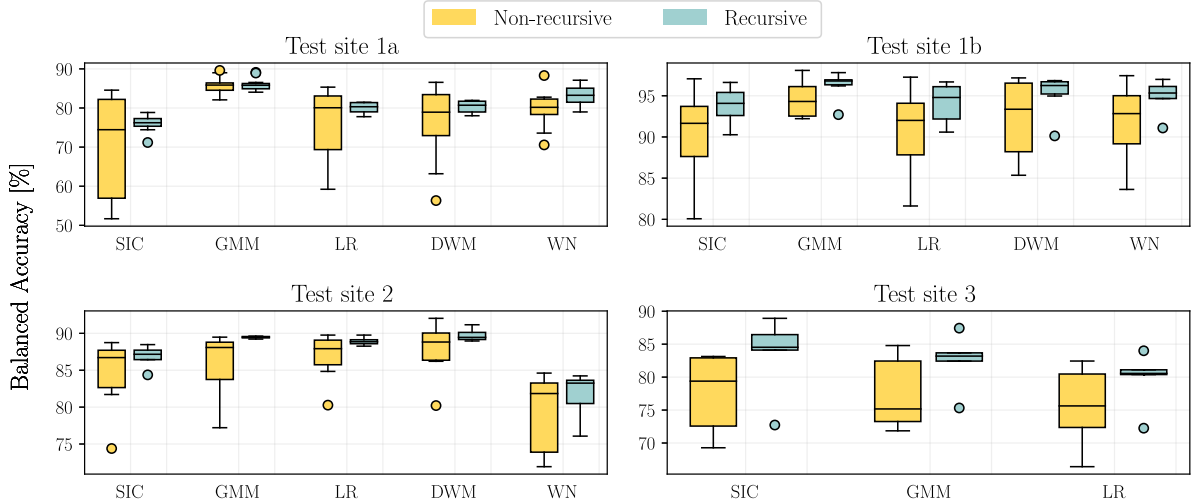


Figure 14: Boxplot showing the distribution of balanced accuracy results for the SIC, GMM, LR, DWM and WN algorithms (yellow boxes), including their recursive counterparts (blue boxes). The boxplot illustrates the median with a line within each box, the interquartile range with the box boundaries, and the data range with whiskers. Outliers are represented as filled individual circles.

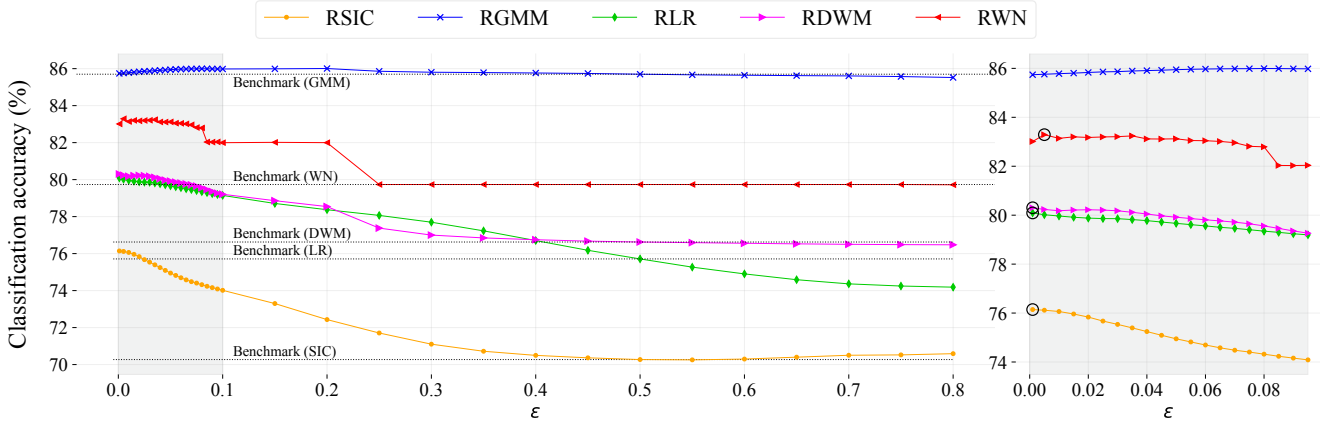
in average balanced classification accuracy of up to 5.87%, 0.3%, 4.38%, 3.68%, and 4.55% for the SIC, GMM, LR, DWM, and WN classifiers respectively. The RGMM algorithm gives a lower improvement because its non-recursive counterpart provides an already good performance, which can be understood from the low variance of the non-recursive GMM boxplot for test site 1a in Fig. 14. For the deforestation detection task, the improvements in accuracy are of up to 5.9%, 4.9%, and 4.2% for the SIC, GMM, and LR classifiers respectively. For $\epsilon > 0.5$, the recursive algorithms demonstrate inferior performance compared to their non-recursive counterparts. This makes sense, as $\epsilon > 0.5$ suggests a higher likelihood for a pixel to transition between classes rather than remaining in the same class, which is an unrealistic hypothesis. Although the accuracy provided by the RSIC model increases slightly for $\epsilon > 0.6$, this result does not appear to be influential in shaping the conclusions.

The best performance is achieved for $0 < \epsilon < 0.1$. This region is shaded in grey and magnified in the right subplot of the figures. The ϵ values yielding the best results are 0.001, 0.2, 0.001, 0.001, and 0.005 for the RSIC, RGMM, RLR, RDWM, and RWN models in the water mapping analysis, and 0.03, 0.04, and 0.04 for the RSIC, RGMM, and RLR models in the deforestation detection analysis. The RGMM algorithm in the water mapping analysis shows a wider range of ϵ values providing classification accuracy results near the best accuracy, and a notably lower sensitivity to ϵ variations. The sensitivity to this hyperparameter is highly influenced by the performance of the instantaneous classifier, in particular to the variance of its balanced classification accuracy results. As an example, the RSIC algorithm shows the highest sensitivity to ϵ in Fig. 15a, being the SIC algorithm the one showing the highest interquartile range in the subplot from Fig. 14 for test site 1a. Interquartile ranges in the subplot for test site 3 are similar between the three algorithms under study, resulting in similar sensitivities to ϵ in Fig. 15b.

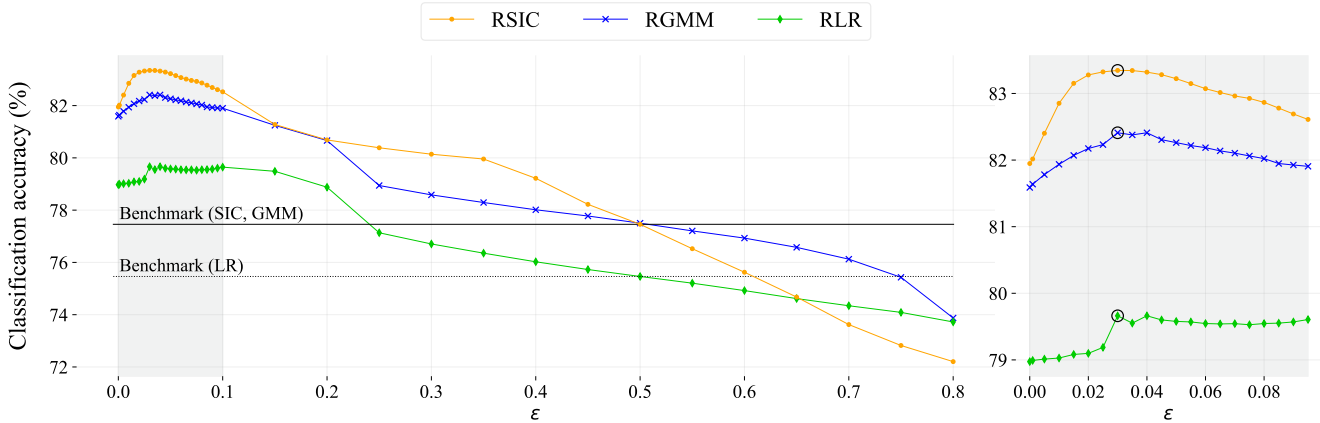
4.4. Computational Cost Analysis

We provide an analysis of the computational cost associated with the RBC framework. Specifically, we examine the computation time dedicated to recursion and the time required by the corresponding instantaneous classifier (SIC, GMM, LR, WN, or DWM) for each test site and algorithm, measured over a single time step. RBC transforms any instantaneous classifier into an online method with minimal overhead added by recursion at each time step, as demonstrated by the magnitudes in Table 4. Using the code in <https://github.com/neu-spiral/RBC-SatImg>, the cost analysis was conducted on an Apple M1 Pro (10 cores, 16-core integrated GPU) with 16 GB RAM, running macOS Sonoma 14.4.1. TensorFlow was used to integrate GPU via the METAL API.

Execution times vary across classifiers due to their complexity, with deep learning models like WN and DWM taking longer (2.49 and 1.49 seconds on test site 1a) than simpler models like SIC and LR (0.005 seconds on test site 1a). Notably, and given the nature of Bayesian recursion explained in Section 3.1, the computational cost of recursion is independent of the complexity of the underlying instantaneous classifier. For instance, across test site 1a, the recursion times for all models are similar, ranging from



(a) Sensitivity analysis for the water mapping task (test site 1a). Maximum accuracy is given by $\epsilon = 0.001, 0.2, 0.001, 0.001, 0.005$ for RSIC, RGMM, RLR, RDWM and RWN.



(b) Sensitivity analysis for the deforestation detection task (test site 3). Maximum accuracy is given by $\epsilon = 0.03, 0.04, 0.04$ for RSIC, RGMM and RLR.

Figure 15: Sensitivity analysis conducted with data from test sites 1a (water mapping) and 3 (deforestation detection). Average balanced classification accuracy results are presented for ϵ values between 0.001 and 0.8. The best performance is achieved for the region $0 < \epsilon < 0.1$, which is shaded in grey and magnified in the right subplots.

0.007 to 0.008 seconds. The cost of recursion increases, however, with image size. For example, test site 2, with the largest images (927×2041 pixels), has recursion times of 0.06 to 0.07 seconds, while test site 1b, with smaller regions (150×110 pixels), shows recursion times as low as 0.002 to 0.003 seconds. The minimal overhead introduced by recursion is mostly evident for more complex models like RWN and RDWM. For example, the recursion time for RWN constitutes only 0.32% of the baseline time on test site 1a, 0.09% on test site 1b, and 1.39% on test site 2. Similarly, for RDWM, the recursion time is 0.54% on test site 1a, 0.15% on test site 1b, and 4.32% on test site 2. Moreover, given the nature of Bayesian recursion introduced in Section 3.1 the RBC cost at each time step remains constant regardless of the time-series length. This reinforces its suitability for real-time remote sensing applications.

5. Discussion

This section discusses the implications of the main findings from this research. For a quantitative summary of the results, please refer to the conclusion in Section 6.

5.1. Study Overview and Significance

Real-time updating of classification maps in remote sensing remains a significant challenge, particularly when training data is limited. This article presents the RBC framework as a solution that enhances robustness and accuracy in land cover classification across multitemporal settings. By using information from previous time steps, RBC successfully handles disturbances present in remote sensing imagery such as illumination and atmospheric interference, e.g., different aerosol concentrations or viewing angles. It can be applied atop any instantaneous classifier based on either a generative or discriminative model.

Table 4: Total computation time in seconds for recursion and instantaneous classification (baseline) across different test sites for RSIC, RGMM, RLR, RWN, and RDWM algorithms.

Metric	Test Site	RSIC	RGMM	RLR	RWN	RDWM
Recursion Time (s)	1a	0.008	0.007	0.007	0.008	0.008
	1b	0.003	0.002	0.002	0.002	0.002
	2	0.06	0.07	0.06	0.06	0.07
	3	0.004	0.005	0.004	N/A	N/A
Baseline Time (s)	1a	0.005	0.12	0.005	2.49	1.49
	1b	0.003	0.07	0.002	2.19	1.32
	2	0.04	0.92	0.02	4.32	1.62
	3	0.002	0.02	0.001	N/A	N/A

Furthermore, classification results obtained with a three-class experiment using data from test site 2, as presented in [Appendix A](#), illustrate the scalability of the RBC framework to more complex classification tasks.

This paper also introduces the SIC classifier, which converts a spectral index value into a probability measure using the mapping in Eq. (7) (please refer to [Fig. 7](#)). Spectral indices are highly regarded in the remote sensing community owing to their simplicity, interpretability and low computational cost. However, they are often sensitive to changes in illumination and pixel disturbances. Applying RBC on top of SIC helps to address these difficulties and improves the overall strength of the classification.

The versatility of RBC is shown in its application to various models, ranging from traditional algorithms for machine learning like GMM and LR to sophisticated deep learning models such as DeepWaterMap and WatNet. Deep learning models, such as deep neural networks, offer great flexibility but can sometimes lead to overconfident classification results. This overconfidence reduces the impact of information from previous time steps and therefore compromises algorithm robustness. To circumvent this phenomenon, we propose to empirically reduce such overconfidence by inserting a positive constant to slightly push the probabilities towards a discrete uniform distribution as in Eq. (6).

Reliability and low computational overhead (as demonstrated in [Section 4.4](#)) make RBC an excellent choice for real-time remote sensing applications where fast updates are essential, e.g., environmental monitoring and disaster management. RBC can enable quick decision-making in the presence of events like deforestation, wildfires, and floods. Being able to manage such diverse and critical tasks highlights the broad importance and potential influence of the proposed framework in many real-world applications.

RBC is simple, easy to use, interpretable, and controlled by a unique explainable hyperparameter ϵ which regulates the probability of transitioning among the different classification labels. As a consequence, ϵ governs the trade-off between adaptability to natural changes in the scene and robustness to outliers caused by illumination of atmospheric interferences. As the class transition probability is specific to each site and the instantaneous classification algorithm, a study has been conducted on the model sensitivity to the hyperparameter ϵ in the context of water mapping and deforestation detection. The following subsection looks into the process of selecting this hyperparameter.

5.2. Hyperparameter Selection

The scarcity of labeled temporal data makes unsupervised strategies particularly appealing for the parameter selection task. One potential approach is to maximize the model evidence or marginal likelihood of the test data ([Barber, 2011](#)). While this method aligns with Bayesian principles, it requires further investigation beyond the scope of this study. Therefore, based on the insights gained from the sensitivity analysis in [Section 4.3](#), we propose the following guidelines for selecting ϵ .

Overall, the analysis shows that to obtain robust classification results, i.e., that are insensitive to undesired abrupt changes in the image, it is best to select small values of ϵ . In the absence of ground truth data, we propose to first assess the evolution of the number of pixels classified as a specific class over time for different ϵ values. For instance, in a water mapping experiment, tracking how the number of pixels classified as water changes over time under different ϵ values may be informative. The sensitivity of the scene to ϵ is best evaluated on dates with significant changes in class distribution, such as those caused by natural phenomena like draining events in preparation for extreme rainfall in the water mapping task. The selection process begins by determining the appropriate increment between tested ϵ values to observe changes in the algorithm behavior. Starting with an ϵ value close to 0 offers high robustness, and the value can be gradually increased in increments of 0.005. If variability is not observed in the curves showing

the number of water pixels over time, the increment can be doubled until changes are detected. From our experience, effective ϵ values typically range between 0 and 0.1, as these values are small enough to ensure robustness while still allowing for adaptability. Importantly, ϵ values greater than 0.5 should not be tested, as they suggest the typically unrealistic scenario where a pixel is more likely to transition between classes than remain in the same class. When there is a change in the scene, such as draining, the selected ϵ value should allow the model to reflect this change smoothly. This means ϵ should be sensitive enough to capture the change, yet robust enough to avoid erratic fluctuations, ensuring both adaptability and stability. If multiple ϵ values achieve this balance, the final choice can be made arbitrarily among them.

5.3. RBC framework limitations and challenges

In this subsection, we discuss the limitations of the proposed framework. These include the lack of ground truth data, which challenges quantitative performance evaluation; missing data due to cloud cover and other atmospheric disturbances; reliance on constant class transition and prior probabilities; and the omission of spatial correlations between classification results across different pixels.

The main limitation of this work is the scarcity of ground truth multitemporal classification maps, which complicates the quantitative assessment of results. For test sites 1a, 1b, and 2, the authors manually labeled a water mapping dataset with the LabelStudio tool. The labeled dataset has been shared with the remote sensing community to support researchers facing similar challenges (Calatrava et al., 2024). This labeling process is both time-consuming and resource-intensive, which limits the extension of the evaluation of the proposed framework under multiple geographic and environmental conditions. The availability of open-source labeled deforestation data in the MultiEarth challenge dataset (Cha et al., 2023) facilitated a quantitative analysis for test site 3. However, this dataset lacks deforestation labels for each acquisition date, allowing quantitative metrics to be computed for only a limited portion of the time series. Specifically, error classification maps were generated for just five images, indicated by filled blue markers in Fig. 8. Additionally, temporal misalignment between image and label dates in the MultiEarth dataset required comparing each label to the nearest classification result in time, which introduced temporal gaps.

Despite the limited availability of temporal datasets with ground truth, we evaluated our algorithm under a variety of settings to demonstrate its full capabilities, versatility, and potential limitations. We carefully selected three geographical regions: the Oroville Dam (California, USA), Charles River (Boston, USA), and the Amazon rainforest (Brazil), each presenting unique challenges (see Fig. 4). Cloud cover and shadows were particularly problematic in the Amazon rainforest, where we used CI_2 and CSI indices to assess cloud contamination (Zhai et al., 2018). On specific dates, such as 2020-06-10 and 2020-08-04, the presence of clouds caused interference, but RBC was able to improve classification accuracy. Significant differences in illumination were observed between the winter of 2020 and the fall of 2021 in the Oroville Dam region, especially at test site 1a. Additionally, natural events such as substantial fluctuations in water levels at Oroville Dam, particularly at test site 1b from April 2021 to September 2021, presented unforeseen changes that challenged the framework. The summer of 2021 also saw the emergence of seasonal algal blooms in the Charles River, creating some of the most challenging conditions.

RBC effectively handles temporary disruptions in remote sensing imagery and resolves missing data from cloud cover by using information from previous time steps. However, prolonged cloud cover and persistent disruptions can negatively impact classification performance. In general, extended cloud cover may hinder the overall efficiency of the classification process with RBC. During data preprocessing for test site 3, 182 of the 225 available images were discarded due to high cloud and cloud shadow percentages as indicated by the CI_2 and CSI indices, and an additional 12 images were filtered out through visual inspection. Although this level of cloud cover did not deny the performance of RBC and allowed us to demonstrate its capability to handle temporary disruptions, continuous cloud cover and a larger extent of missing data would need additional strategies. Temporal interpolation is a potential solution that can be employed to estimate absent values using data from other periods to ensure dataset continuity. However, a recent investigation (Che et al., 2024) that used linear interpolation to fill in missing values in time series data showed only marginal enhancements in classification accuracy when deep learning models were used. The authors emphasized the need for stronger methods to deal with persistent gaps in data, such as data fusion from multiple sensors and RNNs with masking capabilities. Another recent study uses the Whittaker smoothing to reduce noise and produce continuous NDVI time series using satellite data collected from multiple sources. This technique benefits the precision and spatial resolution of the reconstructed images (Liang et al., 2023). Consequently, further research is needed to enhance the

adaptability of RBC to diverse environments and develop improved strategies for managing larger levels of missing data.

Our methodology, which relies on constant transition probabilities, is computationally efficient, scalable across large geographic areas, and relatively easy to tune. However, this simplicity may come at the cost of accurately reflecting real-world dynamics, as natural systems and human activities often experience shifts that influence transition probabilities. Incorporating recursion could help automatically determine these transition probabilities over time, and a similar approach could be applied to class prior probabilities. Another limitation of our work is the absence of spatial correlation between classification results at different pixels. By assuming independence between pixel labels, the RBC framework fails to leverage the spatial structures and patterns that are prevalent in natural environments. Addressing these limitations could lead to meaningful advancements in future research.

6. Conclusion

In this paper, we have introduced the recursive Bayesian classifier (RBC), a framework that converts any instantaneous classifier into a robust online method through a probabilistic approach that is resilient to non-informative image variations. Using Sentinel-2 data, we have applied RBC to GMM, LR, and to our proposed SIC algorithm, which uses standard broadband spectral indices to generate predictive probabilities. The conducted experiments involve water mapping of the Oroville Dam in California, the Charles River basin in Massachusetts, and deforestation detection in the Amazon. The results demonstrate that RBC significantly improves the robustness of classifiers in multitemporal settings under challenging conditions. Specifically, in water mapping, RBC enhances balanced classification accuracy by up to 26.95% for SIC, 13.81% for LR, and 12.4% for GMM. In deforestation detection, the accuracy improvements are 15.25% for SIC, 14.17% for GMM, and 14.7% for LR. Additionally, without requiring additional training data, RBC improves the performance of state-of-the-art deep learning models, with DeepWaterMap showing a 9.62% accuracy increase and WatNet improving by 11.03% thanks to recursion. Despite these significant gains, RBC introduces a low computational cost, with minimal overhead that can be considered negligible when applied to more complex deep learning models. For instance, recursion time for recursive WatNet constitutes only 0.32% of the total algorithm time on test site 1a, 0.09% on test site 1b, and 1.39% on test site 2. Similarly, for recursive DeepWaterMap, the recursion overhead is 0.54% on test site 1a, 0.15% on test site 1b, and 4.32% on test site 2. This overhead, in addition to being small, remains constant for each time step regardless of the image time-series length due to the nature of Bayesian recursion, making the proposed methodology a suitable solution for real-time remote sensing applications. Future work will focus on methods for automatically determining class transition probabilities and addressing the issue of missing data caused by cloud cover and other disruptions in satellite imagery.

Declaration of Competing Interest

The authors of this paper declare that they have no known personal relationships or competing financial interests that could have appeared to influence this research.

Dataset and supplemental results

A Python implementation of the proposed algorithms can be found at <https://github.com/neu-spiral/RBC-SatImg>. The pre-processed data and the manually generated ground truth labels for the water mapping experiments are available at (Calatrava et al., 2024). Supplemental material containing additional experimental results is also available with this paper.

Acknowledgements

This work has been partially supported by the National Geographic Society under grant NGS-86713T-21, the National Science Foundation under Awards ECCS-1845833 and CCF-2326559, and the National Aeronautics and Space Administration under Award 80NSSC20K0742.

Appendix A Land cover classification experiment

Land cover classification results for test site 2 are presented in Fig. 16. Performance evaluation in this experiment is restricted to the visual inspection of the classification maps given the lack of ground truth data for the vegetation class in this test site. Results suggest that seasonal variations in the distribution of land, water and vegetation are captured well by the instantaneous classifiers and their recursive counterparts. The decrease in the amount of vegetation starting from November (through winter) with an increase in dry land (at dates 2020-12-13 and 2021-03-20) is represented by an increase in yellow pixels until May, followed by an increase in the number of vegetation pixels through summer and fall (from 2021-05-27 to 2021-09-14). The advantages of the RBC framework are evident, especially on the date 2020-12-13, where the instantaneous classifiers misclassify a substantial land area as water, while the recursive algorithms properly identify the land pixels. On 2021-09-14, both the SIC and LR algorithms fail to identify a section of the water body, yet their recursive counterparts successfully handle this task.

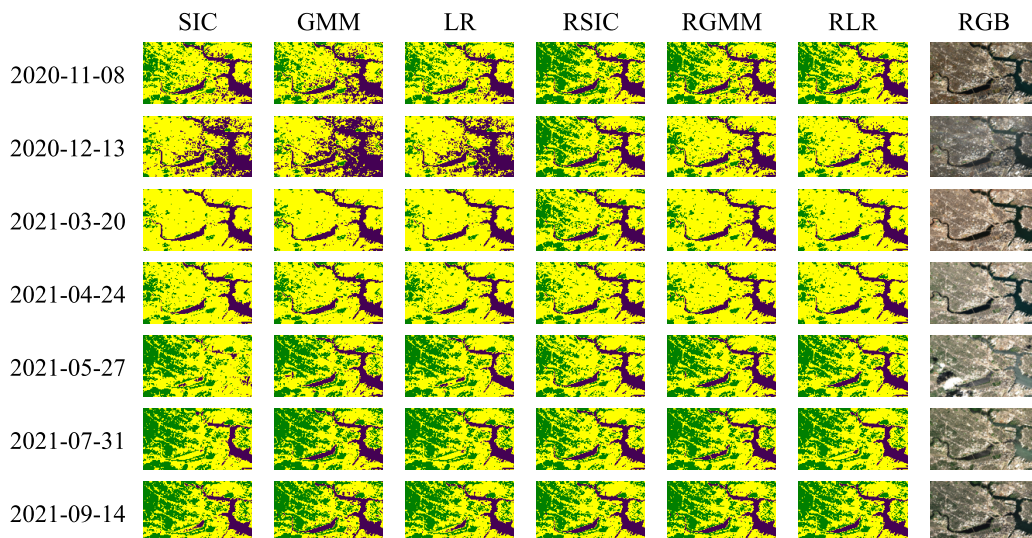


Figure 16: Land cover classification map results obtained for the Charles river basin area covered by test site 2. Purple, green and yellow represent water, vegetation and land, respectively. Images are arranged in chronological order. By visually inspecting these results, it can be assessed that the proposed RBC framework provides adaptability to seasonal changes and robustness to highly reflective surfaces and other disturbances.

Table 5: Parameter settings for the three-class land cover classification experiment with data from test site 2.

Test site 2	$C_t \in \mathcal{C}_{LC} = \{\text{water, land, vegetation}\}$	
	SIC	$\tau_{LC} = [-1, -0.05, 0.35, 1];$ $\mu_{LC} = [-0.525, 0.149, 0.675];$ $\sigma_{LC} = [0.475, 0.19, 0.325];$ $y_{NDVI}(z_t) = \frac{z_{t,NIR} - z_{t,red}}{z_{t,NIR} + z_{t,red}}$
	RSIC	$\epsilon = 0.05; \lambda = 0$
	RGMM	$\epsilon = 0.05; \lambda = 0$
	RLR	$\epsilon = 0.05; \lambda = 0$

References

- Acharya, T., Subedi, A., Lee, D., 2018. Evaluation of Water Indices for Surface Water Extraction in a Landsat 8 Scene of Nepal. *Sensors* 18, 2580. doi:[10.3390/s18082580](https://doi.org/10.3390/s18082580).
- Adugna, T., Xu, W., Fan, J., 2022. Comparison of random forest and support vector machine classifiers for regional land cover mapping using coarse resolution FY-3C images. *Remote Sensing* 14, 574.

- Bai, B., Tan, Y., Donchyts, G., Haag, A., Xu, B., Chen, G., Weerts, A.H., 2023. Naive Bayes classification-based surface water gap-filling from partially contaminated optical remote sensing image. *Journal of Hydrology* 616, 128791.
- Barber, D., 2011. *Bayesian Reasoning and Machine Learning*. 04-2011 ed., Cambridge University Press. URL: <http://www.cs.ucl.ac.uk/staff/d.barber/brml>. in press.
- Borsoi, R.A., Imbiriba, T., Bermudez, J.C.M., Richard, C., 2021a. Fast Unmixing and Change Detection in Multitemporal Hyperspectral Data. *IEEE Transactions on Computational Imaging* 7, 975–988.
- Borsoi, R.A., Imbiriba, T., Bermudez, J.C.M., Richard, C., Chanussot, J., Drumetz, L., Tournet, J.Y., Zare, A., Jutten, C., 2021b. Spectral Variability in Hyperspectral Data Unmixing: A Comprehensive Review. *IEEE Geoscience and Remote Sensing Magazine* 9, 223–270.
- Borsoi, R.A., Imbiriba, T., Closas, P., Bermudez, J.C.M., Richard, C., 2020. Kalman filtering and expectation maximization for multitemporal spectral unmixing. *IEEE Geoscience and Remote Sensing Letters* 19, 1–5.
- Cai, Y., Shi, Q., Liu, X., 2024. Spatiotemporal Mapping of Surface Water Using Landsat Images and Spectral Mixture Analysis on Google Earth Engine. *Journal of Remote Sensing* 4, 0117. URL: <https://spj.science.org/doi/abs/10.34133/remotesensing.0117>, doi:10.34133/remotesensing.0117, arXiv:<https://spj.science.org/doi/pdf/10.34133/remotesensing.0117>.
- Calatrava, H., Duvvuri, B., Li, H., Borsoi, R., Imbiriba, T., Beighley, E., Erdogmus, D., Closas, P., 2024. RBC-SatImg: Sentinel-2 Imagery and WatData Labels for Water Mapping. doi:10.5281/zenodo.13345343.
- Campbell, A., Shi, Y., Rainforth, T., Doucet, A., 2021. Online variational filtering and parameter learning, in: *Advances in Neural Information Processing Systems*. URL: <https://openreview.net/forum?id=et2st4Jqhc>.
- Cao, X., Yao, J., Xu, Z., Meng, D., 2020. Hyperspectral Image Classification With Convolutional Neural Network and Active Learning. *IEEE Transactions on Geoscience and Remote Sensing* 58, 4604–4616. doi:10.1109/TGRS.2020.2964627.
- Cavalli, R.M., 2023. Spatial validation of spectral unmixing results: A systematic review. *Remote Sensing* 15, 2822.
- Cha, M., Angelides, G., Hamilton, M., Soszynski, A., Swenson, B., Maidel, N., Isola, P., Perron, T., Freeman, B., 2023. MultiEarth 2023 – Multimodal Learning for Earth and Environment Workshop and Challenge. arXiv:2306.04738.
- Che, X., Zhang, H.K., Li, Z.B., Wang, Y., Sun, Q., Luo, D., Wang, H., 2024. Linearly interpolating missing values in time series helps little for land cover classification using recurrent or attention networks. *ISPRS Journal of Photogrammetry and Remote Sensing* 212, 73–95. URL: <https://www.sciencedirect.com/science/article/pii/S0924271624001813>, doi:<https://doi.org/10.1016/j.isprsjprs.2024.04.021>.
- Chouteau, F., Gabet, L., Fraisse, R., Bonfort, T., Harnoufi, B., Greiner, V., Le Goff, M., Ortner, M., Paveau, V., 2022. Joint Super-Resolution and Image Restoration for PLEIADES NEO Imagery. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 43B1, 9–15. doi:10.5194/isprs-archives-XLIII-B1-2022-9-2022.
- Chuvieco, E., 2020. *Fundamentals of satellite remote sensing: An environmental approach*. CRC press.
- Constantin, A., Fauvel, M., Girard, S., 2022. Joint Supervised Classification and Reconstruction of Irregularly Sampled Satellite Image Times Series. *IEEE Transactions on Geoscience and Remote Sensing* 60. doi:10.1109/TGRS.2021.3076667.
- Demirkaya, A., Imbiriba, T., Lockwood, K., Rampersad, S., Alhajjar, E., Guidoboni, G., Danziger, Z., Erdoğan, D., 2021. Cubature Kalman filter based training of hybrid differential equation recurrent neural network physiological dynamic models, in: *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, IEEE. pp. 763–766.

- Deng, Z., Zhu, X., He, Q., Tang, L., 2019. Land use/land cover classification using time series Landsat 8 images in a heavily urbanized area. *Advances in Space Research* 63. doi:[10.1016/j.asr.2018.12.005](https://doi.org/10.1016/j.asr.2018.12.005).
- Ellwanger, J.H., Kulmann-Leal, B., Kaminski, V.L., Valverde-Villegas, J., Veiga, A.B.G., Spilki, F.R., Fearnside, P.M., Caesar, L., Giatti, L.L., Wallau, G.L., et al., 2020. Beyond diversity loss and climate change: Impacts of Amazon deforestation on infectious diseases and public health. *Anais da Academia Brasileira de Ciências* 92, e20191375.
- Fang, Z., Wang, Y., Peng, L., Hong, H., 2021. Predicting flood susceptibility using LSTM neural networks. *Journal of Hydrology* 594. doi:[10.1016/j.jhydrol.2020.125734](https://doi.org/10.1016/j.jhydrol.2020.125734).
- Franceschi, J.Y., Delasalles, E., Chen, M., Lamprier, S., Gallinari, P., 2020. Stochastic Latent Residual Video Prediction, in: *Proceedings of the 37th International Conference on Machine Learning*, PMLR. pp. 3233–3246. URL: <https://proceedings.mlr.press/v119/franceschi20a.html>.
- Gao, H., Wang, L., Jing, L., Xu, J., 2016. An effective modified water extraction method for Landsat-8 OLI imagery of mountainous plateau regions. *IOP Conference Series: Earth and Environmental Science* 34, 012010. doi:[10.1088/1755-1315/34/1/012010](https://doi.org/10.1088/1755-1315/34/1/012010).
- Ghezelbash, R., Maghsoudi, A., Shamekhi, M., Pradhan, B., Daviran, M., 2023. Genetic algorithm to optimize the SVM and K-means algorithms for mapping of mineral prospectivity. *Neural Computing and Applications* 35, 719–733.
- Hedhli, I., Moser, G., Serpico, S., Zerubia, J., 2021. Hierarchical Markov random fields for high resolution land cover classification of multisensor and multiresolution image time series.
- Hoberg, T., Rottensteiner, F., Feitosa, R.Q., Heipke, C., 2015. Conditional Random Fields for Multi-temporal and Multiscale Classification of Optical Satellite Imagery. *IEEE Transactions on Geoscience and Remote Sensing* 53, 659–673. doi:[10.1109/TGRS.2014.2326886](https://doi.org/10.1109/TGRS.2014.2326886).
- Hu, Q., Yin, H., Friedl, M.A., You, L., Li, Z., Tang, H., Wu, W., 2021. Integrating coarse-resolution images and agricultural statistics to generate sub-pixel crop type maps and reconciled area estimates. *Remote Sensing of Environment* 258, 112365. URL: <https://www.sciencedirect.com/science/article/pii/S0034425721000833>, doi:<https://doi.org/10.1016/j.rse.2021.112365>.
- Huang, H., Roy, D.P., 2021. Characterization of Planetscope-0 Planetscope-1 surface reflectance and normalized difference vegetation index continuity. *Science of Remote Sensing* 3, 100014. doi:[10.1016/j.srs.2021.100014](https://doi.org/10.1016/j.srs.2021.100014).
- Huang, S., Tang, L., Hupy, J.P., Wang, Y., Shao, G., 2021. A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing. *Journal of Forestry Research* 32, 1–6.
- Imbiriba, T., Closas, P., 2020. Enhancing particle filtering using Gaussian processes, in: *2020 IEEE 23rd International Conference on Information Fusion (FUSION)*, IEEE. pp. 1–7.
- Imbiriba, T., Demirkaya, A., Dunik, J., Straka, O., Erdoğan, D., Closas, P., 2022. Hybrid neural network augmented physics-based models for nonlinear filtering, in: *2022 25th International Conference on Information Fusion (FUSION)*, IEEE. pp. 1–6.
- Isikdogan, F., Bovik, A.C., Passalacqua, P., 2017. Surface Water Mapping by Deep Learning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 10, 4909–4918. doi:[10.1109/JSTARS.2017.2735443](https://doi.org/10.1109/JSTARS.2017.2735443).
- Jacobsen, K., 2022. Mapping with SkySat images. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLIII-B1-2022*, 99–106. doi:[10.5194/isprs-archives-XLIII-B1-2022-99-2022](https://doi.org/10.5194/isprs-archives-XLIII-B1-2022-99-2022).
- Jaynes, E.T., 1957. Information Theory and Statistical Mechanics. *Phys. Rev.* 106, 620–630. URL: <https://link.aps.org/doi/10.1103/PhysRev.106.620>, doi:[10.1103/PhysRev.106.620](https://doi.org/10.1103/PhysRev.106.620).
- Ji, R., Liang, Y., Xu, L., 2022. Recursive Bayesian inference and learning for target tracking with unknown maneuvers. *International Journal of Adaptive Control and Signal Processing* 36, 1032 – 1044.

- Ji, S., Zhang, C., Xu, A., Shi, Y., Duan, Y., 2018. 3D convolutional neural networks for crop classification with multi-temporal remote sensing images. *Remote Sensing* 10, 75.
- Karaman, K., Sainte Fare Garnot, V., Wegner, J.D., 2023. Deforestation Detection in the Amazon with Sentinel-1 SAR Image Time Series. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences X-1/W1-2023*, 835–842. URL: <https://isprs-annals.copernicus.org/articles/X-1-W1-2023/835/2023/>, doi:10.5194/isprs-annals-X-1-W1-2023-835-2023.
- Karmakar, P., Teng, S.W., Murshed, M., Pang, S., Li, Y., Lin, H., 2023. Crop monitoring by multimodal remote sensing: A review. *Remote Sensing Applications: Society and Environment* , 101093.
- Kenduiwo, B.K., Bargiel, D., Soergel, U., 2017. Higher Order Dynamic Conditional Random Fields Ensemble for Crop Type Classification in Radar Images. *IEEE Transactions on Geoscience and Remote Sensing* 55, 4638–4654. doi:10.1109/TGRS.2017.2695326.
- Khalid, H.W., Khalil, R.M.Z., Qureshi, M.A., 2021. Evaluating spectral indices for water bodies extraction in Western Tibetan Plateau. *The Egyptian Journal of Remote Sensing and Space Science* 24, 619–634. URL: <https://www.sciencedirect.com/science/article/pii/S1110982321000703>, doi:<https://doi.org/10.1016/j.ejrs.2021.09.003>.
- Ko Oo, T., Arunrat, N., Sereenonchai, S., Ussawarujikulchai, A., Chareonwong, U., Nutmagul, W., 2022. Comparing four machine learning algorithms for land cover classification in gold mining: A case study of kyaukpahto gold mine, northern myanmar. *Sustainability* 14, 10754. doi:10.3390/su141710754.
- Kok, Z.H., Shariff, A.R.M., Alfatni, M.S.M., Khairunniza-Bejo, S., 2021. Support vector machine in precision agriculture: A review. *Computers and Electronics in Agriculture* 191, 106546.
- Lapola, D.M., Pinho, P., Barlow, J., Aragão, L.E., Berenguer, E., Carmenta, R., Liddy, H.M., Seixas, H., Silva, C.V., Silva-Junior, C.H., et al., 2023. The drivers and impacts of Amazon forest degradation. *Science* 379, eabp8622.
- Li, H., Duvvuri, B., Borsoi, R., Imbiriba, T., Beighley, E., Erdoğan, D., Closas, P., 2022. Online multi-resolution fusion of space-borne multispectral images, in: *2022 IEEE Aerospace Conference (AERO)*, IEEE. pp. 1–12.
- Li, Q., Zhang, J., Jiang, H., 2024. Incorporating multi-source remote sensing in the detection of earthquake-damaged buildings based on logistic regression modeling. *Heliyon* .
- Liang, J., Ren, C., Li, Y., Yue, W., Wei, Z., Song, X., Zhang, X., Yin, A., Lin, X., 2023. Using enhanced gap-filling and whittaker smoothing to reconstruct high spatiotemporal resolution NDVI time series based on Landsat 8, Sentinel-2, and MODIS imagery. *ISPRS International Journal of Geo-Information* 12, 214.
- Likó, S.B., Holb, I.J., Oláh, V., Burai, P., Szabó, S., 2024. Deep learning-based training data augmentation combined with post-classification improves the classification accuracy for dominant and scattered invasive forest tree species. *Remote Sensing in Ecology and Conservation* 10, 203–219.
- Liu, Z., Xu, J., Liu, M., Yin, Z., Liu, X., Yin, L., Zheng, W., 2023. Remote sensing and geostatistics in urban water-resource monitoring: A review. *Marine and Freshwater Research* 74, 747–765.
- Luo, X., Tong, X., Hu, Z., 2021. An applicable and automatic method for earth surface water mapping based on multispectral images. *International Journal of Applied Earth Observation and Geoinformation* 103, 102472.
- Maciej Huk, Marcin Maleszka, E.S., 2020. *Intelligent Information and Database Systems: Recent Developments*. Springer International Publishing. URL: <http://dx.doi.org/10.1007/978-3-030-14132-5>, doi:10.1007/978-3-030-14132-5.
- Martinez, J.A.C., da Costa, G.A.O.P., Messias, C.G., de Souza Soler, L., de Almeida, C.A., Feitosa, R.Q., 2024. Enhancing deforestation monitoring in the Brazilian Amazon: A semi-automatic approach leveraging uncertainty estimation. *ISPRS Journal of Photogrammetry and Remote Sensing* 210, 110–127. URL: <https://www.sciencedirect.com/science/article/pii/S0924271624000765>, doi:<https://doi.org/10.1016/j.isprsjprs.2024.03.004>.

- Martín-Doñas, J., Jensen, J., Tan, Z.H., Gomez, A., Peinado, A., 2020. Online Multichannel Speech Enhancement Based on Recursive EM and DNN-Based Speech Presence Estimation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 28, 3080–3094. doi:[10.1109/TASLP.2020.3036776](https://doi.org/10.1109/TASLP.2020.3036776).
- Mashala, M.J., Dube, T., Mudereri, B.T., Ayisi, K.K., Ramudzuli, M.R., 2023. A systematic review on advancements in remote sensing for assessing and monitoring land use and land cover changes impacts on surface water resources in semi-arid tropical environments. *Remote Sensing* 15, 3926.
- Miller, L., Pelletier, C., Webb, G.I., 2024. Deep Learning for Satellite Image Time-Series Analysis: A review. *IEEE Geoscience and Remote Sensing Magazine* , 2–45doi:[10.1109/MGRS.2024.3393010](https://doi.org/10.1109/MGRS.2024.3393010).
- Miura, T., Tokumoto, Y., Shin, N., Shimizu, K.K., Pungga, R.A.S., Ichie, T., 2023. Utility of commercial high-resolution satellite imagery for monitoring general flowering in Sarawak, Borneo. *Ecological Research* 38, 386–402.
- Mountrakis, G., Heydari, S.S., 2023. Harvesting the Landsat archive for land cover land use classification using deep neural networks: Comparison with traditional classifiers and multi-sensor benefits. *ISPRS Journal of Photogrammetry and Remote Sensing* 200, 106–119. doi:[10.1016/j.isprsjprs.2023.05.005](https://doi.org/10.1016/j.isprsjprs.2023.05.005).
- Pastorino, M., Moser, G., Serpico, S.B., Zerubia, J., 2024. Multimission, Multifrequency, and Multiresolution SAR Image Classification Through Hierarchical Markov Models and Convolutional Networks. *IEEE Geoscience and Remote Sensing Letters* 21, 1–5. doi:[10.1109/LGRS.2024.3414284](https://doi.org/10.1109/LGRS.2024.3414284).
- Pelletier, C., Valero, S., Inglada, J., Champion, N., Sicre, C.M., Dedieu, G., 2017. Effect of training class label noise on classification performances for land cover mapping with satellite image time series. *Remote Sensing* 9. doi:[10.3390/rs9020173](https://doi.org/10.3390/rs9020173).
- Pelletier, C., Webb, G.I., Petitjean, F., 2019. Temporal convolutional neural network for the classification of satellite image time series. *Remote Sensing* 11. doi:[10.3390/rs11050523](https://doi.org/10.3390/rs11050523), [arXiv:1811.10166](https://arxiv.org/abs/1811.10166).
- Qiu, C., Mou, L., Schmitt, M., Zhu, X.X., 2019. Local climate zone-based urban land cover classification from multi-seasonal Sentinel-2 images with a recurrent residual network. *ISPRS Journal of Photogrammetry and Remote Sensing* 154, 151–162. URL: <https://www.sciencedirect.com/science/article/pii/S0924271619301261>, doi:<https://doi.org/10.1016/j.isprsjprs.2019.05.004>.
- Rome, M., Beighley, R.E., Faber, T., 2021. Sensor-based detection of algal blooms for public health advisories and long-term monitoring. *Science of The Total Environment* 767, 144984.
- Rußwurm, M., Korner, M., 2017. Temporal Vegetation Modelling Using Long Short-Term Memory Networks for Crop Identification from Medium-Resolution Multi-spectral Satellite Images, in: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, IEEE Computer Society. pp. 1496–1504. doi:[10.1109/CVPRW.2017.193](https://doi.org/10.1109/CVPRW.2017.193).
- Särkkä, S., Svensson, L., 2023. Bayesian filtering and smoothing. volume 17. Cambridge university press.
- Satir, O., Berberoglu, S., 2012. Land Use/Cover Classification Techniques Using Optical Remotely Sensed Data in Landscape Planning. *INTECH*. pp. 21–54. doi:[10.5772/2761](https://doi.org/10.5772/2761).
- Sharma, A., Liu, X., Yang, X., 2018. Land cover classification from multi-temporal, multi-spectral remotely sensed imagery using patch-based recurrent neural networks. *Neural Networks* 105, 346–355.
- Shi, S., Zhong, Y., Zhao, J., Lv, P., Liu, Y., Zhang, L., 2022. Land-Use/Land-Cover Change Detection Based on Class-Prior Object-Oriented Conditional Random Field Framework for High Spatial Resolution Remote Sensing Imagery. *IEEE Transactions on Geoscience and Remote Sensing* 60, 1–16. doi:[10.1109/TGRS.2020.3034373](https://doi.org/10.1109/TGRS.2020.3034373).
- Silva-Junior, C.H., Silva, F.B., Arisi, B.M., Mataveli, G., Pessôa, A.C., Carvalho, N.S., Reis, J.B., Silva Júnior, A.R., Motta, N.A., e Silva, P.V.M., et al., 2023. Brazilian Amazon Indigenous territories under deforestation pressure. *Scientific Reports* 13, 5851.
- Strahler, A.H., 1980. The use of prior probabilities in maximum likelihood classification of remotely sensed data. *Remote Sensing of Environment* 10, 135–163. URL: <https://www.sciencedirect.com/science/article/pii/0034425780900115>, doi:[https://doi.org/10.1016/0034-4257\(80\)90011-5](https://doi.org/10.1016/0034-4257(80)90011-5).

- Sun, C., Fagherazzi, S., Liu, Y., 2018. Classification mapping of salt marsh vegetation by flexible monthly NDVI time-series using Landsat imagery. *Estuarine, Coastal and Shelf Science* 213, 61–80.
- Sun, W., Chen, C., Liu, W., Yang, G., Meng, X., Wang, L., Ren, K., 2023. Coastline extraction using remote sensing: A review. *GIScience & Remote Sensing* 60, 2243671.
- Swain, P.H., 1978. Bayesian Classification in a Time-varying Environment. Technical Report. Purdue University.
- Tariq, A., Jiango, Y., Li, Q., Gao, J., Lu, L., Soufan, W., Almutairi, K.F., Habib-ur Rahman, M., 2023. Modelling, mapping and monitoring of forest cover changes, using support vector machine, kernel logistic regression and naive Bayes tree models with optical remote sensing data. *Heliyon* 9.
- Tran, T.V., Reef, R., Zhu, X., 2022. A review of spectral indices for mangrove remote sensing. *Remote Sensing* 14, 4868.
- Uslu, F., Bharath, A.A., 2019. A recursive Bayesian approach to describe retinal vasculature geometry. *Pattern Recognition* 87, 157–169. URL: <https://www.sciencedirect.com/science/article/pii/S0031320318303625>, doi:<https://doi.org/10.1016/j.patcog.2018.10.017>.
- Wang, B., Chen, Y., Yan, Z., Liu, W., 2024. Integrating remote sensing data and cnn-lstm-attention techniques for improved forest stock volume estimation: A comprehensive analysis of baishanzu forest park, china. *Remote Sensing* 16. URL: <https://www.mdpi.com/2072-4292/16/2/324>, doi:[10.3390/rs16020324](https://doi.org/10.3390/rs16020324).
- Wang, J., Bretz, M., Dewan, M.A.A., Delavar, M.A., 2022. Machine learning in modelling land-use and land cover-change (LULCC): Current status, challenges and prospects. *Science of The Total Environment* 822, 153559.
- Wang, Q., Zhang, C., Atkinson, P.M., 2020. Sub-pixel mapping with point constraints. *Remote Sensing of Environment* 244, 111817. URL: <https://www.sciencedirect.com/science/article/pii/S0034425720301875>, doi:<https://doi.org/10.1016/j.rse.2020.111817>.
- Wang, Y., Sun, Y., Cao, X., Wang, Y., Zhang, W., Cheng, X., 2023. A review of regional and Global scale Land Use/Land Cover (LULC) mapping products generated from satellite remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing* 206, 311–334.
- Wasehun, E.T., Hashemi Beni, L., Di Vittorio, C.A., 2024. UAV and satellite remote sensing for inland water quality assessments: a literature review. *Environmental Monitoring and Assessment* 196, 277.
- Wei, P., Ye, H., Qiao, S., Liu, R., Nie, C., Zhang, B., Song, L., Huang, S., 2023. Early Crop Mapping Based on Sentinel-2 Time-Series Data and the Random Forest Algorithm. *Remote Sensing* 15. URL: <https://www.mdpi.com/2072-4292/15/13/3212>, doi:[10.3390/rs15133212](https://doi.org/10.3390/rs15133212).
- Wenger, R., Puissant, A., Weber, J., Idoumghar, L., Forestier, G., 2022. MultiSenGE: A multimodal and multitemporal benchmark dataset for land use/land cover remote sensing applications. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences V-3-2022*, 635–640. doi:[10.5194/isprs-annals-v-3-2022-635-2022](https://doi.org/10.5194/isprs-annals-v-3-2022-635-2022).
- Wu, P., Imbiriba, T., LaMountain, G., Vilá-Valls, J., Closas, P., 2019. Wifi fingerprinting and tracking using neural networks, in: *Proceedings of the 32nd International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2019)*, pp. 2314–2324.
- Xu, Y., Yang, Y., Chen, X., Liu, Y., 2022. Bibliometric Analysis of Global NDVI Research Trends from 1985 to 2021. *Remote Sensing* 14. URL: <https://www.mdpi.com/2072-4292/14/16/3967>, doi:[10.3390/rs14163967](https://doi.org/10.3390/rs14163967).
- Yang, H., Kong, J., Hu, H., Du, Y., Gao, M., Chen, F., 2022. A review of remote sensing for water quality retrieval: progress and challenges. *Remote Sensing* 14, 1770.
- Yang, X., Qin, Q., Grussenmeyer, P., Koehl, M., 2018. Urban surface water body detection with suppressed built-up noise based on water indices from Sentinel-2 MSI imagery. *Remote sensing of environment* 219, 259–270.

- Yaseen, Z.M., 2024. Flood hazards and susceptibility detection for Ganga river, Bihar state, India: Employment of remote sensing and statistical approaches. *Results in Engineering* 21, 101665.
- Yu, D., Fang, C., 2023. Urban remote sensing with spatial big data: a review and renewed perspective of urban studies in recent decades. *Remote Sensing* 15, 1307.
- Zanaga, D., Van De Kerchove, R., De Keersmaecker, W., Souverijns, N., Brockmann, C., Quast, R., Wevers, J., Grosu, A., Paccini, A., Vergnaud, S., Cartus, O., Santoro, M., Fritz, S., Georgieva, I., Lesiv, M., Carter, S., Herold, M., Li, L., Tsendbazar, N.E., Ramoino, F., Arino, O., 2021. ESA WorldCover 10 m 2020 v100.
- Zeb, A., Armstrong, G.W., Hamann, A., 2019. Forest conversion by the Indigenous Kalasha of Pakistan: A household level analysis of socioeconomic drivers. *Global Environmental Change* 59, 102004.
- Zhai, H., Zhang, H., Zhang, L., Li, P., 2018. Cloud/Shadow Detection Framework based on Spectral Indices for Multi/Hyperspectral Optical Remote Sensing Imagery. *ISPRS Journal of Photogrammetry and Remote Sensing* 144, 235 – 253. doi:[10.1016/j.isprsjprs.2018.07.006](https://doi.org/10.1016/j.isprsjprs.2018.07.006).
- Zhang, C., Liu, Y., Tie, N., 2023. Forest Land Resource Information Acquisition with Sentinel-2 Image Utilizing Support Vector Machine, K-Nearest Neighbor, Random Forest, Decision Trees and Multi-Layer Perceptron. *Forests* 14, 254.
- Zhang, Z., Wang, L., Chen, Y., Zheng, C., 2024. Crop Identification of UAV Images Based on an Unsupervised Semantic Segmentation Method. *IEEE Geoscience and Remote Sensing Letters* .
- Zhou, Y., Huang, C., Wu, T., Zhang, M., 2023. A novel spatio-temporal cellular automata model coupling partitioning with CNN-LSTM to urban land change simulation. *Ecological Modelling* 482, 110394. URL: <https://www.sciencedirect.com/science/article/pii/S0304380023001254>, doi:<https://doi.org/10.1016/j.ecolmodel.2023.110394>.
- Zhu, Z., Qiu, S., Ye, S., 2022. Remote sensing of land change: A multifaceted perspective. *Remote Sensing of Environment* 282, 113266. doi:[10.1016/j.rse.2022.113266](https://doi.org/10.1016/j.rse.2022.113266).

Supplemental Results

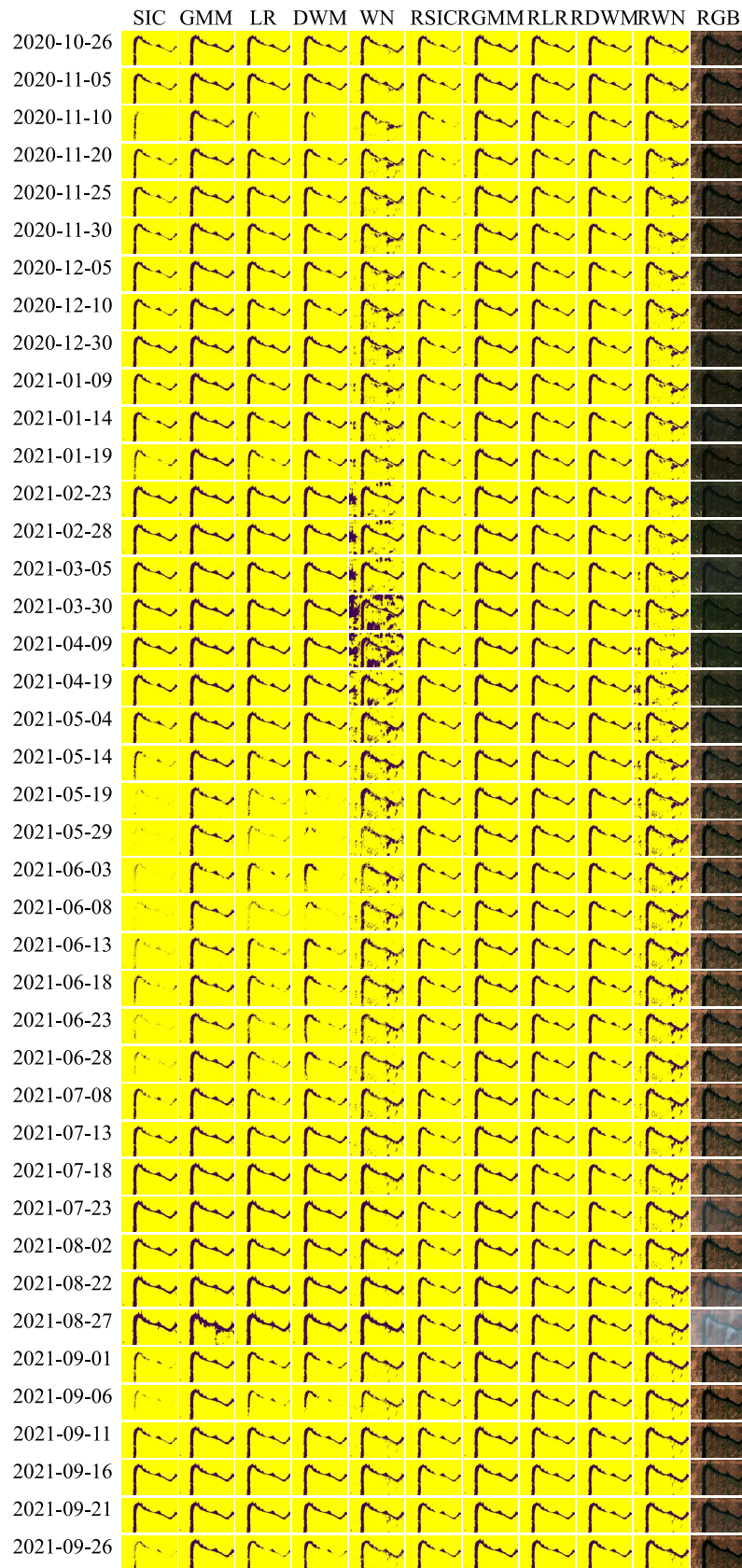


Figure 17: Supplemental results from test site 1a including classification maps for all dates (from 2020-10-26 to 2021-09-26).



Figure 18: Supplemental results from test site 1b including classification maps for all dates (from 2020-10-26 to 2021-09-26).

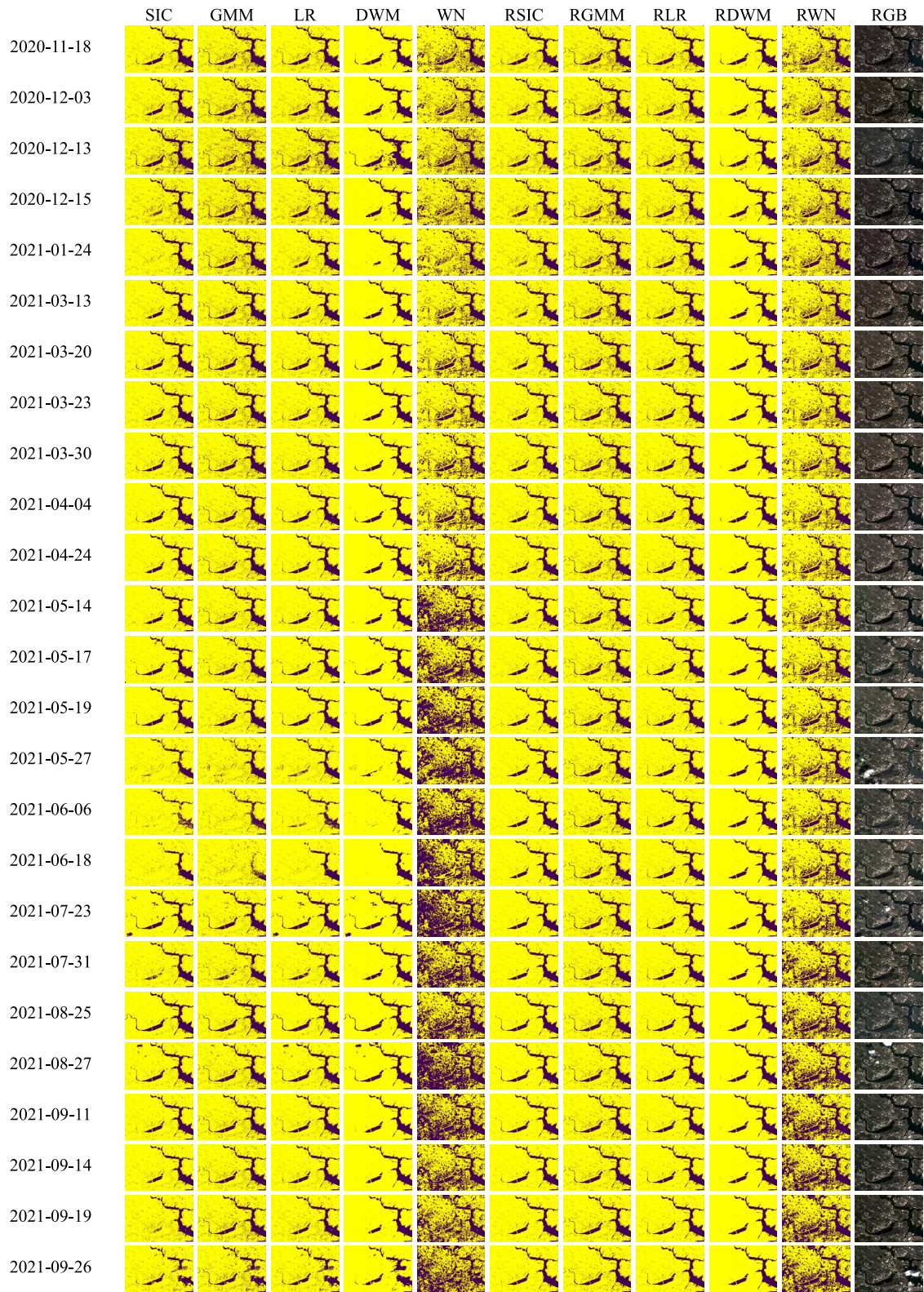


Figure 19: Supplemental results from test site 2 including classification maps for all dates (from 2020-11-18 to 2021-09-26).

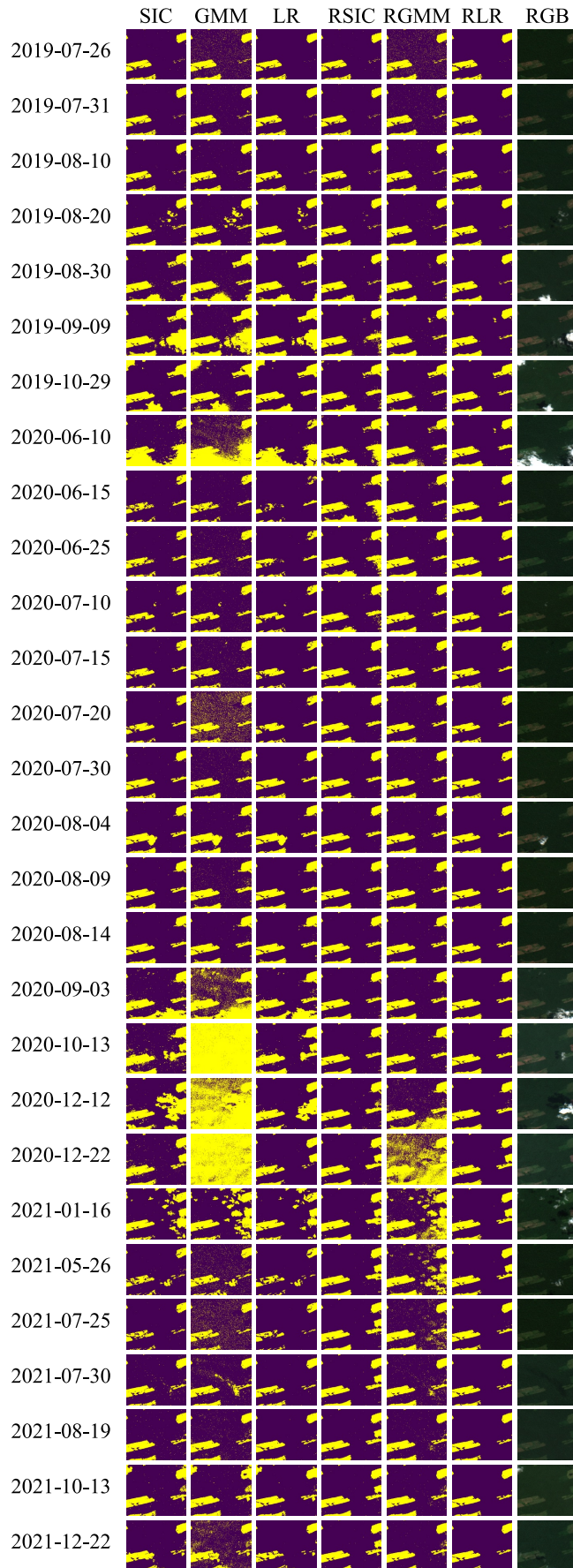


Figure 20: Supplemental results from test site 3 including classification maps for all dates (from 2019-07-26 to 2021-12-22).