Algorithm Theoretical Basis Document (ATBD)

MapBiomas Fire

Collection 4

Version 1

General coordinator

Ane A. Alencar

Team

Dhemerson E. Conciani

Eduardo Reis Rosa

Eduardo Vélez Martin

Gilvan Andrade

Heinrich Hasenack

Luiz Felipe Morais Martenexen

João Paulo Fernandes Márcio Ribeiro

Julia Shimbo

Marcos Rosa

Mariana Dias

Natalia Crusco

Nerivaldo Santos

Newton Coelho Monteiro

Soltan Galano Duverger

Tasso Azevedo

Valderli J. Piontekowski

Vera Laísa da Silva Arruda

Wallace Vieira da Silva

Washington da Franca Rocha

Table of Contents

1.Introduction

1.1. Overview of the MapBiomas Fire

The objective of this document is to describe the theoretical basis, justification, and methodology used to produce the monthly and annual maps of burned areas in Brazil from 1985 to 2024 for the MapBiomas Fire Collection 4. This Algorithm Theoretical Basis Document (ATBD) aims to provide the methodological steps to produce MapBiomas Collection 4 and describe the datasets and analysis. All MapBiomas maps and datasets are freely available on the project platform (http://mapbiomas.org)

The MapBiomas Fire project released its Collection 1 of annual maps of fire scars covering the period from 1985 to 2020 in August 2021. Collection 2 encompassed the years from 1985 to 2022. Collection 3 covered the period 1985 to 2023. The current MapBiomas Fire Collection 4 spans the years 1985 to 2024, with monthly and annual data on burned areas covering the entire Brazilian territory. These maps are based on annual mosaics of images from the Landsat satellites with a spatial resolution of 30 meters.

The entire process was carried out collaboratively by institutions within the MapBiomas network, utilizing machine learning algorithms (deep learning) through the Google Earth Engine and Google Cloud Platforms, which offer immense processing capacity in the cloud, as well as local servers for additional processing.

The classification was organized by biomes and regions, collecting samples of burned and unburned areas for training the algorithm by regions, and using reference maps, such as MODIS Burned Area (MCD64A1 - https://lpdaac.usgs.gov/products/mcd64a1v006/) with 500 m spatial resolution, GABAM (Global Annual Burned Area Map - https://gee-community-catalog.org/projects/gabam/) with 30 m resolution, fire hotspots, and INPE fire scars (https://terrabrasilis.dpi.inpe.br/queimadas/bdqueimadas/).

The products of the MapBiomas Fire Collection 4 include:

- Monthly and annual burned area maps in Brazil from 1985 to 2024;
- Frequency of annual burned areas in Brazil;
- Accumulated burned areas in Brazil;
- Burned areas over Land Use and Land Cover classes of MapBiomas Collection 9;
- Annual burned area by fire scar size interval;
- Year of the last fire occurrence.

The classification algorithms are available on the MapBiomas GitHub (https://github.com/mapbiomas-brazil/fire).

1.2. How we are organized

MapBiomas is a multi-institutional initiative of the Climate Observatory (a network of NGOs working on climate change in Brazil - http://www.observatoriodoclima.eco.br/en/). The co-creators of MapBiomas include NGOs, universities, and technology companies. For the MapBiomas Fire project, the Amazon Environmental Research Institute (IPAM) led the technological and operational development. The geospatial technology company Ecostage is responsible for the backend, dashboard, website, and frontend development of MapBiomas. Expert teams in each biome carried out sampling, evaluation, and refinement of the mapping, as shown in the box below.

Biome coordination:

- **Amazon** – Amazon Environmental Research Institute (IPAM)
- **Atlantic Forest** – SOS Atlantic Forest Foundation and ArcPlan
- **Caatinga** – Geodatin
- **Cerrado** – Amazon Environmental Research Institute (IPAM)
- **Pampa** – Federal University of Rio Grande do Sul (UFRGS) and GeoKarten
- **Pantanal** – SOS Pantanal Institute and ArcPlan

1.3. Historical Perspective: Existent Maps and Mapping Initiatives:

There are few global products that map large-scale burned areas at high temporal resolution (e.g., twice a day), such as the MODIS (Moderate Resolution Imaging Spectroradiometer) based product MCD64A1 Collection 6, with a 500 m pixel resolution provided by the National Aeronautics and Space Administration (NASA). We used the MCD64A1 Burned Area Product as a reference data for burned areas (Giglio et al., 2016).

Additionally, we used the fire hotspots products developed by the National Institute for Space Research (INPE) in Brazil. The INPE fire hotspot product is based on an automatic mapping approach using 1 km x 1 km pixel size and thermal bands of nine satellites, with the AQUA_M-T (Sensor MODIS) as a reference satellite. This product provides daily data of fire hotspots since 2000 and is available at INPE Queimadas (https://terrabrasilis.dpi.inpe.br/queimadas/bdqueimadas/).

We also used the 30-m resolution Global Annual Burned Area Map (GABAM), which defines the spatial extent of fires that occur within a whole year. GABAM is generated via an automated global burned area mapping approach based on all available Landsat images on the Google Earth Engine (GEE) platform (Long, 2019).

These existing mapping initiatives provide essential reference data that support the accuracy and reliability of the MapBiomas Fire project, contributing to a comprehensive understanding of burned area dynamics in Brazil.

2. Methodological description

We used all available Landsat imagery (Landsat 5, 7, 8, and 9) and a Deep Neural Network (DNN) model to detect and map burned areas within the Brazilian biomes from January 1985 to December 2024. The DNN models utilize artificial intelligence and machine learning algorithms to perform deep learning classifications of complex phenomena, resulting in higher performance outcomes, including for fire mapping (Langford, 2018).

The images were processed in Google Earth Engine (GEE) to create annual Landsat quality mosaics, which were used to collect burned and unburned spectral signatures serving as training samples for the classification model. The training samples and annual quality mosaics were exported to a Google Cloud Storage Bucket to be used as input in virtual machines. These were then used to train the DNN models, process the burn scar mapping, and produce a dataset of 40 years of monthly burned area data for all of Brazil from 1985 to 2024.

The image processing and classification routines used to map the monthly burned areas in the Brazilian territory followed six steps:

1. Definition of the classification regions per biome: Biomes were divided into regions to facilitate more accurate classification.
2. Construction of annual Landsat quality mosaics: High-quality annual mosaics were generated from Landsat images to provide the dataset for classification.
3. Collection of training samples: Spectral signatures of burned and unburned areas were collected from the annual quality mosaics to serve as training samples.
4. Training and development of the DNN prediction model: The DNN model was trained using the collected samples and annual mosaics.
5. Use of post-classification routines: Masks and spatial filters were applied to improve the accuracy and reduce noise in the classification results.
6. Validation with reference data and visual checks: The classification results were validated using reference data, along with visual checks of burn scars to ensure their accuracy.

Our approach combined the robust capabilities of deep learning with comprehensive satellite data, enabling the creation of a detailed and reliable burned area map across different biomes in Brazil, illustrated in Figure 1.

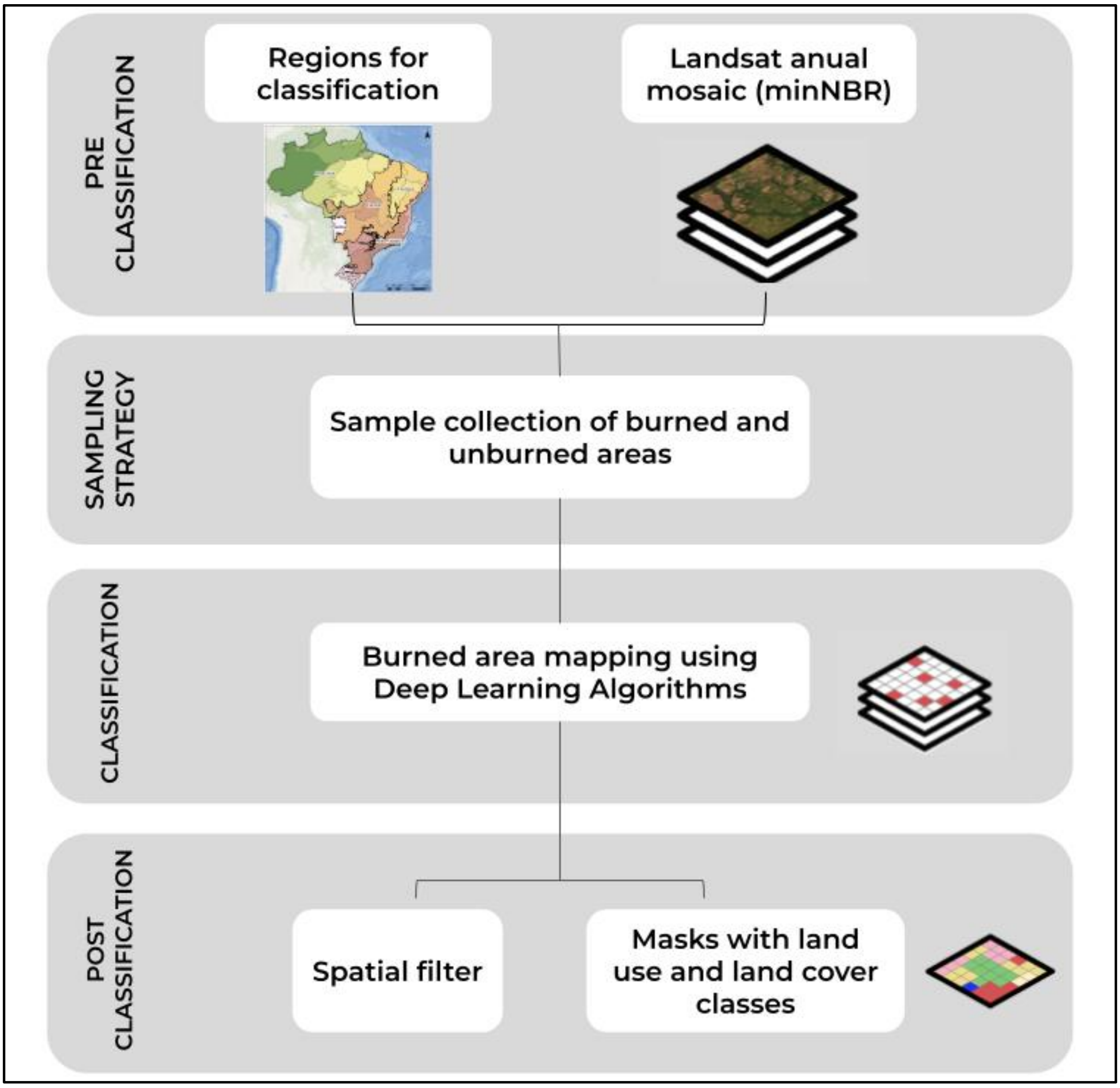

Figure 1. Overview of the method for classifying burned areas in Brazil in MapBiomas Fire Collection 4.

2.1. Definition of regions by biome

Considering that fire regimes and burned area spectral signatures are influenced by climatic conditions as well as land cover and land use types, we combined edaphoclimatic and morphoclimatic data with annual maps of land cover and land use from MapBiomas Collection 9 to segment each biome into classification regions (Figure 2). This process resulted in 28 classification regions, addressing regional patterns and providing a more accurate classification of burned areas.

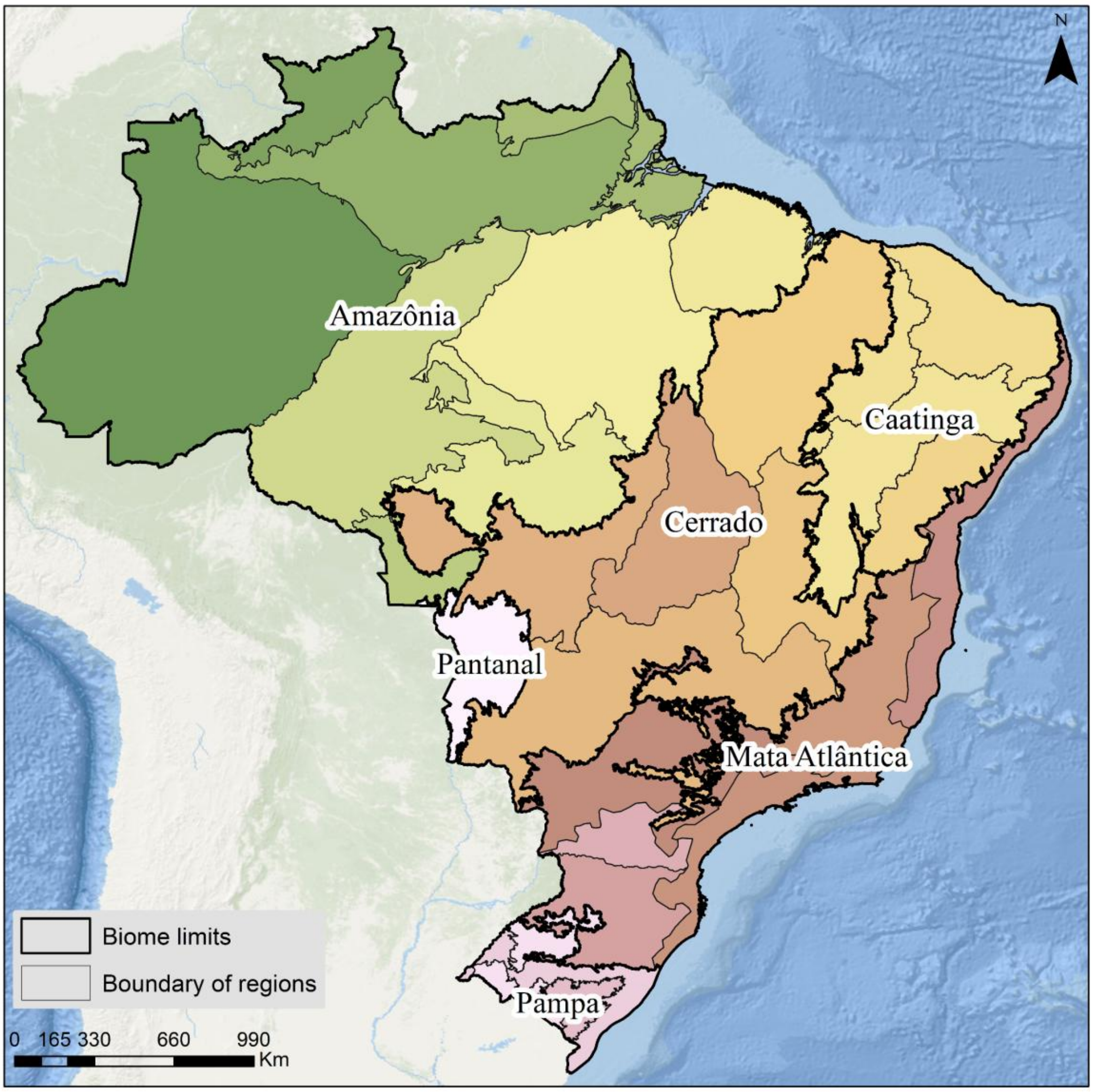

Figure 2. Regions defined for each biome in Brazil to collect training samples and classify burned areas in the MapBiomas Fire Collection 2.

2.2. Annual mosaics

The classification was performed using surface reflectance (SR) USGS Landsat Collection 2 (Tier 1) mosaics (30 m × 30 m) constructed for each year from 1985 to 2023. We assessed all the available Landsat 5 (from 1985 to 2011), Landsat 7 (1999 to 2021), Landsat 8 (2013 to 2023), and Landsat 9 scenes (2022 to 2024) with a 16-day return interval.

Landsat Surface Reflectance is accompanied by two Bitwise Quality Assessment bands (QA_PIXEL and QA_RADSAT) that indicate the pixels with radiometric and instrument-related problems, including a probability flag. We used the QA_PIXEL band to select and mask the pixels with high confidence levels (67–100%) of ‘cloud’ and ‘shadow’. Then, we used the QA_RADSAT to avoid pixels with radiometric saturation in any surface reflectance band. Finally, we discarded pixels with negative values in the surface reflectance in order to eliminate anomalies and noises in annual quality mosaic composition.

We used a per-year statistical approach to summarize this amount of data and optimize the classification without discarding spectral information on a pixel basis. This approach allowed us to create yearly mosaics by performing the composition of all the 16-day images into a single quality mosaic (QM), using the minimum NBR (Normalized Burn Ratio) spectral index (eq. 1 — Key and Benson, 2006) as a per-pixel ordering function. The pixel with the lowest NBR value was selected, and all its spectral reflectance characteristics, including the scene metadata with the date of that selected pixel, were used to create the annual quality mosaic.

$$\lambda QM = [Blue, Green, Red, NIR, SWIR1, SWIR2] = date\ in\ with\ min\left(\frac{\lambda_{NIR} - \lambda_{SWIR1}}{\lambda_{NIR} + \lambda_{SWIR1}}\right) [\, xi \ldots j \,]$$

eq. 1

Where λ represents the reflectance values of the quality bands that compose the quality mosaic (QM), retrieved from the date in which each pixel reached its minimum (min) NBR value in a given year (x), considering the set of all available scenes, from the first (i) to the last (j). The λNIR is the Near-Infrared surface reflectance and $\lambda SWIR1$ is the Short-Wave Infrared surface reflectance used to calculate the NBR spectral index.

In other words, we computed the NBR for each pixel with a valid observation within a specific year and stacked them into a multi-band image. The pixels with the lowest NBR within the multi-band image were selected, and their spectral information (Table 1) was used to compose the annual quality mosaic (QM). In addition to the spectral information, we retained the scene metadata information, including the date on which each pixel showed its lowest NBR value. The NBR quality mosaic created with the spectral information from the minimum NBR performed well in differentiating burned and unburned land use and cover in the Brazilian biomes (Figure 3).

Table 1. Spectral bands used as predictors in the classification process to identify burned areas.

Spectral band	**Landsat 5 and 7**		**Landsat 8**	
	Band number	**Band width (µm)**	**Band number**	**Band width (µm)**
Red	3	0.63 - 0.69	4	0.64 - 0.67
NIR	4	0.76 - 0.90	5	0.85 - 0.88
$SWIR_1$	5	1.55 - 1.75	6	1.57 - 1.65
$SWIR_2$	7	2.08 - 2.35	7	2.11 - 2.29

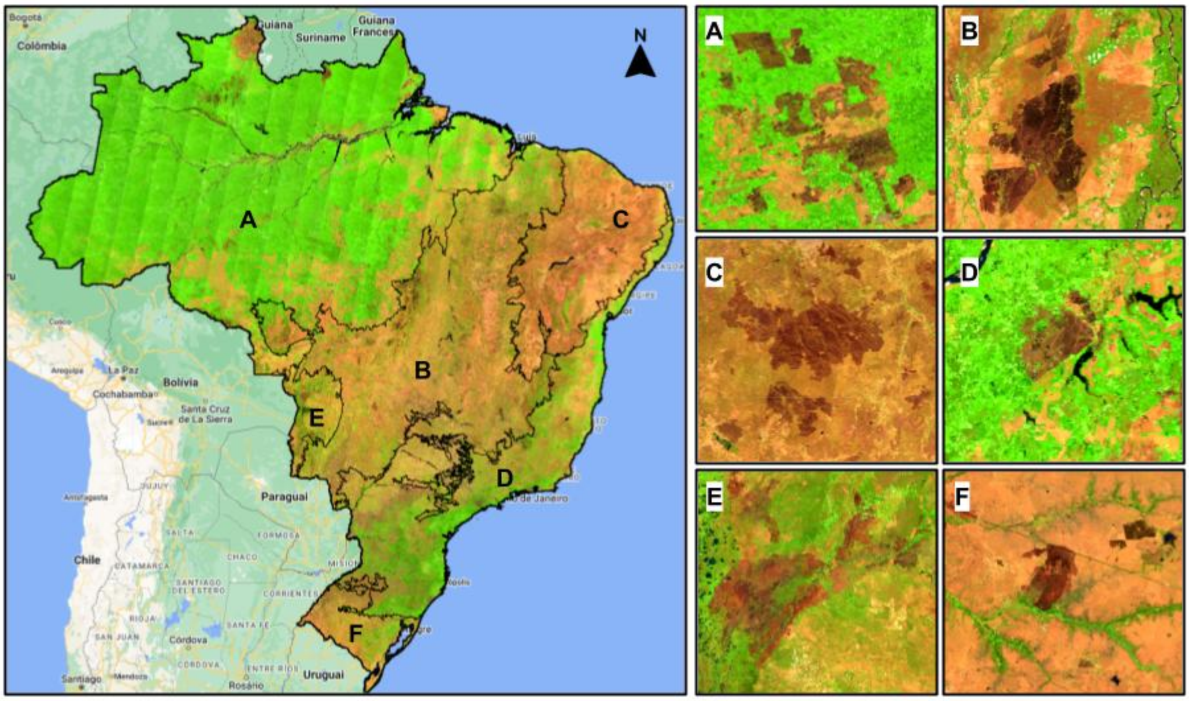

Figure 3. The 2022 quality mosaic (QM) for Brazil (RGB SWIR-1, NIR, RED), created from spectral information retrieved from the minimum NBR pixels in a year, showing examples of the diversity of burn scars by biome: (A) Amazon, (B) Cerrado, (C) Caatinga, (D) Atlantic Forest, (E) Pantanal, and (F) Pampa.

2.3. Training samples

We created a spectral library based on manual delineation of burned and unburned areas to be used as training samples. These samples were stratified by Landsat sensors (collected in different years) and by each biome. The collection of training samples was performed across all 28 classification regions, ensuring representation of the distinct spectral characteristics present in each region. Finally, we divided our spectral library into 28 packs (one for each classification region) and used it as input in the classification step.

2.4. Classification

The classification model used was the Deep Neural Network (DNN), which consists of computational models capable of performing deep learning and visual pattern recognition. The specific structure we used was the Multi-Layer Perceptron Network (MLPN), which incorporates several layers of interconnected computational units. In this structure, each node (neuron) in one layer is connected to a node in the next layer (Hu, Wenk, 2009). The layers are divided into input, hidden, and output layers.

For this DNN model, the input layers were the spectral bands RED, NIR, SWIR1, and SWIR2, and the output layers were the classes burned and unburned. The burned area mapping algorithm consisted of two main steps: training and prediction.

- **Training Phase:**

In the training phase, the following parameters were defined based on prior tests: learning rate (0.001), batch size (1000), number of iterations (7000), and inputs for classification (Arruda et al., 2021). The classification inputs used in this model were the surface reflectance (SR) spectral data retrieved from the annual quality mosaics using the training samples of burned and unburned areas.

Based on the spectral library from the burned and unburned training samples, the following spectral bands were used as inputs for the burned area classification model: Red (RED — 0.65 µm), Near-infrared (NIR — 0.86 µm), Short-wave infrared (SWIR1 — 1.6 µm and SWIR2 — 2.2 µm). These Landsat spectral bands were chosen based on their sensitivity to fire events across distinct land uses and covers.

The training data input was divided into two datasets: 70% of the samples were used for training and 30% for testing, in order to estimate the ability of the DNN algorithm to map burned areas accurately.

- **Prediction Phase:**

The classification was performed using the annual Landsat quality mosaics for each of the 28 regions and for each sensor (Landsat 5, Landsat 7, Landsat 8, and Landsat 9), resulting in 39 maps of burned areas for all of Brazil (Figure 4). This approach allowed us to leverage the powerful capabilities of DNNs for accurate and efficient mapping of burned areas across the diverse biomes of Brazil.

To enhance the quality of the results, a spatial filter was applied to eliminate noise and address small gaps in the classification. Specifically, burned areas equal to or smaller than 1.4 hectares (16 pixels) were removed, while empty gaps located within and fully surrounded by burned areas, and smaller than or equal to 5.8 hectares (64 pixels), were filled and reclassified as burned. This post-processing step significantly improved the overall consistency and accuracy of the final burned area maps.

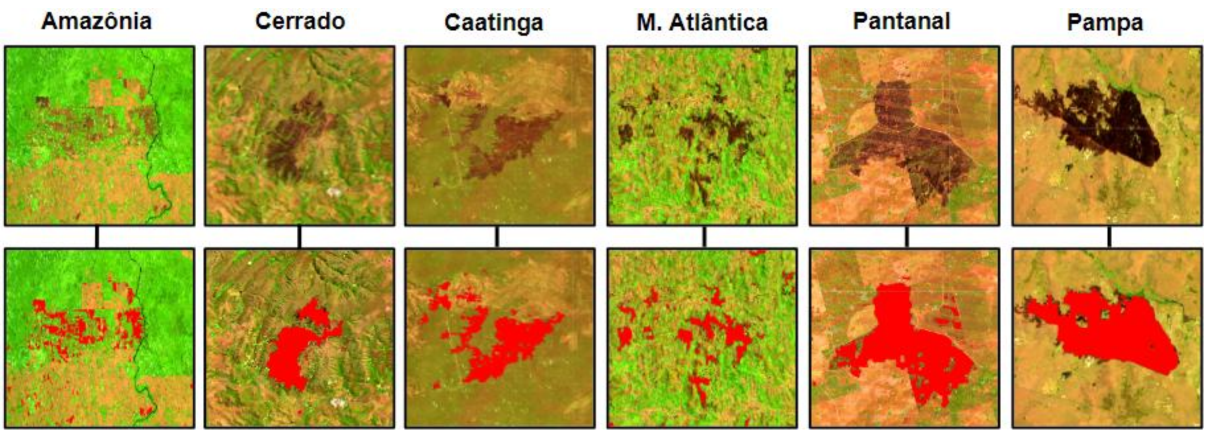

Figure 4. Examples of the burned areas classification for different types of fire, with the Landsat

mosaic used for classification, and the area classified as burned in red.

2.5. Post-classification: Temporal Filtering of Repeated Burned Areas in the Amazon

In dense forest regions of the Amazon, fire scars often remain spectrally visible in Landsat imagery for more than one year, which can lead to the same burned area being mapped in consecutive years. To address this issue, we implemented a temporal filtering step aimed at removing duplicate detections of burned areas across successive years, especially in cases where there is no strong evidence of a new fire event.

This procedure was applied over the entire Amazon biome between 2000 and 2024, focusing on pixels where fire was mapped in consecutive years, and the land cover was classified as Forest Formation (MapBiomas class 3) two years prior to the most recent fire detection. The filtering process followed these main steps (Figure 5):

1. Identification of Consecutive Burned Areas: Annual burned area maps were analyzed to identify pixels mapped as burned in two consecutive years (e.g., year t and year t-1).
2. Forest Cover Validation (Two Years Prior): For a pixel to be eligible for filtering, it must have been classified as Forest Formation in the year t–2 —the year before the first burn event. This ensures that the pixel was part of a stable forest area prior to the burn.
3. Hotspot-Based Fire Activity Validation: To confirm the occurrence of a new fire in year t+1, active fire detections (hotspots) from the current year (*t*) were used. A 1,000-meter buffer was created around each hotspot to define potential fire activity zones.
4. Exclusion of Unsupported Detections: If no hotspot was detected within the buffer in the year t-1, and the area overlapped with a previous burn in year t, the more recent detection was classified as a residual fire scar and excluded from the burned area map.
5. Removal of Small Events: Fire events smaller than 5 hectares were also removed to reduce noise and improve dataset reliability.

Figure 5. Post-classification procedure for removing duplicate detections of burned areas in Forest Formation across consecutive years in the Amazon biome.

This approach helps reduce temporal commission errors and improves the accuracy of the burned area time series by avoiding the double counting of single fire events. The temporal filtering was applied after the spatial and thematic post-classification masks were implemented.

2.6. Post-classification: Land Use and Land Cover Filtering

After evaluating the classification results, post-classification masks were applied to reduce commission errors caused by land use and land cover types that present spectral signatures similar to recently burned areas, such as water bodies, urban zones, and certain crop types. Specific biome-based rules were defined to eliminate pixels incorrectly classified as burned within the land use and cover classes from MapBiomas Collection 9. The exclusion rules for each biome are listed below:

- **Amazon**: Water (33,31), Urban Area (24), Mining (30), Beach, Dune, and Sand Spot (23)
- **Caatinga**: Water (33,31), Urban Area (24), Rocky Outcrop (29)
- **Cerrado**: Water (33,31), Urban Area (24), Mining (30)
- **Atlantic Forest**: Water (33,31), Urban Area (24), Rice (40), Mining (30), Beach, Dune, and Sand Spot (23)
    - Additional for regions 6 and 7: Soybean (39); Temporary Crops (19), Sugarcane (20), and Other Temporary Crops (41)
- **Pampa**: Water (33,31), Urban Area (24), Rice (40), Mining (30), Beach, Dune, and Sand Spot (23), Soybean (39), Other Temporary Crops (41), Mosaic of Uses (21)
    - Additional for 2024: Forest Formation (3), Wooded Sandbank Vegetation (49), Wetlands (11)
- **Pantanal**: Soybean (39), Cotton (62, beta), Other Temporary Crops (41)

In addition to these land use and land cover masks, we used stable water data from MapBiomas Collection 9, which identifies pixels consistently classified as water throughout the time series. A spatial connectivity filter was also applied to remove isolated pixels: areas with six or fewer connected pixels (equivalent to approximately 0.54 hectares) were discarded to reduce classification noise. Finally, post-classification processing retrieved the temporal information of each burned pixel from the annual minimum NBR mosaic, allowing the identification of the month in which the fire scar was detected.

2.7. Classification evaluation

Evaluations of the burn scar classification were conducted using Landsat mosaics through a combination of visual inspection and statistical analysis. Visual inspections were carried out by biome experts, who thoroughly reviewed the classified burn scars against the original Landsat imagery to ensure accuracy. Any discrepancies identified during these inspections were documented and used to refine training samples and improve algorithm performance.

Statistical analyses were also performed to validate the classification results by comparing them with multiple reference burned area products. Biome specialists conducted detailed visual checks, cross-referencing the results with local knowledge and complementary datasets to confirm the presence and extent of fire scars. The classified burn scars were compared against reference datasets from various sources, including: previous collections of MapBiomas Fire, MODIS MCD64A1 (500m) burned area products, INPE's fire hotspot data (1km), ICMBIO's manual mapping (30m), GABAM's high-resolution global burned area maps (30m), FIRMS' near-real-time fire detection (1km), and FireCCI's global burned area products (250m). This comprehensive evaluation process ensured the high quality and reliability of the burned area classification, combining automated assessments with expert-driven validation to generate robust and trustworthy results (Figure 6 and Figure 7).

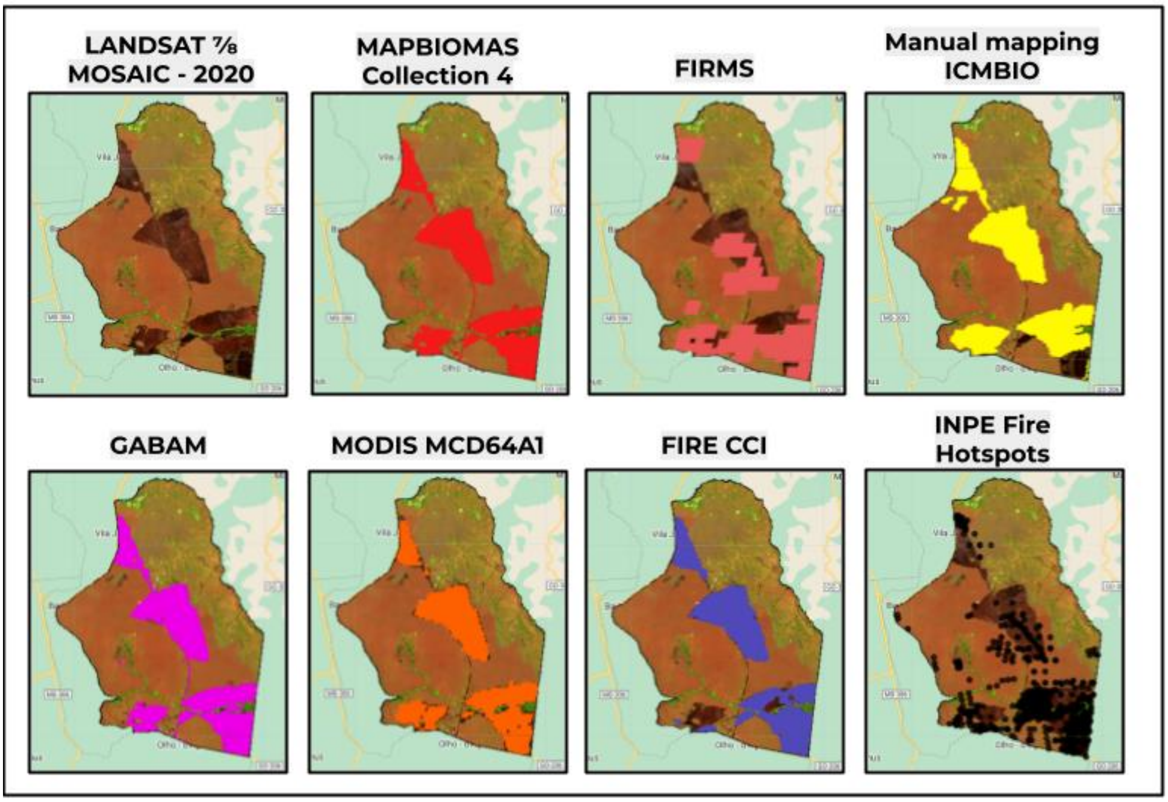

Figure 6. Landsat 7/8 Mosaic - 2020, MapBiomas Collection 4 (30m), FIRMS (1 km), ICMBIO Manual Mapping, GABAM (30m), MODIS MCD64A1 (500m), FIRE CCI (250m), and INPE Fire Hotspots (1km) comparisons.

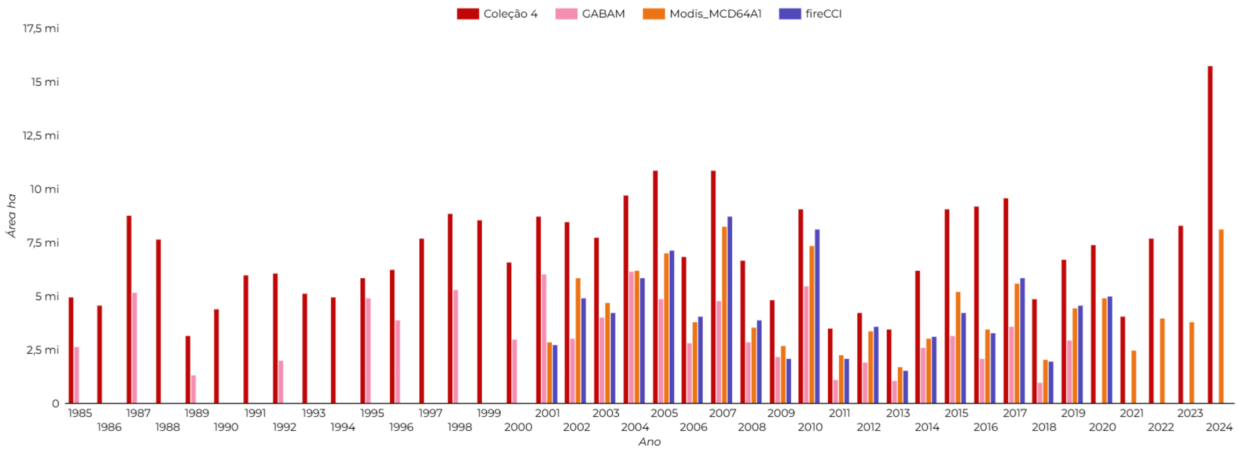

Figure 7. Annual burned area for the following data collections: MapBiomas Collection 4 (30m resolution), GABAM (30m resolution), MODIS MCD64A1 (500m resolution), and FIRE CCI (250m resolution).

3. MapBiomas Fire Products

3.1 Annual Burned Area

This dataset comprises annual maps identifying pixels classified as burned for each year within the 1985–2024 period. Each pixel is attributed with the corresponding land use and land cover (LULC) class from MapBiomas Collection 9 for the year in which the burning occurred. This product facilitates temporal analysis of fire occurrences across different land use categories.

3.2 Monthly Burned Area

Monthly burned area maps indicate the specific month (ranging from 1 to 12) when fire events occurred for each pixel during the 1985–2024 period. The temporal attribution is based on the date of the satellite image in which the pixel exhibited the minimum Normalized Burn Ratio (NBR) value, as derived from the annual quality mosaic. This product supports the examination of seasonal fire patterns.

3.3 Annual Burned Area by Scar Size

This dataset categorizes annual burned areas into ten distinct size intervals, measured in hectares, for the 1985–2024 period. A fire scar is defined as a contiguous cluster of burned pixels identified within the same year. This classification aids in understanding the distribution and scale of fire events.

3.4 Cumulative Burned Area

The cumulative burned area maps represent the total area affected by fire over the entire study period, accounting for each pixel only once, regardless of multiple fire occurrences. Additionally, the cumulative burned area is stratified by LULC classes from the final year of MapBiomas Collection 9, providing insights into long-term fire impacts on various land cover types.

3.5 Fire Frequency Data

Fire frequency maps depict the number of times each pixel was classified as burned throughout the 1985–2024 period. The data is aggregated into a single map with 39 classes, where class 1 indicates a single fire occurrence, class 2 indicates two occurrences, and so on. Each pixel also retains its LULC classification from the final year of MapBiomas Collection 9, enabling analysis of fire recurrence across different land uses.

3.6 Year of Last Fire Occurrence

This dataset records the most recent year in which each pixel was identified as burned, within the 1985–2024 timeframe. The information is derived from the annual quality mosaics, allowing for the assessment of time since the last fire.

4.References

Alencar, Ane A C, Vera L S Arruda, Wallace Vieira, Dhemerson E Conciani, Diego Pereira Costa, Natalia Crusco, Soltan Galano Duverger, et al. 2022. “Long-Term Landsat-Based Monthly Burned Area Dataset for the Brazilian Biomes Using Deep Learning.” Remote Sensing 14 (2510): 29. https://doi.org/10.3390/rs14112510.

Arruda, V. L. S.; Piontekowski, V. J.; Alencar, A.; Pereira, R. S.; Matricardi, E. A. T. An alternative approach for mapping burn scars using Landsat imagery, Google Earth Engine, and Deep Learning in the Brazilian Savanna. Remote Sensing Applications: Society and Environment, [s. l.], v. 22, p. 100472, 2021.

Giglio, L., Schroeder, W., Justice, C.O., 2016. The collection 6 MODIS active fire detection algorithm and fire products. Remote Sens. Environ. 178, 31–41. https:// doi.org/10.1016/j.rse.2016.02.054.

Hu, X.; Weng, Q. Remote Sensing of Environment Estimating impervious surfaces from medium spatial resolution imagery using the self-organizing map and multi-layer perceptron neural networks. Remote Sens. Environ. 2009, 113, 2089–2102.

Key, C., Benson, N., 2006. Landscape assessment: remote sensing measure of severity: the normalized burn Ratio. In: FIREMON: Fire Effects Monitoring and Inventory System. USDA Forest Service, Rocky Mountain Research Station, General Technical Report, RMRS-GTR-164-CD:LA1-LA51, Fort Collins, CO: U.S.

Langford, Z.; Kumar, J.; Hoffman, F. Wildfire mapping in interior alaska using deep neural networks on imbalanced datasets. In Proceedings of the 2018 IEEE International Conference on Data Mining Workshops (ICDMW), Singapore, 17–20 November 2018

Long, Tengfei, Zhaoming Zhang, Guojin He, Weili Jiao, Chao Tang, Bingfang Wu, Xiaomei Zhang, Guizhou Wang, and Ranyu Yin. 2019. "30 m Resolution Global Annual Burned Area Mapping Based on Landsat Images and Google Earth Engine" Remote Sensing 11, no. 5: 489. https://doi.org/10.3390/rs11050489