



MAPBIOMAS
[FOGO]

Algorithm Theoretical Basis Document (ATBD)

MapBiomas Fire

Collection 1.0

Version 1

General coordinator

Ane A. Alencar

Team

Dhemerson E. Conciani

Diego Pereira Costa

Eduardo Reis Rosa

Eduardo Vélez Martin

Heinrich Hasenack

Luiz Felipe Morais Martenexen

Julia Shimbo

Marcos Rosa

Natalia Crusco

Nilson Clementino Ferreira

Noely Vicente Ribeiro

Sarah Moura B. dos Santos

Soltan Galano Duverger

Tasso Azevedo

Valderli J. Piontekowski

Vera L. S. Arruda

Wallace Vieira da Silva

Washington da Franca Rocha

Table of Contents

1.Introduction	2
1.1. Overview of the MapBiomass Fire	2
1.2. How we are organized	2
2. Methodological description	3
2.1. Definition of regions by biome	4
2.2. Annual mosaics	5
2.3. Training samples	6
2.4. Classification	7
2.5. Post-classification	7
2.6. Classification evaluation	8
2.7. Validation	8
2.8. Fire Statistics	10
2.8.1. Accumulated	10
2.8.2 Fire frequency data	10
3.References	10

1.Introduction

1.1. Overview of the MapBiomass Fire

The objective of this document is to describe the theoretical basis, justification and applied to produce the monthly and annual maps of burned areas in Brazil from 1985 to 2020 of the MapBiomass Fire Collection 1.

The MapBiomass Fire project released its beta collection of annual maps of fire scars covering the period of 2000 to 2019 in December 2020. Some improvements in the methodology and the collection of samples was made for the MapBiomass Fire Collection 1 of mapping of fire scars in Brazil, which is based on mosaics of images from the Landsat satellites with a spatial resolution of 30 meters. The mapping period was expanded from 1985 to 2020, now with monthly and annual data of burned areas covering the entire Brazilian territory.

The entire process was carried out collaboratively between MapBiomass institutions and using machine learning algorithms (deep learning) through the Google Earth Engine and Google Cloud Storage platform which offer immense processing capacity in the cloud.

The classification was organized by biomes and regions, collecting burned and unburned area samples 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 of spatial resolution and INPE fire scars (<https://queimadas.dgi.inpe.br/>).

The products of the MapBiomass Fire Collection 1 are the following:

- Monthly and annual burned area maps in Brazil from 1985 to 2020;
- Frequency of annual burned areas in Brazil;
- Accumulated burned areas in Brazil;
- Burned areas over land cover and land use classes of MapBiomass Collection 6.

1.2. How we are organized

MapBiomass is a network of institutions involving NGOs, universities, and technology companies dedicated to understanding the land use dynamic in Brazil. For the MapBiomass Fire, IPAM conducted technological and operational development. Expert teams in each biome carried out sampling, evaluation and refinement of the mapping (Figure 1). The geospatial tech company Ecostage is responsible for the backend and dashboard/website/frontend development of MapBiomass.

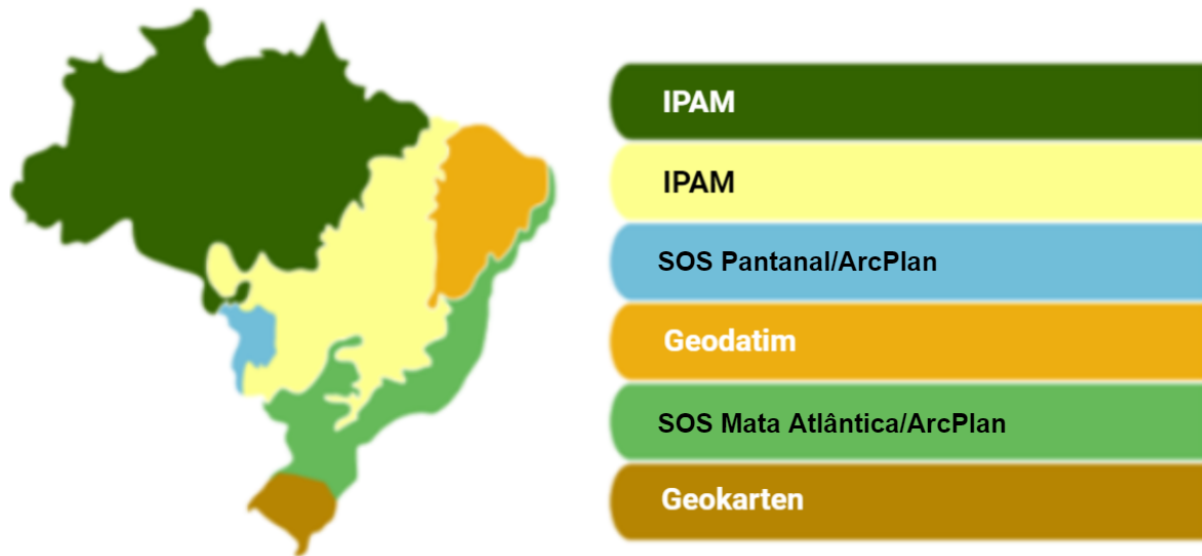


Figure 1. Overview of the teams involved in the MapBiomass Fire.

1.3. Historical Perspective: Existent Maps and Mapping Initiatives:

Global burned area products from the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor aboard the Terra and Aqua satellites provided by the National Aeronautics and Space Administration (NASA) have been globally used, with a coarse spatial resolution (500 m), in 15 days-based products available for worldwide download. We used the product MCD64A1 Burned Area Product as a reference data of burned areas (Giglio et al., 2016). The. 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, and the AQUA_M-T (Sensor MODIS) as a reference satellite, providing daily data of fire hotspots since 2000, available at <http://www.inpe.br/queimadas/bdqueimadas>.

2. Methodological description

The first step consisted of defining the regions by biome for classification. The second step consisted in the construction of annual Landsat mosaics for classification. The third step consisted of collecting training samples of burned and unburned areas in Landsat mosaics on the Google Earth Engine platform. For the classification step, the Google Cloud Storage Bucket platform was used. The mosaics for classification and the training samples were stored, and virtual machines were used for processing the classification algorithms.

The diagram below illustrates the main steps and platforms used to classify fire scars in Brazil (Figure 2).

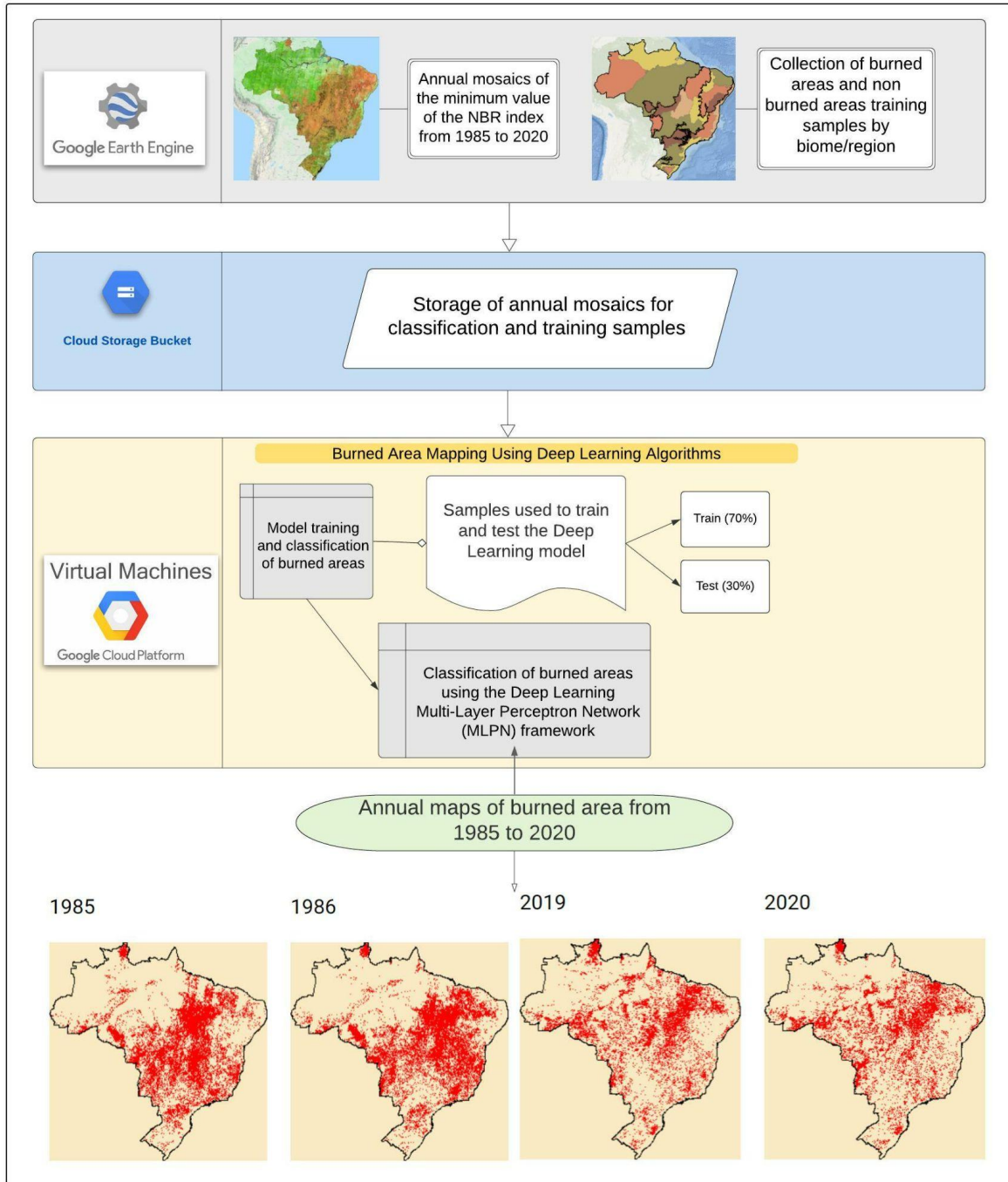


Figure 2. Overview of the method for classifying burned areas in Brazil in MapBiomass Fire Collection 1.0.

2.1. Definition of regions by biome

For each biome, regions were defined to collect training samples and classify burned areas by region to obtain a more precise classification based on edaphoclimatic factors and regional vegetation. The following regions were defined for each biome (Figure 3):

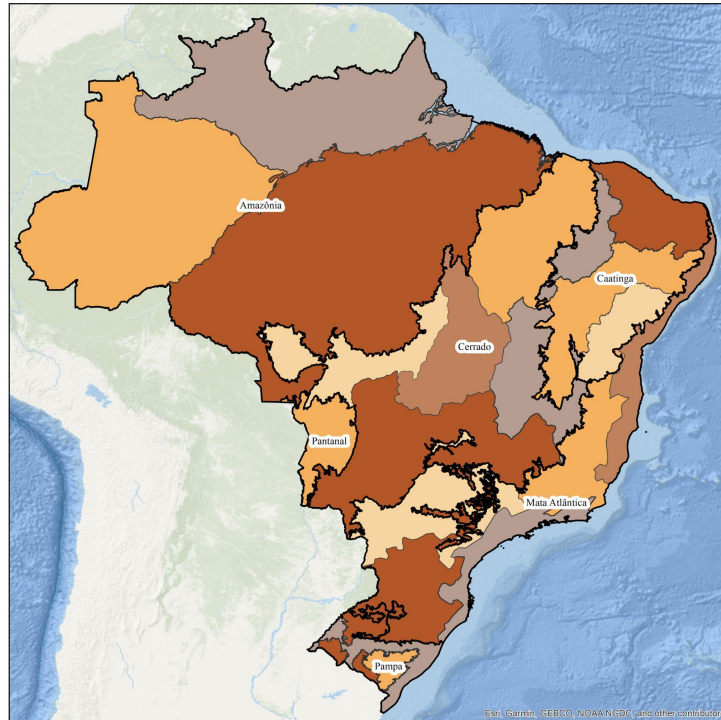


Figure 3. Regions defined for each biome in Brazil to collect training samples and classify burned areas in the MapBiomas Fire Collection 1.

2.2. Annual mosaics

The classification was performed using surface reflectance (SR) Landsat mosaics (30 x 30 meters) constructed for each year from 1985 to 2020. We assessed all the available scenes of Landsat 5 (from 1985 to 1998, and from 2003 to 2011), Landsat 7 (1999 - 2002, and 2012) and Landsat 8 (2013 - 2020), all with a 16-day time return interval. In total, we assessed 154,261 (or ~108 TB of) Landsat scenes distributed over 389 different WRS-2 path/row tiles that overlap the Brazilian territory (on average 4,285 scenes per year) (Figure 4).

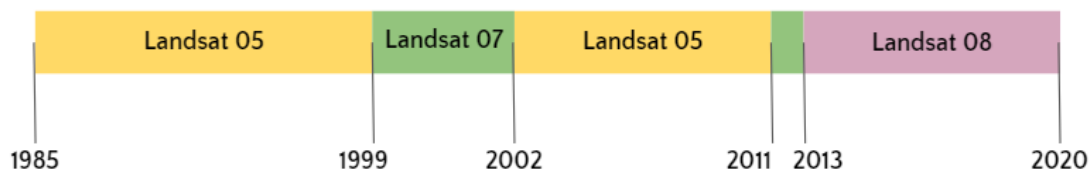


Figure 4. Landsat periods assessed for the MapBiomas Fire Collection 1.

Cloud and cloud shadow were masked for each scene by using their respective 16-bit Quality Assessment Band (BQA). We applied permissive rules and only pixels with high confidence levels of 'cloud', 'shadow' and 'saturation' were masked. We used a per year statistical approach to summarize this amount of data and optimize the classification without discarding spectral information about every pixel. We created yearly mosaics by performing the composition of all the 16-day images into a single image, using the minimum NBR

(Normalized Burn Ratio) spectral index (eq. 1 — Key and Benson, 2006) as a per-pixel ordering function (Figure 5).

eq. 1

$$NBR = \frac{\lambda NIR - \lambda SWIR1}{\lambda NIR + \lambda SWIR1}$$



Figure 5. (a) Example of the multi-temporal mosaic used to perform the burned area classification. Note that this image addresses all the burned areas detected in the monthly scenes from January (b) to December (m).

In other words, we computed the NBR for each one of the 16-day available scenes for a specific year and stacked them into a multi-band image. After this, a per-pixel ordination was applied to construct a single-band minimum NBR image. In addition to the NBR value, we also retained the scene date metadata in which each pixel was shown its lowest value. Finally, we used this metadata to build a multi-temporal mosaic by recovering the per pixel value from SR bands (Table 1) of the scene where each pixel showed its lowest NBR value.

Table 1. Mosaic bands used as predictors in the classification of 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

2.3. Training samples

We created a spectral library with 280,456,236 sampled pixels based on manual delineation of burned areas (89,845,700 sampled pixels; 32%) and unburned areas (190,610,536 sampled pixels; 68%) to be used as training samples. These samples were stratified by Landsat sensors (collected in different years) and each biome (Figure 6). Finally, we divided our spectral library into 21 packs (one for each classification region) and used it as input in the classification step.

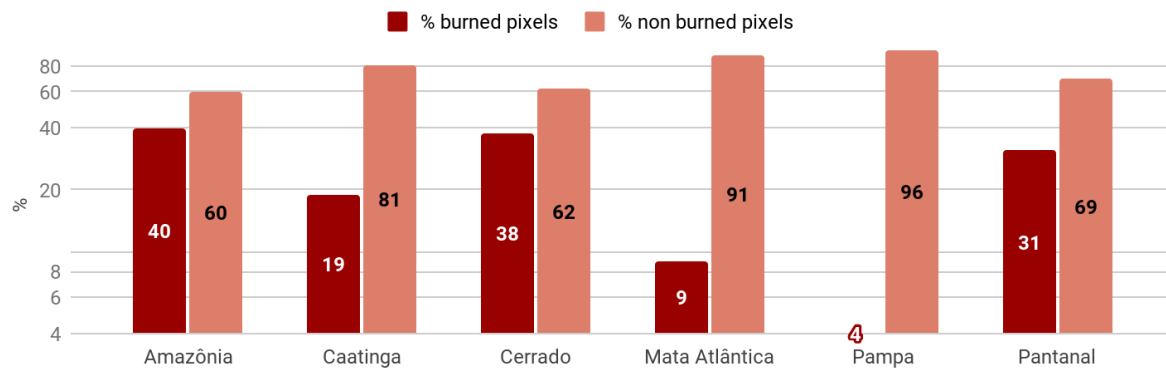


Figure 6. Proportion of burned and unburned pixels sampled by biome.

2.4. Classification

The classification model used was the Deep Neural Network, which consists of computational models based on mathematical calculations capable of performing machine learning and visual pattern recognition.

The burned area mapping algorithm consisted of two steps: training and prediction. Based on the burned and unburned training samples, the following spectral bands were used as input to the burned area classification model: red (RED - 0.65 μm), near infrared (NIR - 0.86 μm), and short wave infrared (SWIR 1 - 1.6 μm and SWIR 2 - 2.2 μm). In addition to these spectral bands, the Landsat bands 2 to 7 and the NDVI (Normalized Difference Vegetation Index), NBR and Delta NBR indices were tested. These spectral Landsat bands were chosen based on their sensitivity to fire events. Training data input was divided into two sets; 70% of the samples were used for training and 30% for testing (Arruda et al., 2021).

2.5. Post-classification

After training and testing the model, the classification was applied with Landsat images fortnightly for the entire period of analysis (1985 to 2020). A spatial filter was applied to remove noise and fill small empty gaps: areas smaller than or equal to 1.4 ha (16 pixels) were removed, and empty gaps smaller than or equal to 5.8 ha (64 pixels) were filled as areas burned.

After evaluating the classification results, post-classification filters were also applied, removing pixels that were in the following land cover and land use classes of MapBiomas Collection 6 and biomes:

- Amazon: Water, Urban Area and Rocky Outcrop;
- Caatinga: Water, and Rocky Outcrop;
- Cerrado: Water and Urban Area;
- Atlantic Forest: Water, Urban Area and Rice;
- Pampa: Rice, Soybeans, Other Temporary Crops, Urban Area and Water;
- Pantanal: Water, Soybean and Other Temporary Crops;

As Deep Learning methods require large computational processing, implementations were based on graphics processing units (GPUs) and specialized hardware components to perform parallel arithmetic operations. Access to GPUs in a virtual machine environment was implemented on the Google Cloud Platform (<https://console.cloud.google.com>), a set of cloud computing services provided by Google.

To obtain the information of the month in which the fire scar was mapped, post-classification processing was performed to retrieve the date information of the pixel that was burned, from the date of the pixel in which the annual mosaic was built from the minimum NBR.

2.6. Classification evaluation

Evaluations of the classification of burn scars were carried out with Landsat images, with visual inspection, statistics and relationship with land cover and land use data of MapBiomass, in addition to comparison with reference maps (MODIS, INPE, FIRMS, FireCCI) (Figure 7).

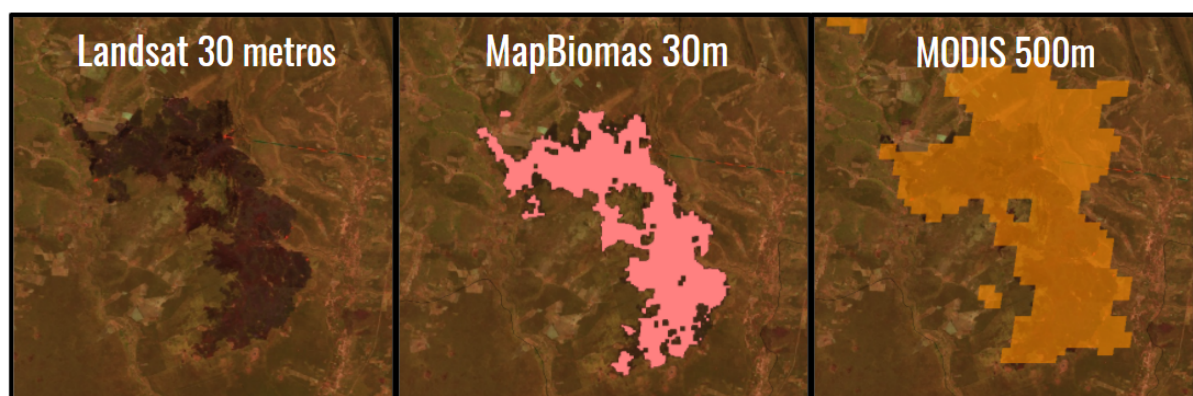


Figure 7: Landsat image with 30 meters resolution, MapBiomass Fire classification and MODIS (MCD64A1) classification at 500 m resolution.

2.7. Validation

The validation of the burned area mapping was performed by biome, considering the years

2007, 2011, and 2019. Each of the biomes was divided into 2 km by 2 km grids and these grids were spatially integrated with the burned area mapping of Fire Information for Resource Management System (FIRMS), for the same years, to consider as squares without fire occurrence, with up to 32% of fire occurrence, with fire occurrence between 32% and 70% and with fire occurrence above 70%. Thus, the squares were divided into four groups for each of the biomes, for each of the years, that is, as squares without fire, with fire, with medium fire, and with high fire.

After this categorization of the grid cells, eq. 2 was applied to obtain the number of sampling units for each group of grid cells, in each of the biomes, for the years 2007, 2011, and 2019.

$$n = \frac{N \cdot Z^2 \cdot p \cdot (1-p)}{(N-1) \cdot e^2 + Z^2 \cdot p \cdot (1-p)} \quad \text{eq. 2}$$

where: n is the sample size, N is the population size, Z is the standardized normal distribution score corresponding to the adjusted confidence level $1-\alpha=0.95$ (95%), and is the maximum margin of error expected (4.18%), p (0.70) the proportion you want to estimate $eq=1-p$.

From the application of equation 1, considering the number of squares in each of the groups, for the years 2007, 2011 and 2019, for the Brazilian biomes, the following numbers of total sampling units were defined:

Table 2. Total sampling units by biome and year used for validation.

	Amazon	Caatinga	Cerrado	Atlantic Forest	Pampa	Pantanal
2007	370	370	370	370	352	364
2011	370	370	370	370	342	361
2019	370	370	370	370	337	366

Then, the random selection of sampling units was carried out, considering the three categories of fire occurrence and also the category of non-occurrence of fire. With this, maps of sampling units were produced for all biomes, for the years 2007, 2011, and 2019.

Subsequently, the maps of sampling units were inserted into the Google Earth Engine platform, so that the compositions of minimum NBR for each year were cut for each sampling unit.

In the next step, the images of the sample units were extracted from Google Earth Engine and segmented using the QGIS software, with the OrfeoToolbox complement. The segments

stored in vector structure were visually interpreted to define whether it was in an area with fire occurrence, or in an area without fire occurrence. Thus, the information on the occurrence or not of fire in the area of each segment was explained based on the value of the “fire” attribute of the vector data, 1 for fire, 0 for non-fire.

After the interpretation of all segments, the centroids with the respective attributes of the segments were extracted. These centroids were then spatially integrated with the burned area mappings, making it possible to count the locations that were mapped and interpreted as being burned sites, the locations that were mapped and interpreted as unburned sites, the locations that were mapped as burned and interpreted as unburned (inclusion error) and, finally, the locations that were mapped as unburned and were interpreted as burned (omission error). In this way, it was possible to assess the amount of inclusion and omission errors, as well as the global accuracy, for each biome, and also for Brazil, for the years 2007, 2011 and 2019.

2.8. Fire Statistics

2.8.1. Cumulative burned area

The accumulated burned area data available in the MapBiomas platform was built from the increment of the burned area for each year; that is, the same pixel is only counted as fire once, regardless of whether there was more than one fire occurrence.

The accumulated burned area data by land cover and land use type was done based on crossing the occurrence of fire with the land cover and land classes of MapBiomas Collection 6, considering the last year of the period.

2.8.2 Fire frequency data

The burning frequency data was produced from the grouping of the burned area in each year, running on a single map with 36 classes for the entire period (1985-2020), where class 1 represents the pixels that burned once, class 2 the pixels that burned 2 times and so on.

The frequency data by land cover type was done by crossing the land cover and land use map of MapBiomas Collection 6 of the last year of the period.

3. References

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>.

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.