MapBiomas 10-meters

Algorithm Theoretical Basis Document (ATBD)

Collection 2 (beta)

Version 1

September, 2025

Table of Contents

Executive Summary

MapBiomas initiative was formed in 2015 by universities, NGOs, and tech companies to develop a fast, reliable, collaborative, and low-cost method to produce time series of annual maps of land cover and land use (LCLU) in Brazil. The network is organized by biomes (Amazon, Atlantic Forest, Caatinga, Cerrado, Pampa, and Pantanal) and cross-cutting themes (Pasture, Agriculture, Coastal Zone, Mining, and Urban Area) and all derived products and scripts are publicly and freely available to all. Since then, MapBiomas has produced nine collections of LCLU annual digital maps from 1985 to 2023 using Landsat satellite images with 30 meters spatial resolution, progressively improving the methods, increasing the number of mapped classes and enhancing the mapping accuracy.

Aiming to provide data with higher spatial resolution, in 2023 MapBiomas launched the first collection in beta version of the land use and land cover annual maps of Brazil at 10 meters resolution using Sentinel-2 satellite images covering the period from 2016 to 2022 (period of availability of Sentinel), named as the initiative MapBiomas 10 m. This new dataset allowed the inclusion of more detailed information in the mapping, for example, in rural properties, urban areas infrastructures and riparian forest in Permanent Preservation Areas (APP) along rivers and springs.

The second collection (beta), launched in 2025, covered the period of 2016 to 2023 and 21 LCLU classes, including three new classes: floodable forest in the Amazon, aquaculture and *apicum* (hypersaline tidal flat) in the Coastal Zone. This collection applied similar methods and the same legend of the land cover and land use classes of the MapBiomas Collection 9 (up to the third level of the legend). Seeking to continually improve methods and update the collection, maps from 2016 to 2022 were reprocessed and the year 2023 was mapped with the same spatial resolution, which resulted in the MapBiomas 10 m Collection 2 (beta). The shorter time series (from 2016 to 2023) decreased the temporal consistency of the Sentinel-2 mapping compared to the longer time series of Landsat mapping.

This Algorithm Theoretical Basis Document (ATBD) aims to provide the methodological steps of the MapBiomas 10 m Collection 2 (beta). All MapBiomas 10 m' maps and datasets are freely available on the project platform (https://plataforma.brasil.mapbiomas.org/cobertura_10m) and to download (https://brasil.mapbiomas.org/colecoes-mapbiomas/), as well as all computational algorithms used in the MapBiomas classifications, which are available on Github (https://github.com/mapbiomas).

1. Introduction

1.1. Scope and content of the document

This document describes the theoretical basis, objectives, and methods applied to produce annual maps of land cover and land use (LCLU) in Brazil from 2016 to 2023 of the MapBiomas 10 m Collection 2 (beta). It covers the classification methods, the image processing architecture, and the approach to integrating the biomes and cross-cutting theme maps. In addition, the document presents a historical context and background information, a general description of the satellite imagery datasets, and feature inputs.

1.2. Overview

The MapBiomas project was launched in 2015, aiming at contributing towards the understanding of LCLU dynamics in Brazil. The LCLU annual maps produced in this project were initially based on the Landsat series archive available in the Google Earth Engine platform, encompassing the years from 1985 to the present. Since then, the MapBiomas published nine LCLU mapping collections in Brazil, launching a new 30-m collection each year.

The MapBiomas collections aim to develop a fast, reliable, collaborative, and low-cost method to process large-scale datasets and generate historical time series of LCLU annual maps. All data, classification maps, codes, statistics, and further analyses are openly available through the MapBiomas platform (https://plataforma.brasil.mapbiomas.org/). All these are possible thanks to: i) Google Earth Engine platform, which provides access to data, image processing, standard algorithms, and the cloud computing facility; ii) freely available Landsat and Sentinel time-series data; iii) MapBiomas collaborative network of organizations and experts that share knowledge and mapping tools; and iv) visionary funding agencies that support the project (Souza Jr et al., 2020).

In 2023, MapBiomas launched the beta version of the first land use and land cover annual maps of Brazil at 10-meter resolution, using Sentinel-2 imagery for the period from 2016 to 2022. In 2025, the second collection extended the mapping to 2023, followed the same legend as MapBiomas Collection 9 at level 3, covering 21 LCLU classes, with three new classes: floodable forest in the Amazon, aquaculture, and *apicum* (hypersaline tidal flat) in the Coastal Zone. The products of the MapBiomas 10 m Collection 2 (beta) were the following:

- Biome maps (Amazon, Atlantic Forest, Caatinga, Cerrado, Pampa, and Pantanal) and cross-cutting theme maps (Pasture, Agriculture, Coastal Zone, Mining, and Urban Area);

- Pre-Processed feature mosaics generated from Sentinel-2 data.

• LCLU statistics and spatial analysis with political territories (e.g. biomes, states, municipalities), watersheds at different levels, protected areas (e.g. conservation units and indigenous territories), and other land tenure categories (e.g. rural setlements, quilombola lands).

Besides these LCLU mappings, the MapBiomas network released MapBiomas Water and Fire collections featuring annual and monthly maps of Brazil's water surface and fire scars from 1985 to the present, respectively. Annual maps of topsoil (0 - 30 cm) organic carbon stocks and soil texture mapping were launched as part of the MapBiomas Soil beta collection, along with the open soil research data repository SoilData. ince MapBiomas operates collaboratively with open data and methods, the network has expanded to other countries, engaging local institutions in mapping across all South American countries and Indonesia (https://brasil.mapbiomas.org/iniciativas-mapbiomas/).

1.3. Region of Interest

MapBiomas was created to produce LCLU annual maps for the entire Brazilian territory, covering all the six official biomes of the country: Amazon, Atlantic Forest, Caatinga, Cerrado, Pampa, and Pantanal (Figure 1). A biome is a geographic region defined based on vegetation types associated with geomorphological and climatic conditions. Our maps are developed per biome and then integrated into a single map of Brazil's LCLU classes in a post-processing step. The official Brazilian biomes map (1:250,000) developed by IBGE (2019) has been used. Classifying LCLU classes at the biome level helps to better discriminate specific LCLU classes and landscape patterns across the country (Table 1). In Brazil, the mapping approach was also divided into cross-cutting themes to improve the classification accuracy and quality: Agriculture, Pasture, Coastal Zone, Mining, and Urban Area.

Figure 1. Brazilian biomes used in the MapBiomas project to generate the Collection 2 (beta) of 10 m land cover and land use mapping products (source: IBGE, 2019).

Table 1. Land cover and land use characteristics of the Brazilian biomes.

Biome	**Area (km^2) (Country %)**	**Land Cover**	**Predominant Land Use**
Amazon	4,196,943 (49.29%)	Evergreen forest, with enclaves of savanna, natural grassland, and extensive wetlands and surface water, with almost 20% of the forest area of the biome cleared.	Cattle ranching, agriculture, mining, logging and non-timber forestry production.

Atlantic Forest	1,110,182 (13.04%)	Isolated forest fragments (Morellato and Haddad, 2000), mostly old secondary growth, surrounded by croplands, pasture, forest plantation, and urban area.	Agriculture, cattle ranching, urban, forest plantation, artificial water reservoir.
Caatinga	844,453 (9.92%)	Woody and deciduous physiognomies, with at least 50% of the original converted (de Oliveira et al., 2012).	Agriculture, cattle ranching, smallholder livestock production, urbanization.
Cerrado	2,036,448 (23.92%)	Mosaic of savanna, grassland, and forest, 50% of the native vegetation cover has already been converted (PPCerrado/INPE).	Agriculture, cattle ranching.
Pampa	176,496 (2.07%)	Natural grassland, with scattered shrub and trees, rock outcrop formations (Roesch et al., 2009).	Agriculture (rice, soy, perennial crops), livestock production (in natural grasslands), forest plantation, and urbanization.
Pantanal	150,355 (1.76%)	Forest, savanna, grassland and wetland.	Agriculture and cattle ranching.

1.4. Key Science Applications

MapBiomas was initially designed to address knowledge gaps in Brazil's greenhouse gas emission estimates from the land-use change sector. However, its annual time-series of land cover and land use (LCLU) maps also enable various scientific applications. In particular, the 10-meter resolution mapping enhances other detailed analyses, such as those focused on rural properties and urban areas. For example, higher-resolution imagery allows for more precise classification of certain land cover types, such as riparian forests in Permanent Preservation Areas (APP) along rivers and springs, among other applications:

- Mapping and quantifying LCLU transitions
- Quantification gross and net forest cover loss and gain
- Monitoring water resources and their interaction with LCLU classes
- Monitoring agriculture and pasture expansion
- Monitoring natural disasters
- Tracking infrastructure expansion and urbanization
- Identifying desertification processes
- Supporting regional planning
- Managing protected areas
- Modeling land-use changes
- Modeling infectious disease risks

- Modeling climate change

2. Overview and Background Information

2.1. Context and Key Information

This section addresses complementary contextual and critical information relevant to understand the MapBiomas products and methods used in the map collections.

2.1.1. MapBiomas Network

MapBiomas is a collaborative network of NGOs, universities, and technology startups committed to mapping land cover and land use changes across Brazil. Each organization plays specific or multiple roles, contributing to the project's overall development (a list of organizations can be found in Annex I). Each biome and cross-cutting theme (Agriculture, Pasture, Coastal Zone, Mining, and Urban Area) has a lead organization, as shown in the box below.

Biome coordination:

• **Amazon** – Institute of People and Environment of the Amazon (IMAZON).

• **Atlantic Forest** – SOS Atlantic Forest Foundation and ArcPlan.

• **Caatinga** – State University of Feira de Santana (UEFS) and Geodatin.

• **Cerrado** – Amazon Environmental Research Institute (IPAM).

• **Pampa** – Federal University of Rio Grande do Sul (UFRGS) and GeoKarten.

• **Pantanal** – SOS Pantanal Institute and ArcPlan.

Cross-cutting theme coordination:

• **Pasture** – Federal University of Goias (LAPIG/UFG).

• **Agriculture and Forest Plantation** – Remap.

• **Coastal Zone and Mining** – Solved and Vale Technological Institute (ITV).

• **Urban Area** – University of São Paulo (USP - QUAPÁ-FAU and YBY), Federal University of Bahia (UFBA), and Federal University of São Carlos (UFSCar - NEEPC).

The geospatial tech company Ecostage is responsible for the dashboard and website / backend and frontend of MapBiomas. The tech-startup Ecode is responsible for integration, post-classification and statistical analysis. Google provides the cloud computing infrastructure that allows data processing and analysis through Google Earth Engine and storage through Google Cloud Storage.

Since the beginning of MapBiomas, the funding for the MapBiomas implementation came from different funders the Skoll Foundation, Woods & Wayside International (WWI), Amazon Fund, Mulago Foundation, Climate and Land Use Alliance (CLUA), Alana Institute, Yield Giving, Ballmer Group, Valhalla Foundation, Sea Grape Foundation, Waverley Street Foundation, The Overbrook Foundation, The Patchwork Collective, Beja Institute, International Center for Tropical Agriculture (CIAT), Umbuzeiro Institute, Norwegian International Climate and Forest Initiative (NICFI), Climate and Society Institute (ICS), Quadrature Climate Foundation (QCF), Montepelier Foundation, Walmart Foundation (USA), Sequoia Climate Foundation, Good Energies Foundation, Gordon & Betty Moore Foundation, Global Wildlife Conservation (GWC), Wellspring philanthropic fund, OAK Foundation, Instituto Humanize, Arapyaú Institute Children's Investment Fund Foundation (CIFF).

Since MapBiomas is not an institution, the initiative received generous institutional management to operational and financing tasks from partners, including Arapyaú Institute, IAMAP and Avina Foundation.

The project also has an independent Scientific Advisory Committee (SAC) in MapBiomas Brazil, presently composed by:

- Dr. Alexandre Camargo Coutinho (Embrapa)
- Dr. Edson Eygi Sano (IBAMA)
- Dr. Gerd Sparovek (University of São Paulo)
- Dra. Leila Maria Garcia Fonseca (INPE)
- Dra. Liana Oighenstein Anderson (CEMADEN)
- Dra. Marina Hirota (Federal University of Santa Catarina)

And also former members who contributed to the project's development on previous collections:

- Dr. Gilberto Camara Neto (INPE)
- Dr. Joberto Veloso de Freitas (Federal University of Amazonas)
- Dr. Matthew C. Hansen (Maryland University)
- Dr. Mercedes Bustamante (University of Brasília)
- Dr. Timothy Boucher (TNC)
- Dr. Robert Gilmore Pontius Jr (Clark University)

2.1.2. Remote Sensing Data

The imagery dataset used in the MapBiomas 10 m Collection 2 (beta) was obtained by the harmonized collection of Sentinel-2 sensor MultiSpectral Instrument (MSI) (https://sentiwiki.copernicus.eu/web/s2-mission), accessible via Google Earth Engine and produced by the European Spatial Agency (ESA).

2.1.3. Google Earth Engine and MapBiomas Computer Applications

MapBiomas data processing is based on Google technology, which includes image processing in cloud computing infrastructure, programming with Javascript and Python via Google Earth Engine, and data storage using Google Cloud Storage. Google defines Google Earth Engine as: “a platform for petabyte-scale scientific analysis and visualization of geospatial datasets, both for public benefit and for business and government users.”

The MapBiomas project has developed the following computer applications based on Google Earth Engine:

• Javascript codes - these scripts were written directly in the Google Earth Engine Code Editor. The classification, post-classification and map integration of MapBiomas 10 m Collection 2 (beta) were written in Javascript.

• Python scripts - This code category was used to optimize image processing of large datasets in Google Earth Engine. In addition, the image mosaicking and statistical analysis were all performed in Google Earth Engine Python API. The python API also is used to train and predict deep learning models, as U-Net.

• Plataforma.brasil.mapbiomas.org (dashboard). The web platform of the MapBiomas initiative presents LCLU annual maps of MapBiomas 10 m Collection 2 (beta), and its graphs and statistics. Besides the 10 m LCLU data (https://plataforma.brasil.mapbiomas.org/cobertura_10m), the MapBiomas dashboard presents other products/modules, such as Collection 9 (Landsat) LCLU, temporal analysis, deforestation, secondary vegetation, agriculture, infrastructure, pasture, fire, mining, soil, water, degradation, urban mappings. All maps, data, and methodological documents of the MapBiomas Collections are open and freely available to download, and more information about the MapBiomas initiatives are available in the website (http://mapbiomas.org/en).

3. General Methodological Description

The general methodological steps of MapBiomas Brazil 10 m Collection 2 (beta) are presented in Figure 2. The first step was to generate annual Sentinel-2 mosaics comprising all images from each year to discriminate the LCLU classes across biomes and cross-cutting themes. The second step was to derive all feature space attributes from the Sentinel-2 bands to train one random forest classifier (feature space definition) for each year (Breiman, 2001). Then, training samples were yearly acquired in each biome and cross-cutting theme according to its information availability and statistical needs. The classification was carried out either using Random Forest, Gradient Tree Boosting or U-Net, depending on the biome or cross-cutting theme (Figure 2). In any case, the output is one LCLU map per year for the entire territory based on the training dataset of that year.

Spatial-temporal filters were applied over the classified data for noise removal and temporal stabilization. Subsequently, the filtered LCLU maps of each biome and cross-cutting themes were hierarchically merged (integrated) based on a set of prevalence rules. A spatial filter was once again applied in the integrated LCLU maps to create the final 10 m Collection 2 (beta) with a minimum mapping unit of 0.5 hectare, after temporal filters application to specifically adjust the temporal consistency of forest plantation and agriculture classes. For the other classes, another temporal filter was applied to eliminate isolated pixels in the time series. MapBiomas Alerta deforestation data was used as an accumulated mask, and when it overlapped with any natural class, it was converted to class 21 (Mosaic of Uses).

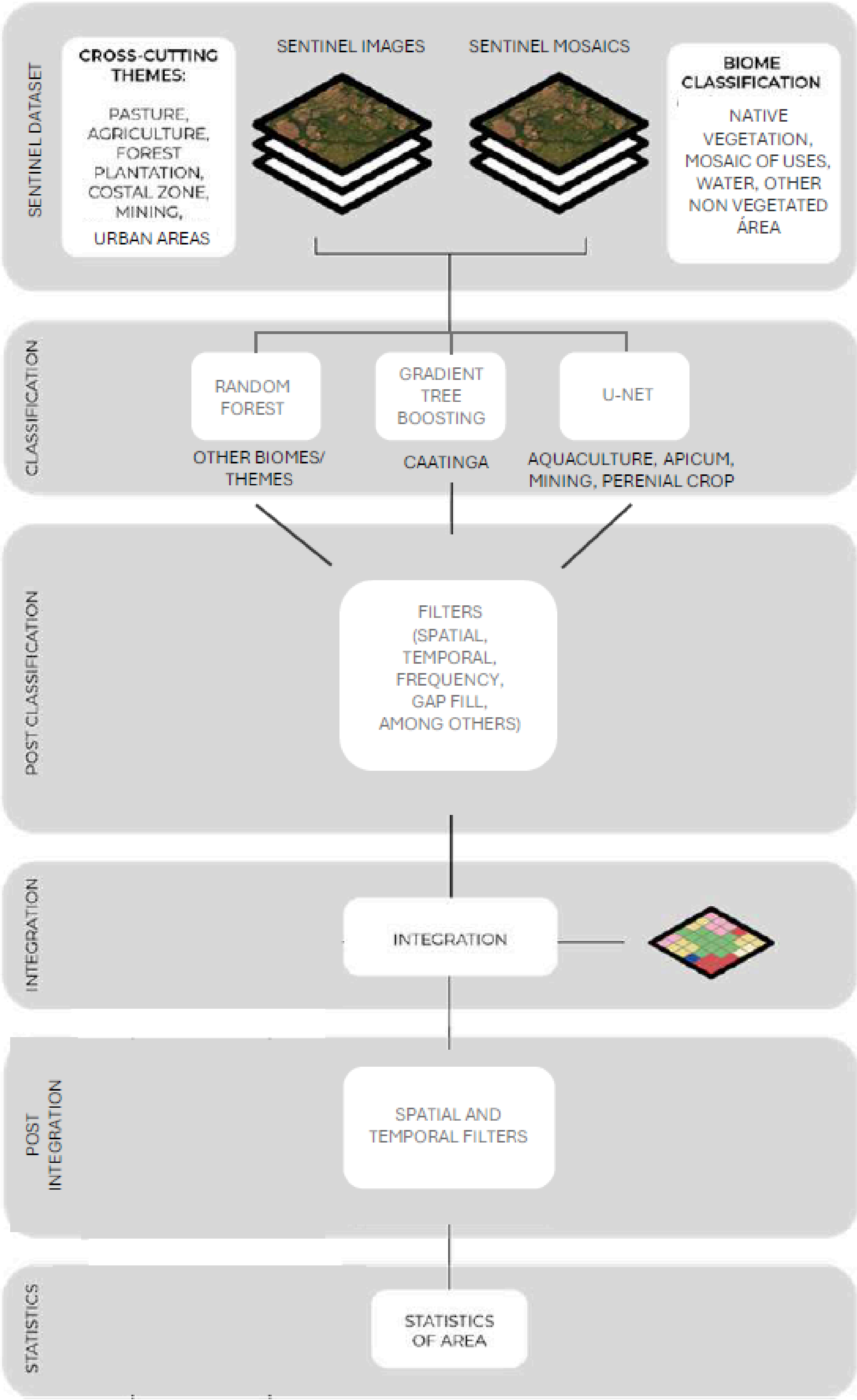

Figure 2. General methodological steps of MapBiomas Brazil 10 m Collection 2 (beta).

3.1. Sentinel-2 Mosaics

All biomes generated cloud-free Sentinel-2 composites based on specific periods of time to optimize the spectral contrast and help with the discrimination of LCLU classes. The cloud/shadow removal script takes advantage of the cloud probability band and the GEE median reducer. The cloud probability can improve data integrity by indicating which pixels might be affected by artifacts or subject to cloud contamination. In conjunction, GEE can be instructed to pick the median pixel value in a stack of images. By doing so, the engine rejects values that are too bright (e.g., clouds) or too dark (e.g., shadows) and picks the median pixel value in each band over time. In the current version, harmonized Collection of Sentinel-2 Level 2A surface reflectance images were used in the classification.

The cross-cutting themes (Pasture, Agriculture, Urban Area, Coastal Zone, and Mining) processed Sentinel-2 mosaics per scene basis. To reduce noise and improve the mosaic quality, a tool was developed to evaluate the images individually, excluding uninformative images (excess cloud cover).

3.2. MapBiomas feature space

The feature space for LCLU classification is composed of original Sentinel-2 bands (blue, green, red, red edge 1, red edge 2, red edge 3, red edge 4, swir 1 and swir 2) and reflectance index (NDVI, NDDI, NDWI, SAVI) per year. In addition, statistical reducers were used to generate temporal features such as:

- Median: median of the pixel values within the defined stack of images
- Median_dry: median of the quartile of the lowest pixel NDVI values
- Median_wet: median of the quartile of the highest pixel NDVI values
- Amplitude: amplitude of variation of the index considering all the year's images
- stdDev: stdDev of the pixel values within the defined stack of images
- Min: the lower annual value of the pixels of each band
- Max: the higher annual value of the pixels of each band

Each biome and cross-cutting theme executed a feature selection algorithm to choose the most appropriate subset of variables to train the respective classifier (e.g. Random Forest, Gradient Tree Boost, U-NET).

3.3. Legend

The MapBiomas 10 m Collection 2 (beta) classification scheme is a hierarchical system comprising three categorical levels (Table 2). At Level 1, there are six classes: 1) Forest, 2) Herbaceous and shrub vegetation, 3) Farming, 4) Non-Vegetated Area, 5) Water, and 6) Not Observed. Level 2 has 20 classes across the six classes of the first categorical level. Agriculture (3.2) is the only class with further subdivisions down to the third categorical level, separating perennial and temporary crops. This collection thus comprises 21 LCLU classes.

Table 2. Classes of land cover and land use of MapBiomas 10 m Collection 2 (beta) in Brazil.

ID	COLLECTION 2 (beta) CLASSES	NATURAL/ ANTHROPIC	LAND COVER/ LAND USER
1	**1. Forest**	NATURAL	COVER
3	1.1. Forest Formation	NATURAL	COVER
4	1.2. Savanna Formation	NATURAL	COVER
5	1.3. Mangrove	NATURAL	COVER
6	1.4. Floodable Forest	NATURAL	COVER
49	1.5. Wooded Sandbank Vegetation	NATURAL	COVER
10	**2. Herbaceous and shrub vegetation**	NATURAL	COVER
11	2.1. Wetland	NATURAL	COVER
12	2.2. Grassland	NATURAL	COVER
32	2.3. Hypersaline Tidal Flat	NATURAL	COVER
29	2.4. Rocky Outcrop	NATURAL	COVER
50	2.5. Herbaceous Sandbank Vegetation	NATURAL	COVER
14	**3. Farming**	ANTHROPIC	USE
15	3.1. Pasture	ANTHROPIC	COVER/USE
18	3.2. Agriculture	ANTHROPIC	USE
19	3.2.1. Temporary Crop	ANTHROPIC	USE
36	3.2.2. Perennial Crop	ANTHROPIC	USE
9	3.3. Forest Plantation	ANTHROPIC	USE
21	3.3. Mosaic of Uses	ANTHROPIC	USE
22	**4. Non Vegetated Area**	NATURAL/ ANTHROPIC	COVER/USE
23	4.1. Beach, Dune, and Sand Spot	NATURAL	COVER
24	4.2. Urban Area	ANTHROPIC	USE
30	4.3. Mining	ANTHROPIC	USE

25	4.4. Other Non Vegetated Areas	NATURAL/ ANTHROPIC	COVER/USE
26	**5. Water**	NATURAL/ ANTHROPIC	COVER/USE
33	5.1. River, Lake and Ocean	NATURAL	COVER
31	5.2. Aquaculture	ANTHROPIC	USE
27	6. Not Observed	NONE	NONE

4. Classification by biomes and cross-cutting themes

Random forest demands the definition of a few parameters, such as the number of trees, a list of variables, and training samples. Each biome and cross-cutting theme map was produced using a particular set of these parameters, variables, and number of training samples.

4.1. Amazon

4.1.1. General map classification algorithm

To generate the time series of land use and land cover of the Amazon biome with 10 m of spatial resolution, we built eight mosaics (2016 - 2023) using a harmonized collection with Surface Reflectance (SR) data from images of satellites Sentinel 2A and 2B. The feature space comprised image bands, vegetation indexes, and spectral mixture analysis (SMA) fractions. All these features are described in Table 3:

Table 3. Feature space used in the Amazon biome land use and land cover mapping in MapBiomas 10 m Collection 2 (beta).

Band or Index	**Metrics**				
	Median	Median (dry season)	Median (wet season)	Standard Deviation	Amplitude
Blue	X	X	X	X	-
Green	X	X	X	X	-
Red	X	X	X	X	-
NIR	X	X	X	X	-
SWIR 1	X	X	X	X	-
SWIR 2	X	X	X	X	-
NDVI	X	X	X	X	X
NDWI	X	X	X	X	X

EVI2	X	X	X	X	X
GV	X	-	-	-	X
NPV	X	X	X	X	X
Soil	X	X	X	X	X
NDFI	X	X	X	X	X
SEFI	X	X	X	X	X
WEFI	X	X	X	X	X

We mapped seven classes using the Random Forest Classifier (RFC): Forest Formation, Savanna Formation, Grassland, Pasture, Agriculture, Other Non-Vegetated Areas, and Water. In the post-classification step, we added three classes: Floodable Forest, Wetlands, and Rocky Outcrop. Floodable Forest and Rocky Outcrop were mapped as new classes in the Amazon in this Collection 2 (beta).

4.1.2. Post-classification

4.1.2.1. Wetlands and Floodable Forest Mapping

We used the Sentinel-2 mosaics and reference maps to sort and stratify samples to map wetlands. To train and calibrate an RFC, we used the Global Ecosystem Dynamics Investigation - GEDI (Potapov et al., 2021), Shuttle Radar Topography Mission - SRTM (Farr et al., 2007), Height Above the Nearest Drainage - HAND (Donchyts et al., 2016), Global Canopy Height (Lang et al., 2023), and SMA fraction imagery dataset (Souza Jr. et al., 2005). The sampled pixels were automatically classified as a binary map, Wetland, and Non-Wetland. We used the trained and calibrated samples to rank the eight annual mosaics. Finally, we analyze all eight annual layers classified as wetlands and apply a maximum reducer to synthesize the layers and define the Maximum Flooded Area (MFA) in the time series for the Amazon biome. Every year, we cross the LCLU map with the MFA layer. When the pixel agrees with Forest Formation and MFA, we remapped it as Floodable Forest. When the pixel agrees with the Savanna or Grassland and the MFA, we remapped it as Wetlands. The MFA mapping is conducted in a Region Of Interest (ROI), defined according to the features used to train and calibrate the RFC. Some features have spatial resolution different from Sentinel-2 data, so in some cases, if the MFA area is greater than ROI, the final mapping can have an appearance of 30 meters of spatial resolution data.

4.1.2.2. Rocky Outcrop Mapping

To map the Rocky Outcrops in the Amazon biome we used the same annual mosaics, plus random stratified samples (with Rocky Outcrop and Non-Outcrop) to train and calibrate an RFC. The steep altitudes and slopes, escarpments, hills and predominantly exposed soil, give a unique spectro-temporal behavior to outcrops. To represent such features, we used fractions derived from spectral mixing models, such as Soil, NPV, GV. Morphological characteristics of the terrain were represented using data such as SRTM and HAND.

4.1.2.3. Filtering

After the classification step, we applied a logic sequence of four filters to improve the RFC results. The filters used are described as follows:

- **Gap Fill Filter**: This filter aims to fill the missing information in the land use and land cover maps caused by cloud cover in the Sentinel-2 mosaics. The filter replaces each pixel that has missing information in the map with its class mode throughout the time series;
- **Spatial Filter:** This filter aims to eliminate isolated pixels lower than one hectare that are highly likely to be misclassification problems.
- **Native Vegetation Stability**: This filter acts in pixels that, during the time series, vary its classification only among native vegetation classes (Forest, Savanna and Grasslands). The shift among these classes has a low probability of occurring in the Amazon biome, and the most common change in natural vegetation areas is the conversion to farming use. This filter identifies this behavior and establishes the classification through the years using the mode of these three classes throughout the period;
- **Temporal Filter**: We applied temporal rules to avoid undesirable transitions during the time series. We identified a few transitions misunderstood by the RFC, mainly in cases when Forest was converted to Pasture. To adjust these cases, we used a three years window to change the classification in the third year, from 2016 to 2023, as shown in Table 4:

Table 4. Temporal filter rules applied in the Amazon biome mapping in the MapBiomas 10 m Collection 2 (beta).

Rule	Class Year 1	Class Year 2	Class Year 3	New Class Year 3
1	Forest	Pasture	Grassland	Pasture
2	Forest	Pasture	Savanna	Pasture
3	Forest	Forest	Grassland	Pasture
4	Forest	Forest	Savanna	Pasture
5	Floodable Forest	Floodable Forest	Savanna	Pasture

4.2. Caatinga

4.2.1. General map classification algorithm

The process of generating the MapBiomas Collection 2 (beta) maps in the Caatinga biome, based on annual Sentinel 2A mosaics with a 10 m spatial resolution, utilized several steps of image classification processes (Figure 3). More details about the improvements made are described in Figure 4.

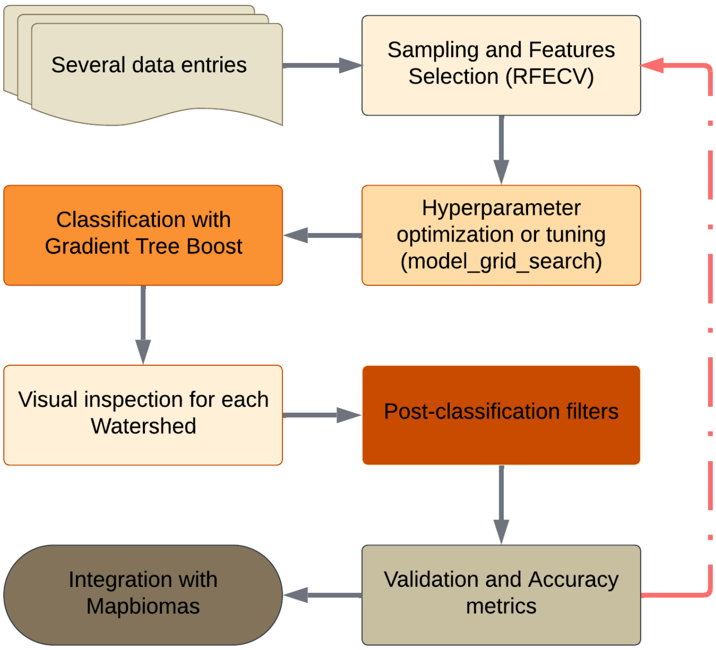

Figure 3. Overview of the main steps in the Caatinga biome classification process in the MapBiomas 10 m Collection 2 (beta).

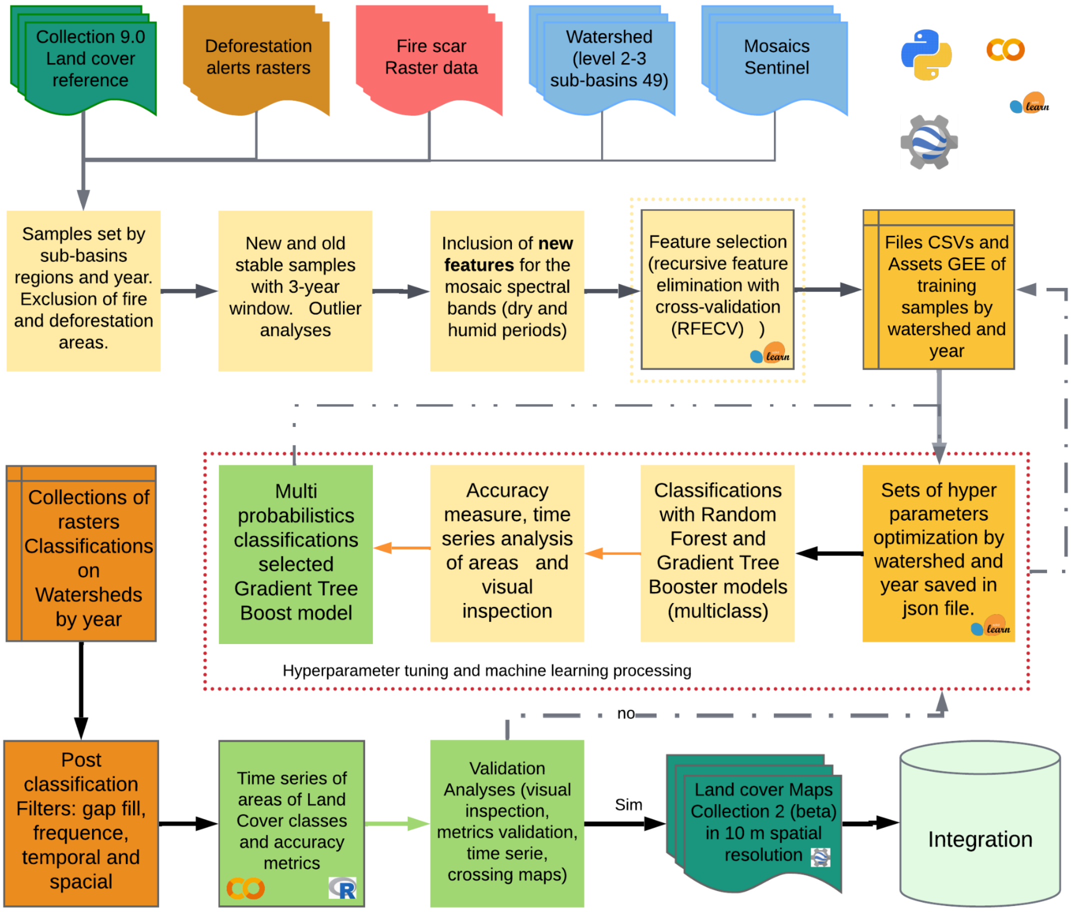

Figure 4. Detailed steps of MapBiomas 10 m Collection 2 (beta) (2016-2023) in the Caatinga biome.

The MapBiomas Caatinga team, in alignment with other project teams, employed a stratified sampling strategy, segmenting the biome into base regions. To enhance the reliability of automatically collected samples, a mask was generated to exclude areas affected by temporal phenomena, such as wildfires and deforestation. Subsequently to the mask application, sample collection was restricted to those pixels maintaining consistent classification in the pre- and post-collection periods, and selecting information from the mosaic bands only within the mask-permitted areas. The resulting samples underwent a feature selection process, wherein spectral bands that optimized classifier performance in land cover class prediction were identified.

The base mosaic comprised six spectral bands from the Sentinel-2 sensor, specifically Blue, Green, Red, NIR, SWIR1, and SWIR2, across three distinct temporal periods: wet, dry, and annual. Each band was identified by a composite name, such as 'blue_median_dry', indicating the spectral band, the statistic (median), and the period (dry). Additionally, several spectral indices were calculated for the mosaics of each period and incorporated as additional bands. The calculated indices are presented in Table 5.

Table 5. Complementary bands added to the Caatinga feature space in the MapBiomas 10 m Collection 2 (beta).

Name	**Formula**	**Descriptions**	**Reference**
RVI	(N * R) / (G ** 2.0)	Ratio Vegetation Index	Jordan (1969)
RATIO	N/R	Ratio	Pearson and Miller (1972)
NDWI	(G - N) / (G + N)	Normalized Difference Water Index	McFeeters (1996)
AWEI	B + 2.5 × G − 1.5 × (N + S1) − 0.25 × S2	Automated Water Extraction Index	Feyisa et al. (2014)
IIA	R/N	Inverse Intensity Index	
EVI	2.5 × (N + R) / (N + 6R - 7.5B + 1)	Enhanced Vegetation Index	Huete et al. (1994)
GCVI	(N/G) - 1	Green Chlorophyll Vegetation Index	Gitelson et al. (2005)
GEMI	(2 X (N^2 - R^2) + 1.5N + 0.5R)/(N + R + 0.5)	Global Environmental Monitoring Index	Pinty and Verstraete (1992)
CVI	(N x R) / (G^2)	Chlorophyll Vegetation Index	Vincini et al. (2008)
GLI	(2G - R - B)/(2G + R + B)	Green Leaf Index	Lourenço et al. (2021)
AVI	(N x (1- R) x (N - R))/ 1000	Advanced Vegetation Index	Loi et al. (2017)
BSI	((S1 + R) - (N + B))/((S1 + R) + (N + B))	Bare Soil Index	Rikimaru et al. (2002)
BRBA		Broadband Reflectance-Based Albedo	Liang (2001)
DSWI5	G / (N + S1 + R)	Dynamic Surface Water Index 5	Fisher et al. (2016)
LSWI	(N - S1)/(N + S1)	Land Surface Water Index	Xiao et al. (2002) -
MBI	((S1 - S2 - N)/(S1 + S2 + N)) + 0.5	Modified Bare Soil Index	Nguyen et al. (2021)
UI	(S2 - N) / (S2 + N)	Urban Index	Kawamura et al. (1997)

OSAVI	(N - R) / (N + R + 0.16)	Optimized Soil Adjusted Vegetation Index	Rondeaux et al. (1996)
RI	(R - G)/(R + G)	Redness Index	*Mathieu et al. (1998)*
BRIGHTNESS		*Tasseled Cap Transformation for Landsat TM*	Crist (1985)
WETNESSS		*Tasseled Cap Transformation for Landsat TM.*	Crist (1985)
GVMI	((N + 0.1) - (S1 + 0.02)) / ((N + 0.1) + (S1 + 0.02))	Global Vegetation Moisture Index	Ceccato et al. (2002)
NIR_CONTRAST	1/14 GLCM metrics proposed by Haralick, Textural Features for Image Classification	CONTRAST NIR bands	Haralick et al. 1973
RED_CONTRAST		CONTRAST RED bands	
NDDI	(NDVI - NDWI) / (NDVI + NDWI)	Normalized Difference Drought Index	Gu et al. (2007)
NDVI	(N - R) / (N + R)	Normalized Difference Vegetation Index	Rouse et al. (1974)

The feature selection process utilized the Recursive Feature Elimination with Cross-Validation (RFECV) methodology. RFECV's goal is to select the most important features, those that contribute the most to the model's performance. The recursive process of RFECV started with all features, training of the model, ranking features by importance, removing the least important, and repeated all steps with cross-validation. The cross-validation part was crucial because it helped in determining the right number of features without overfitting (Chen and Guestrin, 2016).

To optimize the classifier's parameters for maximum accuracy, the samples containing the information from the columns or bands selected in the previous process were used as a basis for hyperparameter tuning, aiming to determine the optimal parameters for each classification region. The method employed was Grid Search, a classic hyperparameter tuning technique used to identify the ideal combination of machine learning model hyperparameters. This technique evaluated the model's performance for each hyperparameter combination, identifying the one with the best performance (e.g., highest accuracy, lowest error). The combination with cross-validation enhanced the robustness of the results, as demonstrated by Bergstra and Bengio (2012).

Gradient Tree Boosting was employed as the classification algorithm. Annual mosaics were classified for the period from 2016 to 2023, encompassing eight land cover and land use (LCLU) classes: Forest Formation, Savanna Formation, Grassland, Mosaic of Uses, Other Non-Vegetated Areas, and Water, aligning with MapBiomas Collection 9.

Gradient Tree Boosting is an ensemble learning method that sequentially aggregates multiple weak decision trees, typically weak, where each subsequent tree corrects the errors of its predecessors (Friedman, 2001; Natekin and Knoll, 2013; Abdi, 2020; Ou et al., 2023). The algorithm takes advantage of the gradient method and the derivatives of the loss function to perform an iterative optimization of the model. This classifier was selected over Random Forest for the Caatinga biome due to its superior performance, attributed to specific characteristics relevant to the region:

- **Sequential Learning**: Each tree is trained to minimize the residual (error) of the previous tree, progressively refining predictions.
- **Customizable Loss Function**: It can be adapted to regression, binary classification, or multiclass classification problems.
- **Regularization**: Hyperparameters such as learning_rate, max_depth, and subsample prevent overfitting by controlling model complexity.
- **Feature Importance**: Identifies which input variables (e.g., spectral bands, vegetation indices) most significantly influence predictions.

The Remote sensing data have complex characteristics that align well with Gradient Boosting strengths:

1. **Nonlinearity**: Captures intricate relationships between spectral bands and target variables (e.g., land cover classes).
2. **High Dimensionality**: Effectively handles datasets with many features, such as spectral bands, indices, and texture metrics.
3. **Imbalanced Data**: Addresses scenarios like detecting deforestation in small areas versus intact forests by weighting classes or using sampling techniques.

Gradient Tree Boosting is a powerful tool for remote sensing due to its flexibility, robustness to noise, and capacity to handle complex datasets. By combining sequential error correction with regularization techniques, it delivers high accuracy in tasks ranging from land cover mapping to environmental monitoring.

4.2.2. Post-classification filtering

The mosaics used in the classification process exhibit noise and pixel gaps, which are more frequent at the beginning of the series with Landsat 5 data. These gaps, predominantly in the early years of the series, are corrected through temporal, frequency, and spatial filters, applied in the following order:

- Gap-Fill Filter: This operation fills pixels with invalid values, using valid pixels from adjacent times (t+1 or t-1), following the chronological sequence of years (ascending or descending).
- Frequency Filter: Applied to natural classes, this filter stabilizes small seasonal changes, especially in savannas and grasslands, which historically show greater confusion in the collections. It works by applying rules that stabilize the time series: pixels classified as savanna in 85% of the series force adjacent pixels of Forest Formation or Grassland Formation to be reclassified as savanna. Similar rules are applied for pixels with a frequency of 90% forest and 80% Grassland Formation.
- Temporal Filter: This filter corrects short-term inconsistencies (1-2 years) in the time series. Pixels with atypical coverages, relative to the environmental pattern, are corrected by a moving window of 3, 4, or 5 pixels.
- Spatial Filter: This filter eliminates "salt and pepper" noise, which are isolated pixels in areas of homogeneous classes. The operation ensures spatial coherence, using the "connectedPixelCount" function of Google Earth Engine (GEE) to assign a pixel the predominant class among its neighbors.

After each cycle of applying these filters, one or more steps may be repeated in the post-classification, to refine the correction."

4.3. Cerrado

4.3.1. General map classification algorithm

In developing the MapBiomas 10 m Collection 2.0 (beta) for the Cerrado biome, we employed a random forest based classification of Sentinel-2 yearly mosaics from 2016 to 2023, encompassing eight land cover and land use (LCLU) classes: Forest Formation, Savanna Formation, Wetland, Grassland, Agriculture, Pasture, Other Non-Vegetated Areas and Water, consistent with the MapBiomas Collection 9. This classification approach not only captures native vegetation (NV) and water bodies but also extends to anthropogenic classes such as Agriculture and Pasture, offering a comprehensive landscape representation of the Cerrado. Notably, Agriculture and Pasture classes are identified in the initial classification stages and later transformed into a “Mosaic of Uses” class during the post-processing step. This approach was essential to mitigate omission or commission errors in the NV classes, thus ensuring classification accuracy. Figure 5 provides an overview of the method for Cerrado native vegetation classification in Collection 2 (beta).

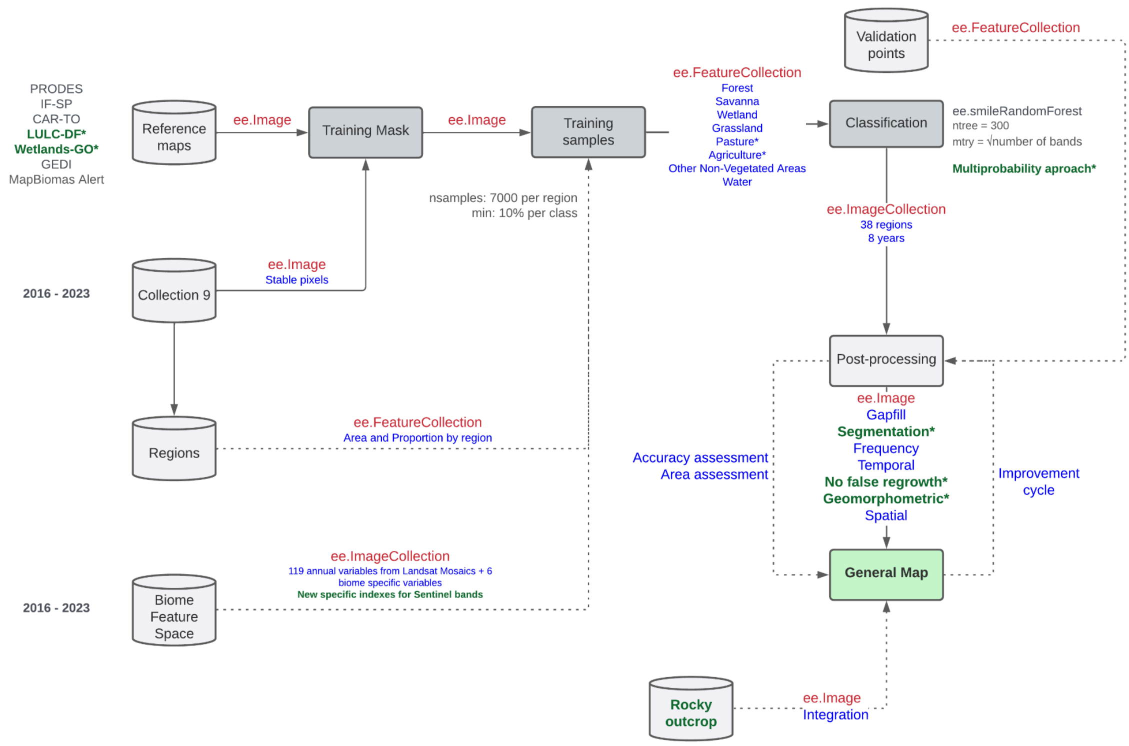

Figure 5. The general map classification flowchart in Collection 2 (beta). Each gray cylinder and rectangle represents a key step in the classification schema. The gray text near databases and processes offers a short description of the step, while the green text highlights the main innovations. Arrows with a continuous black line connecting the key steps represent the main direction of the processing flux, while arrows with dotted black lines represent the databases that feed the main processes. Red text inside arrows refers to the asset type in the Google Earth Engine, while blue text offers a short description of the asset content.

The feature space incorporated both common (all biomes) and biome-specific variables for the Cerrado. Standard variables included annual median reflectance, median values from wet and dry periods, and the standard deviation of annual Sentinel-2 bands (blue, green, red, red-edge 1/2/3, SWIR1, and SWIR2). The Cerrado-specific variables included several vegetation indices from Collection 9, supplemented by new indices based on red-edge bands. Table 6 details these additional variables.

Table 6. Complementary bands added to the Cerrado feature space in the MapBiomas 10 m Collection 2.

Type	Name	Formula	Statistics	Reference
Sentinel 2 -based indexes	Normalized Difference Red Edge Index (NDVI 705)	NDVI 705 = (Red edge 1 - Red) / (Red edge 1 + Red)	median, median_dry, median_wet, stdDev	Gitelson and Merzlyak (1994)
	Vegetation Index 700	VI700 = (Red Edge 1 - Red) / (Red Edge 1 + Red)	median, median_dry, median_wet, stdDev	Gitelson et al. (2002)
	Inverted Red-Edge Chlorophyll Index (IRECI)	IRECI = (Red Edge 3 - Red) / (Red Edge 1 / Red Edge 2)	median, median_dry, median_wet, stdDev	Frampton et al. (2013)
	Chlorophyll Index Red-Edge (CIRE)	CIRE = (NIR / Red Edge 1) - 1	median, median_dry, median_wet, stdDev	Gitelson et al. (2003)
	Transformed Chlorophyll Absorption Reflectance Index (TCARI)	TCARI = 3 * ((Red Edge 1 - Red) - 0.2 * (Red Edge 1 - Green) * (Red Edge 1 / Red))	median, median_dry, median_wet, stdDev	Haboudane et al. (2002)
	Spectral Feature Depth Vegetation Index (SFDVI)	SFDVI = ((NIR + Green) / 2) - ((Red + Red Edge 2) / 2)	median, median_dry, median_wet, stdDev	Baptista (2015)
	Normalized Difference Red Edge Index (NDRE)	NDRE = (NIR+ Red Edge 1) / (NIR−Red Edge 1)	median, median_dry, median_wet, stdDev	Gitelson and Merzlyak (1994)

In addition to standard predictors, the Cerrado feature space was enriched with geolocation (latitude and longitude data) and Time Since the Last Fire (TSLF), calculated as the difference between the current year and the year of the last fire event (Alencar et al., 2022). Geomorphometric data were sourced from Geomorpho 90m dataset (Amatulli et al., 2020), which provided elevation, aspect, slope, ruggedness, east-west second order partial derivative, north-south second order partial derivative, profile curvature, tangential curvature, eastness, northness. We also incorporated the ANADEM/UFRGS (Laipelt et al., 2024) and the Multi-Error-Removed Improved-Terrain DEM (MERIT; Yamazaki et al., 2017) datasets for elevation and slope. This extensive feature set aimed to capture the spectral, temporal, and contextual complexities of the Cerrado.

The Cerrado biome was divided into 38 classification regions in order to represent its environmental heterogeneity and improve the classification. The delineation of these areas was based on seasonal regional variation in vegetation based on normalized difference vegetation index (NDVI), as well as the 19 ecoregions proposed by Sano et al. (2019). Training samples for each one of the 38 classification regions were based on stable areas from Collection 9, enhanced by reference maps for native vegetation (state-level data from São Paulo, Tocantins, Federal District and Goiás) and deforestation (PRODES and MapBiomas Alerta) along with a GEDI-based methodology to exclude outliers. Cerrado's annual classification was performed using the Random Forest algorithm, implemented in the 'ee.Classifier.smileRandomForest' function with the output mode set to "Multiprobability" on the Google Earth Engine (GEE) platform.

4.3.2. Post-classification filtering

The pixel-based classification method required post-classification spatial and temporal filtering to better consistency and reduce errors. This process included multiple filters as follows:

- Gap-Fill filter: The Gap-Fill Filter addressed temporal data gaps from cloud cover and shadows, filling no-data values with the nearest valid temporal classification. If no future valid classification was available, the preceding classification was used.

- Segmentation filter: The segmentation process combined spectral, spatial, textural, and contextual information to refine LCLU classification. Using a Sentinel-2 mosaic (SWIR1, NIR, and Red bands) for each year from 2016 to 2023, we applied the Simple Non-Iterative Clustering (SNIC) algorithm. This algorithm was set to a 5-pixel grid from Image. Segmentation.seedGrid to aggregate similar pixels based on spectral characteristics and texture within local neighborhoods, improving spatial coherence by creating homogeneous segments. Each segment was assigned the most frequent LCLU class within it (mode filter) to align the classification with segment boundaries, thus reducing noise and enhancing overall classification consistency.

- Frequency filter: Frequency filtering was applied to stabilize the NV classes by applying criteria to pixels classified as NV in at least 85% of the time series. Subsequently, specific criteria were used for each native vegetation class to achieve stability. In cases where the Forest Formation class was present for more than 75% of the time series, this class was confirmed for the pixel in question for the entire time series. For Wetland, a minimum frequency of 85% was required, while for Savanna Formation and Grassland, the minimum frequency was 40% and 50%, respectively. These values were empirically set to achieve a more stable classification of the native vegetation, minimizing the uncertainties associated with temporal fluctuations.

- Temporal filter: The temporal filter ensures consistent LCLU change analysis by minimizing classification errors due to invalid transitions. In this step, Agriculture and Pasture classes were reclassified to “Mosaic of Uses” (21) to filter farming areas as a unique class. Then, it followed a series of sequential steps. The first step consisted of a 3-year moving window from 2017 to 2022 (excluding first and last years) that corrected for all intermediate years, considering previous and subsequent years (-1 and +1 years). Each transition was evaluated according to an order of priority, being: Savanna Formation (4), Grassland Formation (12), Forest Formation (3), Wetland (11), Mosaic of Uses (21), Other Non-vegetated Areas (25) and River, Lake, and Ocean (33). The second step involves checking the values of pixels that were not classified as Mosaic of Uses (21) in 2023 (last year) but were classified as such in 2022 and 2021. The value in 2023 is corrected to be consistent with previous years to avoid uncorrected regeneration in the recent year. Finally, the filter verified the regeneration of native vegetation (NV) in the last year. Pixels indicating regeneration between 2022 and 2023 were evaluated, and areas smaller than 1 ha were discarded to ensure classification consistency.

- No false-regrowth filter: The false-regrowth filter was applied exclusively to the Forest Formation (3) class. This filter prevents artificial Forest Formation expansions in silviculture areas, identifying abrupt transitions to Forest Formation as potential classification errors. Pixels that were initially classified as “Mosaic of Uses” (21) in 2016-2017 but were subsequently classified as Forest Formation in the following years were adjusted to retain the anthropogenic designation. This procedure guarantees that the artificial expansion of forest areas was kept to a minimum, thus providing a more precise representation of land use and land cover dynamics.

- Geomorphometric filter: The geomorphometric filter was applied only to the Wetland class (11) to enhance classification consistency by mitigating erroneous classifications in areas characterized by unsuitable terrain conditions. This filter removed Wetland pixels located in regions with slopes exceeding 10 degrees. Pixels within these conditions were remapped to the most frequent neighboring LCLU cover class, considering a kernel of 24 pixels.

- Spatial filter: The spatial filter implemented in this collection was similar to the 30 m Collection 9 and enhances spatial consistency along pixel boundaries. The "connectedPixelCount" function, inherent to the Google Earth Engine platform, was employed to identify connected components (neighbors) sharing the same pixel value. The spatial filter established a minimum connection value of six adjacent pixels, corresponding to a minimum mappable area of approximately 0.54 hectares.

Figure 6 illustrates the percentage of classification changes due to post-processing filters. For the Collection 2 (beta), on average, 89% of the pixels remained the same throughout the annual processing flow, while 11% of the pixels were modified at some stage of post-processing. Among the applied filters, the segmentation filter was the one that most modifies the classification, affecting 4% of the pixels per year, followed by the frequency filter, which modifies about 2% of the pixels. Thus, post-processing filters increased the spatial and temporal consistency of land cover and land use data, while keeping fewer changes in the final maps.

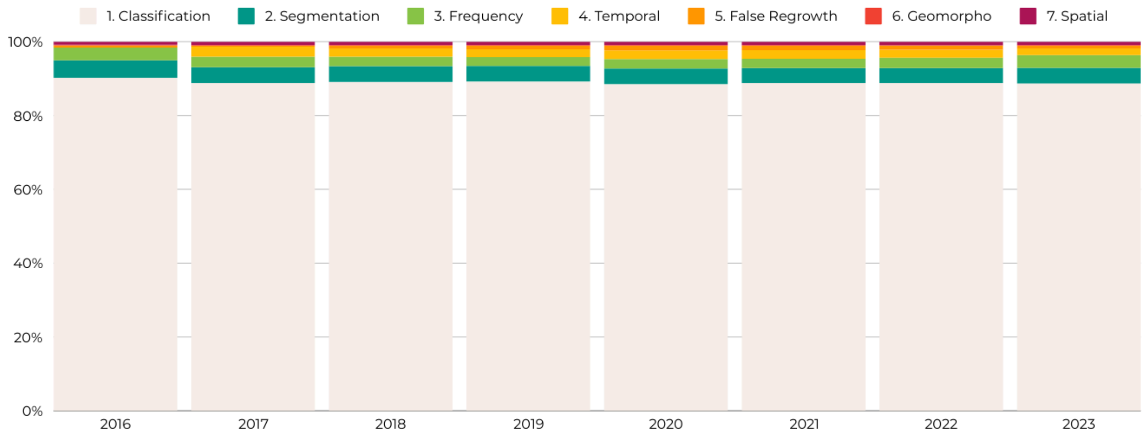

Figure 6. Annual percentage of pixel changes per post-classification filtering step in the Cerrado classification in the MapBiomas 10 m Collection 2 (beta).

4.3.3. Rocky outcrop map classification algorithm

The classification process for rocky outcrop followed a distinct approach compared to that used for the Cerrado's general map, aiming to avoid overestimation of the extent of rocky outcrop areas. The rocky outcrop was a new class in MapBiomas 10 m Collection 2 for the Cerrado biome. The classification flowchart is shown in Figure 7. The feature space used in this classification include the same spectral bands and indexes described in section 4.3.1. However, three additional terrain-related predictor variables were included: relative relief (the difference between the highest and lowest contour values in a given area), valley depth (the elevation difference between a valley and its upstream ridge), and the topographic position index (TPI; which identifies topographic slope positions) (Ganerød et al., 2023).

Training samples included those visually collected by an interpreter, as well as samples provided by the Brazilian Geological Service (SGB/CPRM), subsequently validated by the interpreter. In total, 2,868 samples were collected. The model was trained in Google Earth Engine using the function ee.Classifier.smileRandomForest with ntree = 300, and applied to the Sentinel mosaics through a "Multiprobability" approach, similar to the general classification workflow.

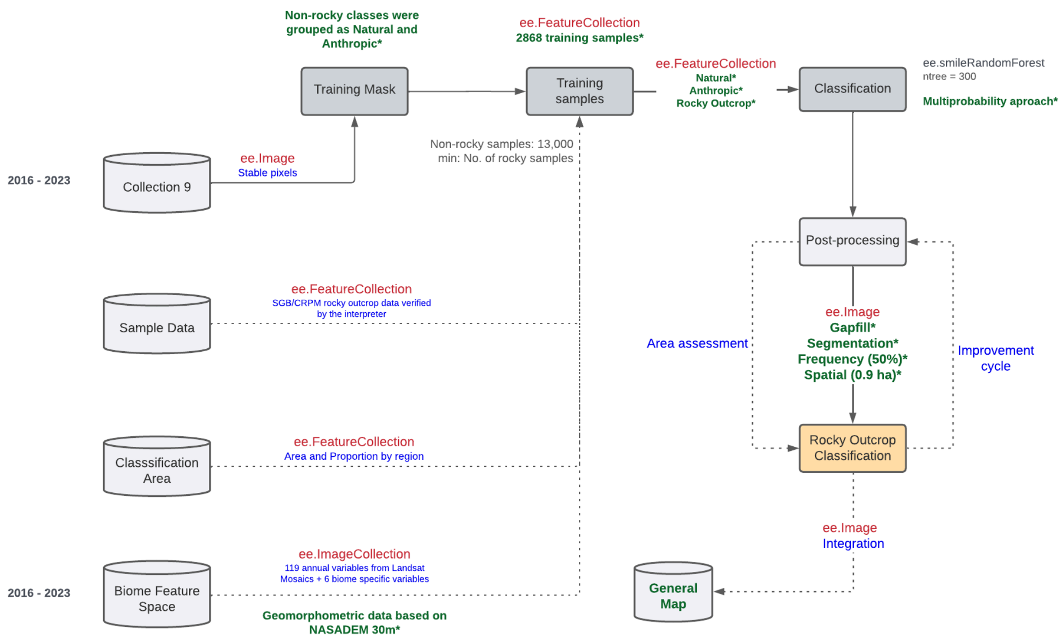

Figure 7. The rocky outcrop classification steps in the Cerrado of MapBiomas 10 m Collection 2 (beta). Each gray cylinder and rectangle represents a key step in the classification schema. The gray text near databases and processes offers a short description of the step, while the green text highlights the main innovations. Arrows with a continuous black line connecting the key steps represent the main direction of the processing flux, while arrows with dotted black lines represent the databases that feed the main processes. Red text inside arrows refers to the asset type in the Google Earth Engine, while blue text offers a short description of the asset content.

Post-processing filters were implemented in the rocky outcrop classification to ensure consistency, following description above:

- Gap-Fill filter: A temporal filter that uses classifications from adjacent years to fill pixels with missing data, ensuring continuity and consistency for rocky outcrop classification.

- Segmentation filter: This filter was set to a 5-pixel grid from Image.Segmentation.seedGrid to aggregate similar pixels based on spectral characteristics and texture within local neighborhoods, improving spatial coherence by creating homogeneous segments.

- Frequency filter: Aims to regulate the rocky outcrop class over time, given that this class does not exhibit significant land cover and land use changes. Consequently, a pixel was classified as a rocky outcrop if it was present in at least 50% of the time series observations.

- Spatial filter: Used to remove spurious isolated pixels that may appear in the classification. This filter eliminated isolated pixels based on a connectivity threshold of 10 pixels, equivalent to an area of 0.9 ha.

After classification and post-processing, the rocky outcrop classification layer is superimposed on the land use and land cover maps in the Cerrado.

4.4. Atlantic Forest

4.4.1. General Map Classification Algorithm

The MapBiomas 10 m Collection 2 (beta) LCLU dataset for the Atlantic Forest biome classified Sentinel-2 annual mosaics from 2016 to 2023 using the Random Forest algorithm (Figure 8).

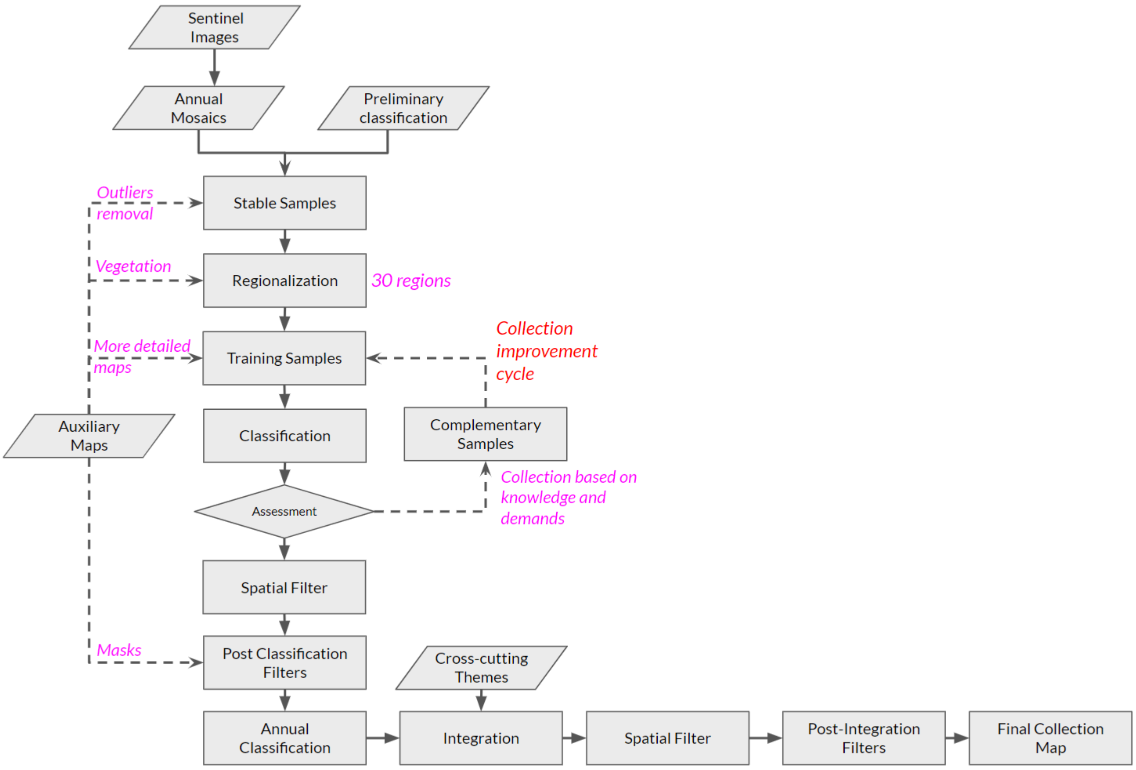

Figure 8. Classification process of MapBiomas 10 m Collection 2 (beta) in the Atlantic Forest biome. Inclined gray rectangles represents databases, while linear rectangles point key-steps in

the workflow. Solid arrows indicate main flow, and dashed arrows point to assessment and evaluation cycles. Pink and red text shows additional information for specific steps.

The LCLU classes mapped with Random Forest classification in the Atlantic Forest were: Forest Formation, Savanna Formation, Grassland, Mosaic of Uses, Non-vegetated areas, Rock Outcrop, Forest Plantation, and Water. Additionally, an extra classification for the Agriculture class was carried out in specific regions of the biome: Southeast and Northeast. With the defined regions, samples were collected in agricultural areas and then classified following the same process as the other classes using Random Forest. The Agriculture and Forest Plantation classification aimed to reduce potential errors in the classification of natural areas and minimize omissions. Later, these two classes were converted to class 21 (Mosaic of uses) in the final Atlantic Forest dataset.

The classification was performed in homogeneous regions to reduce sample and class confusion and to achieve a better balance between samples and results. The biome was divided into 30 regions based on native vegetation types in the Atlantic Forest (IBGE, 2021). For each region, the number of stable samples used in the classification was defined, considering the relevance of each class in each region. Classes were assigned as main, secondary, or rare.

In the classification process, stable samples based on Sentinel data were used for the first time. Additionally, Global Forest Canopy Height (GFCH) data based on GEDI was used to filter stable areas for two classes: Forest Formation (height ≥ 8) and Grassland (height ≤ 7). Reference maps of the states of São Paulo (IF, 2020), Minas Gerais (IEF), Paraná (IAT, 2020) https://geopr.iat.pr.gov.br/portal/apps/dashboards/1eca83bf72e44193ae62f282574da52f), and Espírito Santo (IEMA, 2015 - https://geobases.es.gov.br/links-para-mapes1215) were also used as sources to filter stable samples in natural areas. All reference maps are available on the MapBiomas Brazil website (https://brasil.mapbiomas.org/en/mapas-de-referencia/).

The feature space used to classify the Atlantic Forest biome comprised a subset of the 60 most relevant bands for each region, out of a group of 140 variables. These variables include original Sentinel reflectance bands, vegetation indices, spectral mixture modeling-derived variables, and terrain morphometry (e.g., slope). Additionally, bands resulting from the image segmentation function (ee.Algorithms.Image.Segmentation.SNIC) with the original median bands were used. This function generated a "clusters" band, "clusters_green_text" band, and "clusters_ndfi_median" band for each year. The definition of the subset was made based on a feature importance analysis produced with Random Forest classification using all bands and 500 interactions. All codes used are available on our public GitHub (https://github.com/mapbiomas/brazil-atlantic-forest).

The need for complementary samples was evaluated through visual inspection and by comparing the preliminary accuracy outputs for each region. Complementary sample collection was also done by drawing polygons using the Google Earth Engine Code Editor. The same concept of stable samples was applied, checking the false-color composites of the Sentinel mosaics for all 8 years while drawing polygons. Based on expert knowledge of each region, polygon samples from each class were collected, and the number of random points within these polygons was defined to balance the samples.

4.4.2. Wetland Classification

To reduce confusion with other natural vegetation classes and avoid temporal variations that do not exist in the wetland class, a distinct classification was carried out for wetlands. This classification followed the same steps as the classification of other classes using Random Forest, but it considered the Height Above Nearest Drainage (HAND) product as a proxy to represent groundwater depth and one classification for all the biome.

The result of the classification was added to the land use and land cover maps of the biome and underwent the same post-classification filters. Wetland areas were only overlaid on pixels mapped as class 21 (Mosaic of uses) in the main classification.

4.4.3. Post Classification

Due to the pixel-based classification method and the extended temporal series, a list of post-classification spatial and temporal filters was applied as follows:

- Gap-Fill Filter: No-data values (gaps) are theoretically not allowed and were replaced by the temporally nearest valid classification. In this procedure, if no "future" valid position was available, the no-data value is replaced by its previous valid class. Therefore, gaps should only exist if a given pixel has been permanently classified as no-data throughout the entire temporal domain.

- Spatial Filter: The spatial filter avoids unwanted modifications to the edges of pixel groups (blobs). A spatial filter was built based on the "connectedPixelCount" function. Native to the GEE platform, this function locates connected components (neighbors) that share the same pixel value. Thus, only pixels that do not share connections to a predefined number of identical neighbors are considered isolated. In this filter, at least 25 connected pixels are needed to reach the minimum connection value. Consequently, the minimum mapping unit is directly affected by the spatial filter applied, which was defined as 25 pixels (~0.25 ha).

- Temporal Filter: The temporal filter uses subsequent years to replace pixels that have invalid transitions. The first process checks any native vegetation class (3, 4, 11, 12, 29) that is not this class in 2016 but is equal in 2017, 2018, and 2019, and then corrects 2016 values to avoid regeneration in the first year. The second process looks for pixels in 2023 that are equal to 21 (Mosaic of Uses) but are not 21 in 2022. The value in 2023 is then converted to natural vegetation to avoid any regeneration in the last year. These two processes only occur in areas where the change is greater than 1 hectare. The same rule is applied to avoid deforestation (less than 1 ha) in the first and last year of the series.

- Frequency Filter: Frequency filters were applied only to pixels considered "stable native vegetation." In stable native vegetation all years are changed to the most common class. The result of these frequency filters is a classification with no change between native classes (e.g., Forest and Savanna). Another frequency filter affects transitions in natural vegetation classes in the middle of the historical series. Considering that the class (3, 4, 11, 12, 29) occurs in the first and last years of the series, any classification of Mosaic of Uses (class 21) in intermediate years will be converted to natural vegetation. Similarly, if natural vegetation does not occur in the first and last years but appears in intermediate years, it is remapped to 21 (Mosaic of Uses), thus avoiding false regenerations and deforestation.

- Wooded Sandbank Vegetation Classification: This class was mapped from the post-classification. The ALOS DSM Global 30m was used to identify coastal forest areas with less than 25m altitude, which were converted to this class using a spatial mask to exclude certain regions in the northeast of Brazil.

- Herbaceous Sandbank Vegetation Classification: This class was mapped from the post-classification. The IBGE Soil Map (IBGE, 2021) was used as a reference. Class 13 (Other non-forest formations), classified through the general process using Random Forest, was reclassified into this category in areas with Spodosols (Podzol) and Neosols soils.

4.5. Pampa

4.5.1. General map classification method

The MapBiomas 10 m Collection 2 (beta) for the Pampa biome was developed using the Random Forest machine learning algorithm for annual Sentinel-2 mosaics from 2016 to 2023. For the classification, seven LCLU were considered, including: Forest Formation, Wetland, Grassland, Mosaic of Uses, Non-Vegetated Areas, Rocky Outcrop and Water. Subsequently, during the post-classification stage, two other classes were added: Wooded and Herbaceous Sandbank Vegetation, through reclassification of classes Forest Formation and Grassland, respectively, whenever located over sandy soils in the coastal zone.

The classification process involved three main stages: generating training samples, incorporating variables into the feature space and classifying the mosaics (Figure 9). Random training samples were generated based on the 30-m MapBiomas 9 Collection, sampling only those areas that have remained in the same class from 1999 to 2023 (stable areas).

The classification was carried out independently in seven homogeneous regions within the biome, an adaptation of the former nine ecological systems proposed by Hasenack *et al*. (2010) for the Brazilian Pampa biome, using vegetation, relief and soils data. For each region we generated 2,000 samples for each class.

Annual image mosaics were produced based on annual median reflectance and the annual standard deviation of Sentinel-2 bands calculated over all the images available in a previously defined "optimal period" within the year (September-November) and median values of the wet (scenes with higher NDVI values) and the dry periods (scenes with lower NDVI values) along the year. We also included spectral indexes supplemented with geolocation (latitude and longitude), the ANADEM/UFRGS slope datasets and the Global Height Above the Nearest Drainage (HAND) (Donchyts, et al., 2016). Finally, two segmentation bands generated with the Simple Non-Iterative Clustering (SNIC) algorithm were added to the feature space (Table 7). The median values of the NDVI and EVI bands and a grid of 20 and 80 pixels as the location spacing of the superpixel seed, respectively, were used to cluster pixels with similar spectral characteristics. The average values per cluster resulting from the NDVI and EVI bands were added to the feature space.

The final feature space comprised 145 variables. However, from this dataset, we used only those considered as the most important for classification (70 in total). This selection took into account the most important variables considering the performance within distinct regions of the Pampa and the ones that showed high importance in most years of the collection in a previous classification. All codes used are available on our public GitHub (https://github.com/mapbiomas/brazil-pampa).

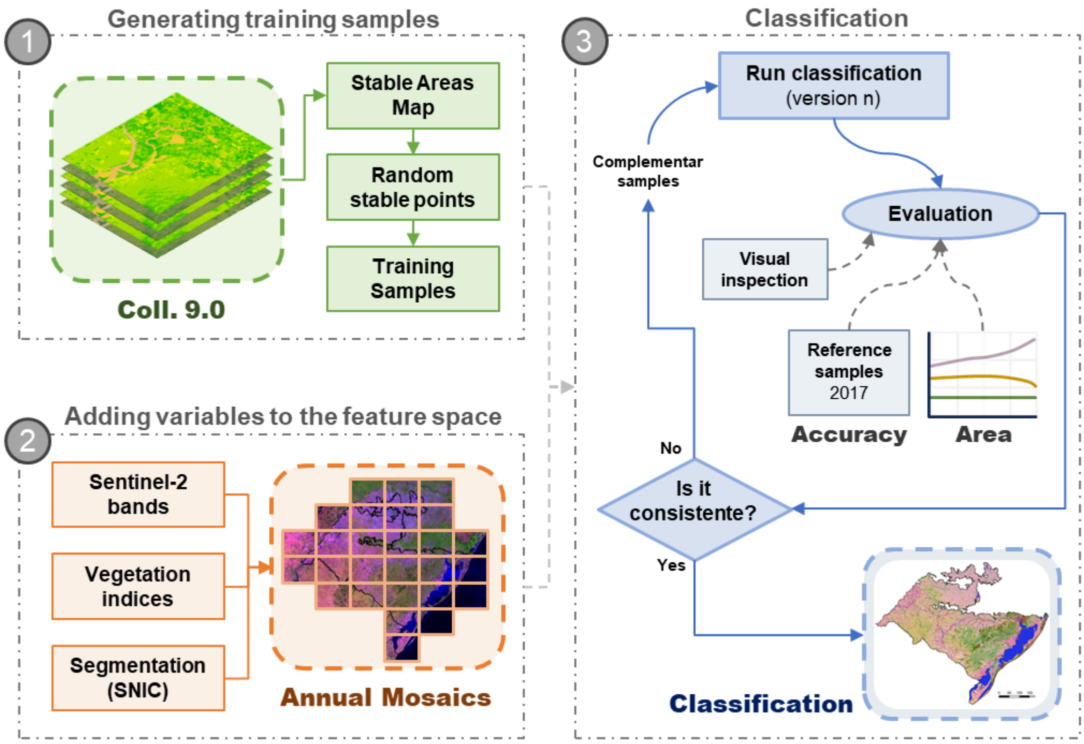

Figure 9. Flowchart outlining the main steps in the Pampa biome classification process of Collection 2 MapBiomas 10 m.

Table 7. Bands added to the Pampa feature space in the Collection 2 MapBiomas 10 m.

Band	Stats	Name
avi	median dry	Advanced Vegetation Index
brba	median, median dry and wet	Band Ratio for Built-up Area
blue band	median	
brightness	median	Tasselled Cap - brightness
bsi	median	Bare Soil Index
cvi	median	Chlorophyll Vegetation Index
dswi5	median wet	Disease-Water Stress Index 5
EVI	median, median dry and wet	Enhanced vegetation index
EVI cluster band	median	Enhanced vegetation index
gcvi	median dry and wet	Green Chlorophyll Vegetation Index
gemi	median, median dry	Global Environment Monitoring Index
gli	median wet	Green Leaf Index
green band	min	
green texture	median	
gvmi	median, median dry and wet	Global Vegetation Moisture Index
hand		Heigth Above Nearest Drainage (GENA)
iia	median, median dry and wet	Indicator of water index
lai	median	Leaf Area Index
latitude		
longitude		
lswi	median dry and wet	Land Surface Water Index
mbi	median, median dry and wet	Modified Bare Soil Index

msi	median dry	Moisture Stress Index
nddi	median and median dry	Normalized Difference Drought Index
NDVI	median, median dry and wet	Normalized Difference Vegetation Index
NDVI cluster band	median	Normalized Difference Vegetation Index
ndwi	median wet	Normalized Difference Water Index
osavi	median wet	Optimized Soil-Adjusted Vegetation Index
ratio	median, median dry and wet	Ratio Vegetation Index
red band	median dry	
red edge 1 band	stdDev, median dry and wet	
red edge 3 band	median wet	
red edge 4 band	stdDev and median dry	
ri	median	Normalized Difference Red/Green Redness Index
rvi	median	Relative Vigor Index
shape	median and median wet	Shape Index
slope		Slope (ANA)
spri	median	Photochemical Reflectance Index
swir 1 band		
swir 2 band		
ui	median and median dry	Urban Index
wetness	median and median dry	Tasselled Cap - wetness

The number of samples of each class to be used in the classification process was established through a temporal weighted balance. The general idea was that the sample size of each class must be proportional to the area occupied by the class *i* in the year *j*. Each class weight was calculated as proportion and multiplied by the number of 2,000 points available in the training data set to establish the number of samples to use in the classification of each specific year. To calculate these annual weights we first converted the class area values observed in MapBiomas Brazil Collection 9 land cover and land use classification, for each year, to relative proportions of the region of interest. Then, we fitted a linear regression, for each class, considering the relative proportions (y) along the 8 years (x) and extracted the intercept (b_0) and the slope of the regression line (b_1). For the year to be classified, each class weight (dependent variable) was calculated using the year as the independent variable. These weights corresponded to proportions (0-1) for each class that were multiplied by the total available samples to set the number of samples in use. We also set a minimum sample size of 100 training points to ensure sufficient representation for those classes with lowest area proportion within the regions.

The final classification process involved repeated classifications with the evaluation of the results using a set of reference samples collected for 2017, visual contrast to Sentinel mosaics and the area of each mapped class. According to the need, in each reclassification, adjustments were made to the number of samples per class and also the addition of complementary samples.

4.5.2. Post classification

The classification was post-processed using seven different filters designed to correct the residual classification errors.

- Gap-fill filter: Designed to fill in pixels classified as Not Observed in a given year. Filling was based on a forward procedure, repeating the information from the previous year. In cases of no-data values remaining in the classification, a backward procedure was carried out, using the following year's information to fill in the previous year.
- Spatial filter: This filter used a mask to modify isolated pixels or very small patches (less than six pixels) of a class, replacing each one with the most frequent value in the eight corresponding neighbors. The filter used the "connectedPixelCount" function in Google Earth Engine and produced a result where the minimum mapped area was a patch with at least six pixels of the same class (~0.54 ha).
- Temporal filter: The temporal filter used information from previous and subsequent years to identify and correct pixel misclassification in a given year, eliminating transitions considered invalid. The rules were different for the first, last and intermediate years of the collection. The process began by analyzing the first three years of the collection, comparing the class of 2016 with those of the following two years. Every pixel classified with a certain class in the first year (2016) and assigned to another class in the following two years (2017 and 2018) was reclassified to the class of the subsequent years. For the last three years, it compared the year 2023 with the previous two and, whenever a pixel was classified as 21, 29 or 33 in both years, but was different in the last year, then it was replaced by the same class as the previous ones. Both procedures aimed to avoid cases of false positives. The last step applied a 3-year moving window to correct the remaining intermediate years. Whenever the first and the third year of the window had the same class and the middle year was different, it was replaced by their class. This procedure had the purpose of fixing abrupt transitions that were unlikely to happen. The filter was applied, step by step, respecting the following sequence of classes: [29, 22, 21, 11, 3, 12, 33]. In addition, a filter for intermediate years was developed to avoid false transitions between natural classes.
- Frequency filter: Different frequency filters were developed to correct cases of false positives, using information from each pixel over the years. Their general logic consisted of searching for a specific combination of classes for each pixel over the 8 years, producing a subset of pixels considered eligible for correction. The filter then detected and replaced only the years in which cases were considered potential false positives, using a fixed class value, which was usually the mode of the classifications detected along the temporal range. This type of filter was used with restraint to solve only very specific cases. The filters involved correcting misclassifications of wetlands in other classes, confusion of rice with temporary water or wetlands, confusion of rocky outcrops and other

non-vegetated areas, false positives of water and wetlands on shaded slopes in regions with undulating relief and confusion between agriculture and rocky outcrops and agriculture and non-vegetated areas.

- Particular cases filter: This filter was developed to correct false positives of water in the 2023 classification year due to excessive rainfall amount. Therefore, temporarily submerged areas classified as water in 2023 were reclassified to the same class in 2022.
- Time-series start/end filter: This filter smoothed the transitions between the penultimate and final years of the time series. Evaluations of the collection showed an unexpected increase in anthropic classes in the last year and a decrease in natural classes. Therefore, this filter was developed to smooth out this abrupt transition, avoiding all transitions from natural areas to anthropic areas, and vice versa, in patches equal to or smaller than 2 hectares. In these cases, the pixels corresponding to the last year received the same classification as the penultimate year.
- Ending filter: The same 3-year temporal filter described above was applied to remove accidental and unwanted effects in the filtered classification resulting from the combined application of the filters.

4.6. Pantanal

4.6.1. General map classification algorithm

In the development of the MapBiomas 10 m Collection 2 (beta) for the Pantanal biome, classification based on Random Forest was used. The classification was performed on annual mosaics from Sentinel-2 from 2016 to 2023. The land use and land cover classes mapped included Forest Formation, Savanna Formation, Wetland, Grassland, Agriculture, Pasture, Other Non-Vegetated Areas, Rocky Outcrops, and Water. The Rocky Outcrop class was added after a binary classification to identify these specific areas. Wetland and Water were added during the post-classification stage through reclassification of classes based on a threshold of maximum flooded area using a humidity index.

The Pantanal biome was divided into seven regions based on drought and flooding patterns, sub-basin watersheds, and the distribution of native vegetation as presented in different regionalization approaches (Silva and Abdon, 1998; Assine, 2015). The goal of this process was to reduce confusion and noise in the classification, improving sample balance in more homogeneous regions.

Training samples for each of the seven classification regions were based on stable areas from the first MapBiomas 10 m Collection (beta), based on the

MapBiomas Brazil Collection 9 sampling areas that remained stable between 2013 and 2023. The classification process involved three main stages: generating training samples, incorporating variables into the feature space and classifying the mosaics.

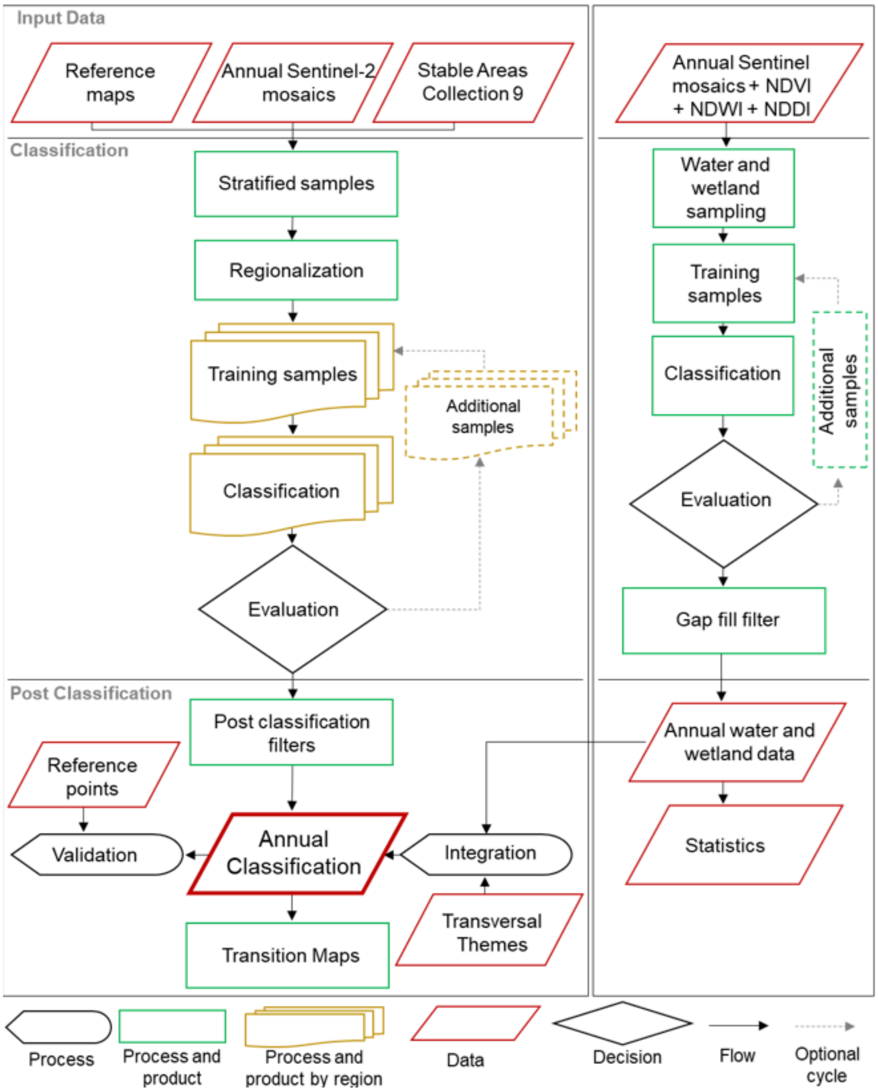

Figure 10. Flowchart outlining the main steps in the Pantanal biome classification process of Collection 2 MapBiomas 10 m.

The variables included annual median reflectance, median values for dry and wet periods, and the standard deviation of the annual bands of Sentinel-2 (blue, green, red, red-edge 1/2/3, SWIR1, and SWIR2) and some mainly indexes as 'evi', ndvi' and 'ndwi'. Additionally, bands resulting from the image segmentation function (ee.Algorithms.Image.Segmentation.SNIC) with the original median bands were used. This function generated a "clusters" band, "clusters_green_text" band, and "clusters_ndfi_median" band for each year. All codes used are available on our public GitHub (https://github.com/mapbiomas/brazil-pantanal).

4.6.2. Post-classification

4.6.2.1. Masks

Some natural grassland areas in the western portion of the Pantanal can be easily confused and classified as exotic pastures by RF because of the overgrazing or the soils exposition during a dry period. To avoid this misclassification, a mask of 'non-pasture' was manually drawn based on historical reference maps (CI et al., 2009), PRODES data (Assis et al., 2019) and high resolution images, and then applied for excluding pixels misclassified as pasture. For the same purpose, a second 'non-pasture' mask was applied considering flooded pixels in more than 75% of the entire Collection.

Also, in order to improve the classification for the most recent years, a mask of 'deforested areas' was built from the MapBiomas Alertas validated polygons. Following, all native vegetation pixels intercepting the mask were reclassified to farming class in the deforestation event year and in the following years. It is worth remembering that these deforestation polygons are available from 2019.

4.6.2.2. Water and Wetland data integration

To generate Water and Wetland data, the Normalized Difference Dynamic Index (NDDI) was calculated for the dry and wet periods of the annual mosaic. Once a threshold has been defined, it differentiates flooded areas from the water surface and non-flooded areas. The intention with this processing is to map the maximum flooded area each year, considering that the Pantanal is one of the largest wetlands in the world subject to a significant flood pulse that varies intra-annually and multi-annually.

This annual water and wetland data were added to the map of 'non-wet' classes only in areas classified as grasslands, considering that the methodology for the latter is not adapted to identify flooded forest and savannah. This is also a strategy to avoid false positives from humid areas in the shade of the relief, urban areas or roads.

4.6.2.3. Rocky outcrop data integration

Rocky outcrops are very limited in the Pantanal biome, primarily concentrated in the Serra do Amolar region. Due to their rarity and spatially isolated nature, a supervised binary classification was conducted specifically for this class, separate from the broader classifications of forest, savanna, and pasture. Based on manually collected training samples derived from high-resolution imagery (Planet and Google Images), areas were classified as either 'rocky outcrop' or 'non-rocky outcrop'. This classification of rocky outcrop areas was then integrated into the existing land use and land cover map, overlaying areas previously identified as grasslands.

4.6.2.4. Filtering

To enhance the temporal consistency of the mapping and minimize errors from the classification method, several filters were applied:

- Gap Fill filter: In this filter, no-data values ("gaps") were theoretically not allowed and were replaced by the temporally nearest valid classification. In this procedure, if no "future" valid position was available, then the no-data value was replaced by its previous valid class. Therefore, gaps should only exist if a given pixel was permanently classified as no-data throughout the entire temporal domain.
- Trajectory filter: To avoid false transitions between savanna formation and grassland in short time intervals, a filter was applied that considered the trajectory of the pixel according to Pontius (2022). Considering the trajectory of absence - alternation - absence, the number of changes that the class was involved in and also the pixel mode along the time series, the filter stabilizes these two classes. This filter was also implemented to prevent areas of Forest or Savanna Formations that were converted to anthropic uses from regenerating as Grassland. It addressed part of the confusion between pasture areas and Grassland. Specifically, the filter ensured that any area converted to pasture remained classified as pasture for at least seven consecutive years, reducing classification errors.
- Spatial Filter: A spatial filter was applied to prevent undesired changes along the edges of pixel groups using the "connectedPixelCount" function. Native to the Google Earth Engine (GEE) platform, this function identifies connected pixels (neighbors) with the same value. Isolated pixels—those not sharing connections with a predefined number of identical neighbors—were considered separately. Similar to the implementation in Collection 9, this filter improved spatial consistency along pixel boundaries.

4.6.3. Integration with cross-cutting themes

The final map generated by the Pantanal biome team was integrated with maps from some cross-cutting themes, which represent rare classes or classes that demand a specific mapping strategy. These external themes comprised urban areas (24), mining (30), agriculture (41) and forest plantation (9). They were superimposed on the Pantanal data, resulting in final maps of the biome containing 13 classes.

4.7. Agriculture

4.7.1 Methodology overview

For this collection the agriculture classes follow the general methodology for agricultural classes in MapBiomas' Collection 9, although without specific crop differentiation. For this collection the available agricultural classes are, at legend level 2, Forest Plantation and Agriculture, the last being further divided into Temporary Crop and Perennial Crop at legend level 3. Codes are available on our public Github (https://github.com/mapbiomas/brazil-agriculture).

4.7.2 Classification with Random Forest

The agriculture mosaic was divided into an annual and seasonal mosaic. The annual mosaic was composed from images from the entire calendar year, while the seasonal mosaic was constructed based on the annual vegetative peak per mosaic tile. The vegetative peak month was identified based on the maximum value of EVI2 in the crop year, which starts at October of the previous calendar year and ends at September of the target calendar year, as shown in Figure 11. The seasonal mosaic took into account the period between three months before and three months after the vegetative peak month.

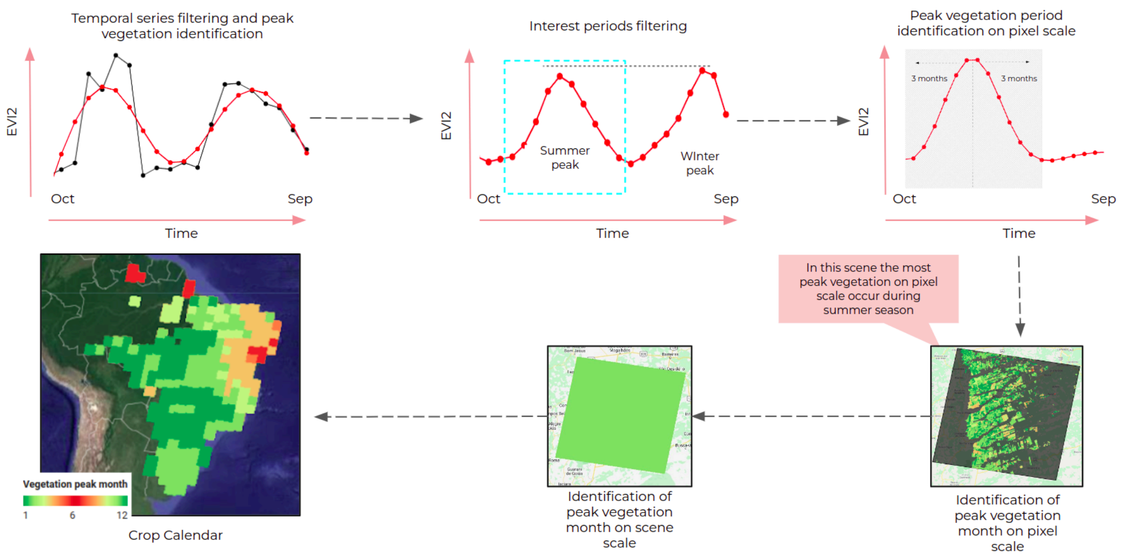

Figure 11. Scheme to obtain vegetation peak month, year by year, per mosaic tile in the agriculture classification.

Considering both seasonal and annual mosaics, the total feature space consisted of 80 variables. From Sentinel-2, the Green, Red, Red Edge1, NIR, SWIR1 and SWIR2 bands were used, and the EVI2 and NDWI indexes were calculated. For each band and index, the median, standard deviation, percentile 20 and 85 were calculated, in addition to a Quality Mosaic band.

For the classification of each class, samples were obtained from stable samples of MapBiomas' Collection 9 and other reference maps, listed in Table 8. A simple sample approach was used for temporary crops, where samples are selected at random throughout the reference mask, with no restriction to sample balance. An stratified sample approach was used for Forest Plantation and Perennial Crops, with samples being balanced by the percentage of the class in each tile.

Table 8. Reference maps used in the classification for the agriculture classes in the MapBiomas 10 m Collection 2 (beta).

Class	Training samples per tile	Sampling Approach	Source
Temporary Crop	10,000	Simple	MapBiomas Brazil Collection 9 stable samples; Agência Nacional de Águas (ANA, 2020) and Companhia Nacional de Abastecimento (Conab), Terraclass.
Perennial Crop	10,000	Stratified	MapBiomas Brazil Collection 9 stable samples; Companhia Nacional de Abastecimento (Conab); Quarta comunicação nacional do Brasil à UNFCCC .
Forest Plantation	10,000	Stratified	Global Forest Watch, Transparent World (2015)

Each class was classified separately, using a Random Forest model with 100 trees. The classification for Temporary Crop used both the seasonal and annual mosaic, while Forest plantation used the annual mosaic. The Perennial Crop class used the annual mosaic but was only partially classified using a Random Forest model, mainly in areas of coffee production.

4.7.3 Classification with Deep Learning

The Perennial Crop class was made using two Deep Learning models for different regions. The samples used in this classification were manually collected using high resolution images and visual interpretation. The model training and predictions were made in Google Colab, with results imported back to Google Earth Engine.

For the north region of Brazil, mainly in areas of oil palm, an adaptation from the deeplabv3_resnet50 architecture (TORCHVISION CONTRIBUTORS, 2023) was used. In its default configuration, the dimensions received were (256, 256, 3), which corresponded to an 256x256 pixels image with three bands. The adaptation was made to allow it to receive six bands, including Blue, Green, Red, NIR, SWIR1 from Sentinel-2 and the VH band from Sentinel-1. The model was trained from scratch for 150 epochs.

For other regions of Brazil, a model adapted from the U-Net architecture was used. The model was trained with tiles of 260x260 pixel size, with data from Sentinel-2

Red and NIR bands, as well as Sentinel-1 VH band.

4.7.4 Filters

Post processing consisted in a temporal and spatial filter. The temporal filter used was a moving windows filter with a window and a threshold. This filter goes step by step at the time series looking for a window of values centered at the current step, meaning it also considers the previous and next values. The threshold indicates the minimum occurrences of the class in the window to assign the step value as the class. This was used to fill gaps due to misclassifications and to reduce omissions. The parameters for Temporary Crop were window of 3 and threshold of 2, while Perennial Crop and Forest Plantation used a window of 5 and threshold of 3. An additional filter in the final year was also applied that acts as a fill, to prevent sudden reduction in area caused by omissions in the last years and the inability of the window to go beyond the end of the series.

A spatial filter was used to remove clusters of isolated pixels, and was applied before and after the temporal filter. This filter aims to reduce inclusion errors and noise from the classifier and filters. The minimum number of connected pixels was set to 36.

4.8. Coastal Zone

4.8.1 Methodology overview

For this MapBiomas 10 m Collection 2 (beta), the classes presented in the Coastal Zone followed the same general methodology described and applied to these classes in the MapBiomas 30 m Collection 9, but without the use of the frequency filter (final filter). The classes in this collection are listed in Table 9, along with their respective references.

Table 9 - Reference datasets to guide training samples of Coastal Zone classes in the MapBiomas 10 m Collection 2 (beta).

Class	References
Mangrove	MapBiomas Collection 9, Giri et al., 2011, ICMBio Mangrove Atlas (ICMBio, 2018), Global Mangrove Watch (Bunting et al., 2018; Thomas et al., 2018), Diniz et al., 2019, Panorama da Conservação dos Ecossistemas Costeiros e Marinhos no Brasil (MMA, 2010), and visual inspection.
Aquaculture/Salt-Culture	MapBiomas Collection 9, Atlas Dos Remanescentes Florestais da Mata Atlântica (SOS Mata Atlântica, 2020), Barbier and Cox, 2003; Guimarães et al., 2010; Prates, Gonçalves and Rosa, 2010, Queiroz et al., 2013; Tenório et al., 2015; Thomas et al., 2017, Diniz et al., 2021, and visual inspection

Apicum/Hypersaline tidal flat	MapBiomas Collection 9, Atlas Dos Remanescentes Florestais da Mata Atlântica (SOS Mata Atlântica, 2020), Prates, Gonçalves and Rosa, 2010, Panorama da Conservação dos Ecossistemas Costeiros e Marinhos no Brasil (MMA, 2010), and visual inspection.
Beaches, Dunes and Sand Spots	MapBiomas Collection 9, Atlas Dos Remanescentes Florestais da Mata Atlântica (SOS Mata Atlântica, 2020), Prates, Gonçalves and Rosa, 2010, Panorama da Conservação dos Ecossistemas Costeiros e Marinhos no Brasil (MMA, 2010), and visual inspection.
Shallow Coral Reef	Áreas Prioritárias para Conservação da Biodiversidade (MMA), Panorama da Conservação dos Ecossistemas Costeiros e Marinhos no Brasil (MMA, 2010), Atlas dos Recifes de Corais nas Unidades de Conservação Brasileiras (Prates, 2006), Allen Coral Reef Atlas, and UNEP-WCMC Global Distribution of Coral Reefs.

The general detection methodology is shown in Figure 11, showing two possible paths; machine learning or deep learning detection. The steps of image processing, acquisition of training datasets, classification and integration and accuracy are also shown. It is worth noting that machine learning methods were performed on the GEE platform, while deep learning methods were performed locally.

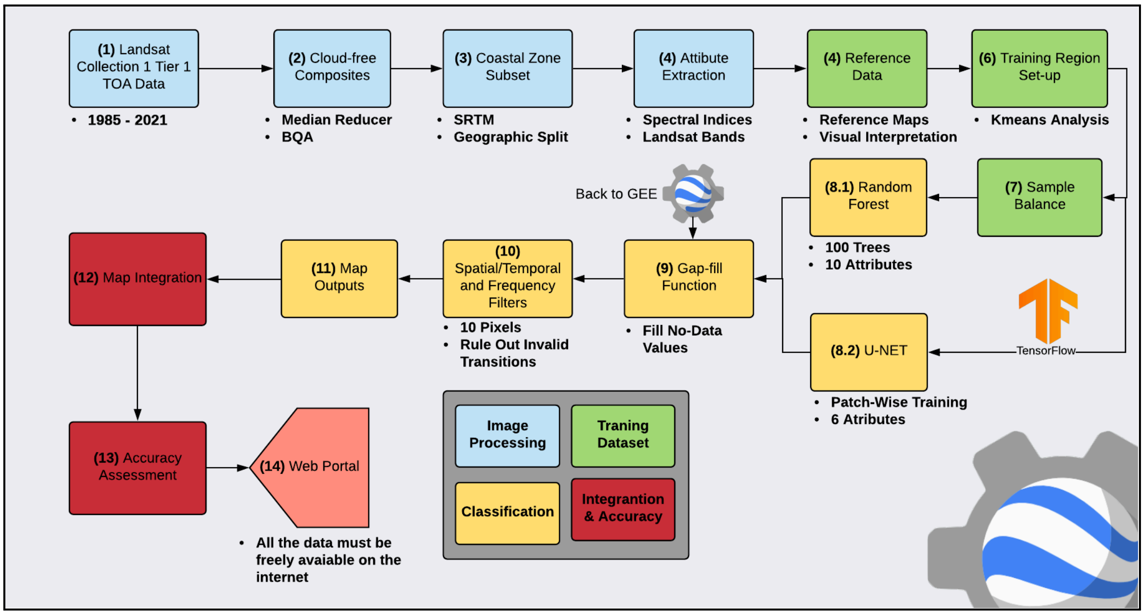

Figure 12. Workflow of Coastal Zone mapping, validation, and publication in the MapBiomas 10 m Collection 2 (beta). All data processing occurs within the Google Earth Engine - GEE platform, except for the aquaculture/saline pattern and Hypersaline Tidal Flat classification, dependent on the TensorFlow library. In green are steps related to sampling design. In yellow are steps related to classification.

4.8.2 Mosaic preparation

The annual Sentinel 2 mosaic provided by MapBiomas was used. The values were normalized by dividing by 37, with subsequent selection of the spectral bands: green, red, nir, swir1. With this, the spectral indices NDVI, MNDWI, MMRI were calculated and aggregated (Table 10). The mosaic was also normalized to include values in the range of unsigned integers of 8 bits. These were stored as assets in the Google Earth Engine platform.

Table 10. Spectral Indices used for Coastal Zone classification in the MapBiomas 10 m Collection 2 (beta).

Index	Expression	Reducer	Reference
NDVI	(NIR - RED) / (NIR + RED)	Median and Standard Deviation	Tucker, 1979
MNDWI	(GREEN - SWIR1) / (GREEN + SWIR1)	Median and Standard Deviation	Xu, 2006
MMRI	Modular Mangrove Recognition Index	Median and Standard Deviation	Diniz et al., 2019

4.8.3 Classification with Random Forest

The mapping of classes using the Random Forest algorithm (Breiman, 2001) occurred in its entirety on the Google Earth Engine platform, using mosaics stored in asset format. The selected classes were mangroves and beaches, dunes and sandbanks. The classification occurred separately, resulting in binary maps indicating the presence and absence of the respective target class. Table 11 presents the configurations of the Random Forest algorithm used.

We have selected training points based on the availability of reference maps and the previous MapBiomas 10 m Collection 1 (beta) (with Sentinel 2 images). When reference maps that match the classes and/or year to be classified, reference maps of the closest possible timeframe to the median composites were used. When no reference map was available, then the classification results of the previous year were used for subsequent training.

Table 11. Random Forest parameters used to classify each one of the years. Mangroves, beaches, dunes, sand spots, and shallow coral reefs.

Parameter	Value
Number of trees	100
Number of points	100000
Number of Variables	25 (Coastal Zone + Coastal Waters (Coral Reefs Area)
Classes	2 (binary classification)

4.8.4 Segmentation with Deep Learning

The detection of the aquaculture and *apicum* (hypersaline tidal flat) classes was performed using convolutional neural networks, specifically a model based on U-net. The need to map using Deep Learning techniques occurs due to the high dependence of the target on its spatial context: aquacultures have as a fundamental characteristic a grouping of rectangular pools; *apicuns* are spatially linked to the presence of mangroves, and are spectrally exposed soil.

The mosaics used for deep learning were stored locally, and with the selection of the attributes present in Table 12. After that, the supervised layer was developed for each class and sample collection geometries for training and validation were designed, with the samples being transferred to local storage. The years selected for the composition of the samples were 2023 for aquaculture and 2021 for *apicum*.

The training was carried out using the Keras library, with Tensor Flow as the processing backend. The other training characteristics are present in Table 12.

Table 12. U-Net parameters used to classify each year. The U-Net-derived classes are the aquaculture and Hypersaline Tidal Flat classes.

Parameter	Value
Classifier	U-Net
Epochs	100
Tile-Size	256 x 256 px
Optimizer	Nadam (Adam with Nesterov momentum)
Learning Rate	5e-6
Samples	70% (training) / 30% (validation)
Attributes	Red, Green, Nir, Swir 1, MNDWI, NDVI
Classes	2 (probabilistic segmentation)

After training, the trained model performed predictions on mosaic patches from each respective year, resulting in probabilistic maps of target existence. These images were sent to GEE for subsequent binarization and post-processing.

4.8.5 Filters

Due to the classification method's pixel-based nature and the very long temporal series, a set of post-classification filters was applied. The post-classification process includes applying a gap-fill, a temporal and a spatial filter.

4.8.5.1 Gap-Fill filter

The chain starts by filling in possible no-data values. In a long-time series of severely cloud-affected regions, such as tropical coastal zones, it is expected that no-data values may populate some of the resultant median composite pixels. In this filter, no-data values (“gaps”) are theoretically not allowed and are replaced by the temporally nearest valid classification. In this procedure, if no “future” valid position is available, the no-data value is replaced by its previous valid class. Up to three prior years can be used to fill in persistent no-data positions. Therefore, gaps should only exist if a given pixel has been permanently classified as no-data throughout the entire temporal domain. A mask of years was built to keep track of pixel temporal origins.

4.8.5.2 Temporal filter

After gap-filling, a temporal filter was executed. The temporal filter uses sequential classifications in a 3-year unidirectional moving window to identify temporally non-permitted transitions. Based on a single generic rule (GR), the temporal filter inspects the central position of three consecutive years (“ternary”). If the extremities of the ternary are identical, but the center position is not, then the central pixel is reclassified to match its temporal neighbor class, as shown in Table 13.

Table 13. The temporal filter inspects the central position for three consecutive years, and in cases of identical extremities, the center position is reclassified to match its neighbor. T1, T2, and T3 stand for positions one (1), two (2), and three (3), respectively. GR means “generic rule”, while Tg and N-Tg represent target class and non-target class pixels.

Rule	Input (Year)			Output		
	T1	**T2**	**T3**	**T1**	**T2**	**T3**
GR	Tg	N-Tg	Tg	Tg	Tg	Tg
GR	N-Tg	Tg	N-Tg	N-Tg	N-Tg	N-Tg

4.8.5.3 Spatial filter

Posteriorly, a spatial filter was applied. To avoid unwanted modifications to the edges of the pixel groups (blobs), a spatial filter was built based on the "connectedPixelCount" function. Native to the GEE platform, this function locates connected components (neighbors) that share the same pixel value. Thus, only pixels that do not share connections to a predefined number of identical neighbors are considered isolated. This filter needs at least ten connected pixels to reach the minimum connection value. Consequently, the minimum mapping unit is directly affected by the spatial filter applied, and it was defined as 10 pixels (~1 ha).

4.9. Mining

4.9.1 Methodology overview

Mining mapping in the new MapBiomas 10 m Collection carried the same method as in the MapBiomas 30 m Collection 9 in most of Brazil's territory. It included updates on the quantity and quality of the training samples, and a more significant number of activation grids were used in the processing steps of the mining recognition algorithm, which is based on a U-Net classifier.

The reference dataset used in our classification consisted of multiple data sources, as specified in Table 14.

Table 14. Mining class references used in MapBiomas 10 m Collection 2 (beta).

Mining class reference dataset	Mining Deter: http://terrabrasilis.dpi.inpe.br/ MapBiomas Alert: http://alerta.mapgiomas.org RAISG: http://www.amazoniasocioambiental.org ISA: https://www.socioambiental.org/ CPRM-GeoSGB: https://geosgb.cprm.gov.br/ Ahkbrasilien: https://www.ahkbrasilien.com.br/ AMW: https://amazonminingwatch.org/ and Additional visual interpretation.

The multiple reference data were visually analyzed and converted to bounding boxes, which were overlaid on grids used to process the deep-learning mining recognition algorithm in a parallel fashion.

The entire process is structured in 4 steps (Figure 13). First (1), GEE generates the cloud-free composites and creates the initial training dataset. Second (2), the mosaics and training data are downloaded and stored locally. Three (3) initiate patch-wise training and classification. In the fourth step (4), the classified product is spatially and temporally filtered. The filtered product is visually and statistically inspected. Multiple iterations may be used until a satisfactory spatial and temporal quality is achieved

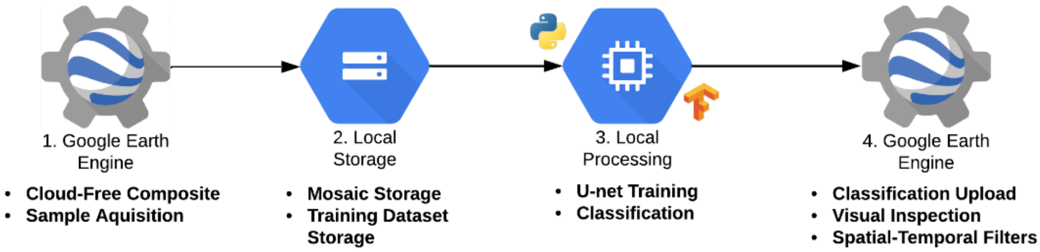

Figure 13. Mining Detection Earth Engine-TensorFlow pipeline.

4.9.2 Classification with Deep Learning

For the supervised classification of the Landsat mosaics, we selected training samples geometries) from the previously generated bounding boxes (grids). Like any supervised algorithm, our U-net-based approach depends on human-labeled training data, categorized as mining (Mi) and non-mining (N-Mi) and guided by the reference dataset, the samples are visually delineated. The labels generated for Sentinel data differ from the ones made for Landsat, accounting for change in resolution and which patterns / mining structures are visible to the specialists in each image.

Images were processed on 512x512 chips, making the U-net network 2x larger than that used in the regular Collection classification.

It is essential to highlight that no differentiation was made between artisanal or industrial mining samples during the classification process: they include both artisanal and industrial patterns. The dissociation between such patterns, garimpo or industrial, and the exploited main substance is a post-classification step that is not present in the Sentinel Collections.

Once the sample collection is finished, the U-net classification results in the pre-filtered classification product. The classified data is injected back into GEE, where spatial-temporal filters and visual inspection occur.

4.9.3 Filters

In GEE, the resulting classification undergoes a chain of filters with the goal of reducing salt-and-pepper effect and adding spatiotemporal consistency. This process includes the application of the following filters: gap-fill, temporal and spatial.

4.9.3.1 Gap-Fill filter

The post-processing steps start by filling in possible no-data values. In a series of severely cloud-affected regions, such as forested areas of tropical countries, pixels with no-data values are expected to be present in median composite mosaics. The gap-fill filter replaces the no-data values (i.e., image "gaps") with a classified pixel from the nearest date available. In this procedure, if no "future" valid class is available, the no-data value is replaced by the nearest previous valid class. Up to three prior years can fill in persistent no-data pixels. Therefore, gaps should only exist if a given pixel has been permanently classified as no-data throughout the entire temporal series.

4.9.3.2 Temporal filter

Next, we applied a temporal filter that uses sequential classifications in a 3-year unidirectional moving window to identify temporally non-permitted transitions. Based on a single generic rule (GR), the temporal filter inspects the central position of three consecutive years ("ternary"). It changes its value if it differs from the first and last years in the ternary, which must have identical classes. The central year of the ternary is then reclassified to match its temporal neighbor class, as shown in **Table 15**.

Table 15. The temporal filter inspects the central position for three consecutive years, and in cases of identical extremities, the center position is reclassified to match its neighbor. T1, T2, and T3 stand for positions one (1), two (2), and three (3), respectively. GR means "generic rule," while Mi and N-Mi represent mining and non-mining pixels.

Rule	**Input (Year)**			**Output**		
	T1	**T2**	**T3**	**T1**	**T2**	**T3**
GR	Mi	N-Mi	Mi	Mi	Mi	Mi
GR	N-Mi	Mi	N-Mi	N-Mi	N-Mi	N-Mi

4.9.3.3 Spatial filter

Then, a spatial filter was applied to avoid unwanted modifications on the edges of grouping pixels (clusters) by using the "connectedPixelCount" function. Native to the

GEE platform, this function locates connected components (neighbors) that share the same pixel value. Thus, only pixels that do not share connections to a pre-defined number of identical neighbors are considered isolated, as shown in Figure 10. This filter needs at least ten connected pixels to reach the minimum connection value. Consequently, the minimum mapping unit is directly affected by the spatial filter applied, which was defined as 10 pixels (~1 ha).

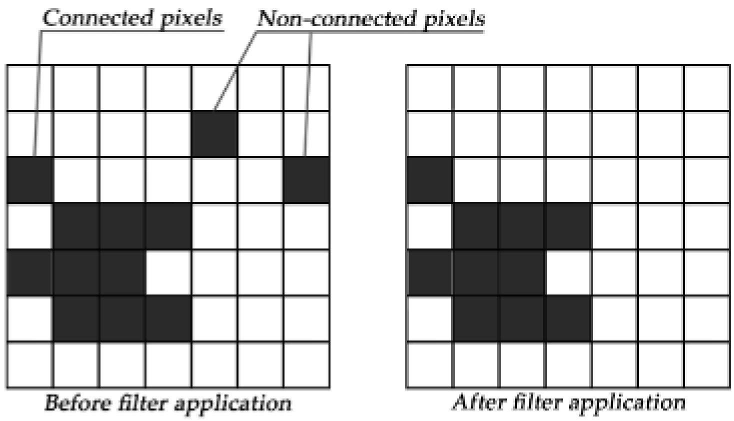

Figure 14. The spatial filter removes pixels that do not share neighbors of identical value. The minimum connection value was 10 pixels.

4.9.3.4 Frequency filter

The last post-processing filter step is the frequency filter. This filter considers the frequency of a given class throughout the entire time series. Thus, all class occurrences with less than 25% temporal persistence (2 years or out of 8) are filtered out and incorporated into the non-class binary. This mechanism contributes to reducing the temporal oscillation in the classification, decreasing the number of false positives, and preserving consolidated classes.

4.10. Pasture

4.10.1 Methodology overview

Using data from ESA's Sentinel 2 constellation at 10 m resolution for the period of 2016 to 2023, the MapBiomas 10 m Collection 2 (beta) pasture map follows the general approach (Parente et al., 2019) for pasture mapping in the MapBiomas 9 Collection, however with some minimal changes, focusing mainly on ensure a resolution leap with a better precision in pastureland detection (Figure 15).

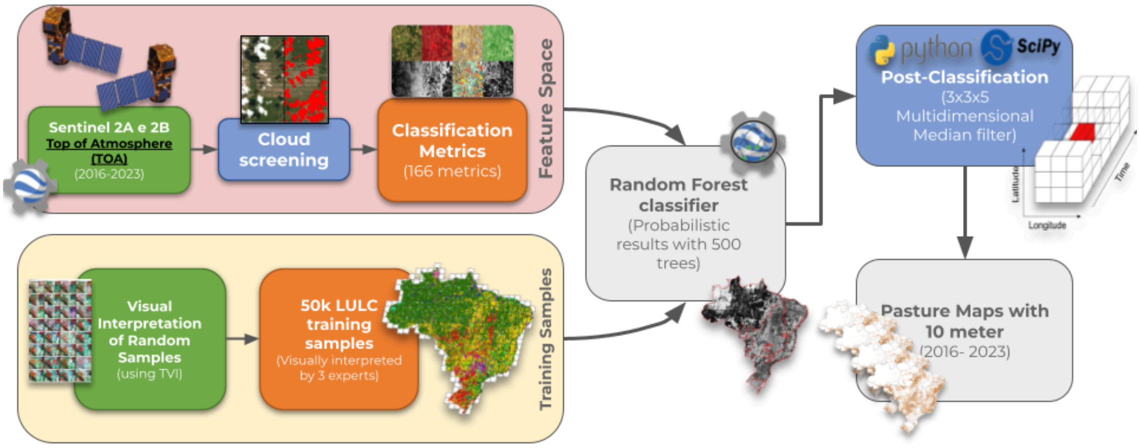

Figure 15 - Pasture mapping workflow regarding the MapBiomas 10 m Collection 2 (beta).

With a broad option of spectral bands, ESA'S Harmonized Sentinel-2 MSI *Top of Atmosphere* data, stored in Earth Engine Platform, were used to build a brand new feature space, considering the same features used by Parente et al. (2019) with the addition of newer ones.

ESA data makes available the Red Edge spectral region, which is crucial for a better understanding of plant health and behavior. Also, these new spectral bands increase the range of possible spectral indexes that could be estimated with satellite data and some of them, like Soil-Adjusted Total Vegetation Index - SATVI (Marsett et al., 2006), are specific for pasture/grassland areas. In total, 166 metrics were used as feature space in this collection (Table 16), adding 88 new metrics and keeping the 78 used in MapBiomas Col. 9.

Table 16. Table exemplifying the feature space used in the pasture mapping with Sentinel 2 constellation, composed of 162 spectral-temporal metrics (Sentinel 2 based) and 4 time independent metrics (geographic coordinates and SRTM data).

Bands/Indexes	Reducers
Blue (B2)	Minimum, Median, Maximum, Amplitude, Standard Deviation, Percentile (10%,25%,75%,90%)
Green (B3)	
Red (B4)	
Red Edge 1 (B5)	
Red Edge 2 (B6)	
Red Edge 3 (B7)	
Near Infrared	
Red Edge 4 (B8A)	
Short Wave Infrared 1 (B11)	
Short Wave Infrared 2 (B12)	
Normalized Difference Vegetation Index - **NDVI** (Tucker et al., 1979)	
Normalized Difference Water Index - **NDWI** (Gao, 1996)	
Cellulose Absorption Index - **CAI** (Nagler et al., 2000)	
Carotenoid Reflectance Index - **CRI1** (Gitelson et al., 2007)	
Anthocyanin Reflectance Index - **ARI1** (Gitelson et al., 2007)	
Simple Ratio Red/Green Red-Green Ratio - **RGR** (Gamon and Surfus, 1999)	
Plant Senescence Reflectance Index - **PSRI** (Merzlyak et al., 1999)	
Soil-Adjusted Total Vegetation Index - **SATVI** (Marsett et al., 2006)	
Geographic Coordinate - **Longitude**	-
Geographic Coordinate - **Latitude**	-
SRTM - **Elevation** (Farr et al., 2007)	-
SRTM - **Slope** (Farr et al., 2007)	-

The Random Forest classifier training considered the same 50,000 random samples visually interpreted by experts using TVI plus 4,664 intervention points, used to ensure that the classifier will not fail to detect pasture areas not covered by the main sampling approach. Furthermore, the classifier hyperparameters were kept the same except for the number of variables per split, which follows the square root value of the feature space size (approximately 13).

4.10.2 Post-processing

The spatiotemporal filter used in the probabilistic pasture maps, before collection integration, was kept the same as in the other MapBiomas collections due to its simplicity and quality of results, which are even superior to some complex 3D Savitzky-Golay filters. It consists of a multidimensional median filter with a 3 x 3 window in space (X and Y) for 5 years (Z) packed in the Scipy library from the Python language. After the filtering process, the maps are transformed from probabilistic maps to discrete maps by establishing the cut-off point for pasture areas for probabilities greater than 51%.

4.11. Urban Area

4.11.1. Classification algorithm

The methodology for this collection follows the methodology for Urban Area data of Collection 9 and the general methodology of Collection 10 m (Figure 16). To reduce computational cost for urban area mapping, the classification was performed only in “search areas”, defined by polygons where urban areas were likely to be found. A uniform hexagonal polygon grid was created over Brazilian territory and intersected with urban census tracts (IBGE, 2021)[1], resulting in a search area of 226 million ha, covering 27% of the Brazilian territory.

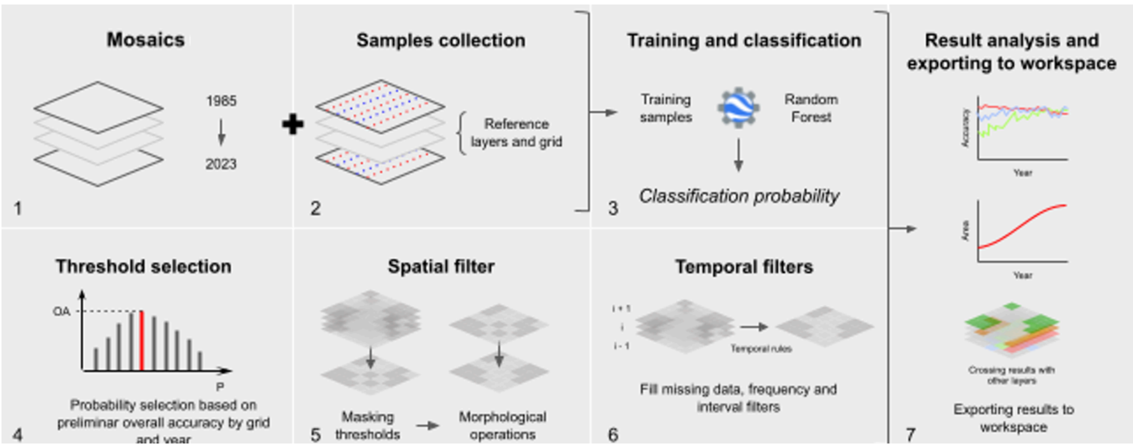

Figure 16. Basic scheme of the production of Urban Area maps in MapBiomas

[1] Census tracts are classified according to their situation. To build the search areas, we considered tracts in the following situations: (1) urban area with high density of buildings, (2) urban area with low density of buildings and (3) urban nucleus.

For this theme, Brazil’s territory was divided into 558 tiles that correspond to charts with a scale of 1:250.000, derived from the International Map of the World (IMW). Tiles with no search area were discarded, resulting in 522 valid tiles. Then, a specific classifier was trained to each of these tiles of each year of the 39 years of the Collection 9. Random Forest parameters for Urban Area mapping were set to 500 trees and 20 minimum leaf populations. Samples were taken from Collection 9 and considering all tiles, the average balance of samples was: between 2016-2017, 1 urban sample to 1.9 non-urban samples; and in the final years (2018-2023), 1 urban sample to 1.7 non-urban samples.

4.11.2. Spatial Filter

The classifier may incorrectly assign high urban area (UA) probability values to non-urban features like mining, sands, and rural structures, while urban areas with trees or parks might receive low probability values. Universal probability thresholds for defining urban areas would result in errors due to the diverse characteristics of different cities. To solve this problem, an algorithm for choosing the best probability threshold is applied, which defines the best cut-off value for each grid.

To refine urban area (UA) classification, data on urbanized and slum areas from the 2022 census were combined with the Index of Roads and Infrastructure (IRS) developed by Justiniano et al. (2022). These datasets provide a comprehensive view, enabling the delineation of the maximum extent of urban pixels based on recent historical patterns, improving accuracy in identifying and classifying urban spaces.

Isolated pixels or small clusters, often errors in classification, are addressed by spatial filters using morphological operations. For urban areas, small zero-value clusters might represent squares, parks, or water features, while in non-urban areas, isolated one-value pixels could represent agricultural or rural structures. The spatial filters apply circular kernels, performing closing operations to remove small holes (fewer than 10 pixels) and opening operations to eliminate noise (fewer than 10 pixels).

4.11.3. Temporal Filter

Temporal filters (TF) were applied as rules to check classification consistency over time, observing the conceptual aspects delimited to the mapped category. For this purpose, the sequence of filters indicated and described in Table 17 was developed. General rules (GR), Persistent Noise (PN) for middle years, and specific rules for the first years (FYR) and last years (LYR).

Table 17. Descriptions of Rules.

Rule	Years (i)	Kernel						Conditionals
		i-2	i-1	i	i+1	i+2	i+3	
PN	1986 to 2021		x	x	x	x		If the pixel under analysis is classified as 'UA' within three or more years of the interval, then the 'UA' is validated
GR	1986 to 2022		x	x	x			If the pixel under analysis is classified as 'UA' within two or more years of the interval, then the 'UA' is validated
FYR	1985			x	x			If the pixel under analysis is classified as 'UA' within two years of the kernel, then the 'UA' is validated
LYR	2023		x	x				If the pixel under analysis is classified as 'UA' within two or more years, then the 'UA' is validated

5. Integration and Post-classification

5.1 Integration

The maps of each biome and of cross-cutting themes were integrated on a pixel-by-pixel basis through the hierarchical overlap of each mapped class, following prevalence rules defined by experts. Certain prevalence rules may show exceptions for one or more classes. Some classes present specific prevalence rules or exceptions in certain biomes or regions. The prevalence rules and its exceptions are listed in Annex II.

5.2. Filters on Integrated Maps

In the integrated maps four spatial and temporal filters were applied. Two filters were applied to adjust the temporal consistency of the forest plantation and agriculture classes specifically. For the other classes, another temporal filter was used to remove isolated pixels in the time series (e.g. an isolated class pixel between two different classes pixels). A spatial filter similar to the one described was applied on the integrated maps to remove isolated classes with less than half hectare as well as noise resulting from integration. MapBiomas Alerta deforestation data was used as an accumulated mask, and when it overlapped with any natural class, it was converted to class 21 (Mosaic of Uses).

5.3. Statistics

Zonal statistics of the mapped classes were calculated for different spatial units, such as the biomes, states, municipalities, watersheds, protected areas (including indigenous lands and conservation units) and can be found in https://brasil.mapbiomas.org/en/estatisticas/.

6. Concluding Remarks and Perspectives

The algorithms developed for pre-processing and classifying Sentinel-2 imagery hold promise for expanding the possible applications of MapBiomas LCLU maps. Thanks to Google Earth Engine and open source technology, it is possible to access and process large-scale satellite imagery datasets such as the one generated by the MapBiomas project. The replication of this type of project is viable for other areas of the planet. The MapBiomas initiative has expanded to all South American countries and Indonesia. In addition, the MapBiomas team will keep improving the following collections in subsequent years. The open-access MapBiomas LCLU dataset allowed several scientific publications in Brazil and abroad. Policymakers and stakeholders also use the dataset for public policies and decision-makers in the country.

7. References

ABDIBDI, A. M. Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data. *GIScience & Remote Sensing*, v. 57, n 1, p 1-20. 2020.

ALENCAR, A., ARRUDA, V.L.S., SILVA, W., CONCIANI, D., COSTA, D., CRUSCO, N., DUVERGER, S., FERREIRA, N., FRANCA-ROCHA, W., HASENACK, H., MARTENEXEN, L.F., PIONTEKOWSKI, V.J, RIBEIRO, N., ROSA, E.R., ROSA, M., DOS SANTOS, S.M.B, SHIMBO, J.Z., VÉLEZ-MARTIN, E. Long-Term Landsat-Based Monthly Burned Area Dataset for the Brazilian Biomes Using Deep Learning. *Remote Sensing*, v. 14, 2510. 2022.

Allen Coral Atlas. Imagery, maps and monitoring of the world's tropical coral reefs, 2024. doi.org/10.5281/zenodo.3833242

AMATULLI, G., MCINERNEY, D., SETHI, T., STROBL, P., DOMISCH, S. Geomorpho90m - Global High-Resolution Geomorphometry Layers. Distributed by OpenTopography. https://doi.org/10.5069/G91R6NPX. 2020.

ASSIS, L. F. F. G. et al. TerraBrasilis: A Spatial Data Analytics Infrastructure for Large-Scale Thematic Mapping. ***ISPRS International Journal of Geo-Information,*** v. 8, n 513, 2019. DOI: 10.3390/ijgi8110513

ASSINE, M. L. et al. The Quaternary alluvial systems tract of the Pantanal Basin, Brazil. Brazilian Journal Of Geology, [S. L.], v. 45, n. 3, p. 475-489, 2015. DOI: 10.1590/2317-4889201520150014.

BAPTISTA, G. Aplicação do Índice de Vegetação por Profundidade de Feição Espectral (SFDVI - Spectral Feature Depth Vegetation Index) em dados RapidEye. *Anais XVII Simpósio Brasileiro de Sensoriamento Remoto - SBSR*, João Pessoa-PB, Brasil, 25 a 29 de abril de 2015, INPE.

BARBIER, E. B.; COX, M. Does Economic Development Lead to Mangrove Loss? A Cross-Country Analysis. *Contemporary Economic Policy*, v. 21, n. 4, p. 418–432, 2003.

BERGSTRA, J.; BENGIO, Y. Random search for hyper-parameter optimization. *The journal of machine learning research,* v. 13, n1, 281-305, 2012.

BRAZILIAN GEOLOGICAL SERVICE SGB/CPRM. *Rocky Outcrop features.* Available at:https://geoportal.sgb.gov.br/server/rest/services/geologia/afloramentos/MapServer/0

BREIMAN, L. Random forests. *Machine learning,* v. 45, n. 1, p. 5-32, 2001.

BROWN, C. F. *et al.* Dynamic World, Near real-time global 10 m land use land cover mapping. *Scientific Data,* v. 9, 251, 2022. https://doi.org/10.1038/s41597-022-01307-4.

BUNTING, P. et al. The Global Mangrove Watch—A New 2010 Global Baseline of Mangrove Extent Remote Sensing , 2018.

CECCATO, P.; FLASSE, S.; GREGOIRE, J. M. Designing a spectral index to estimate vegetation water content from remote sensing data: Part 2. Validation and applications. *Remote sensing of Environment*, v. 82, n 2-3, p. 198-207, 2002.

CHEN, T.; GUESTRIN, C. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining.* p. 785-794. 2016.

CI – Conservação Internacional, ECOA - Ecologia e Ação, Fundación AVINA, Instituto SOS Pantanal, WWF-Brasil. *Monitoramento das alterações da cobertura vegetal e uso do solo na Bacia do Alto Paraguai – Porção Brasileira – Período de Análise: 2002 a 2008.* Brasília, 2009. 58 p.; Il.; 23 cm. ISBN 978-85-86440-25-0.

Crist, E. P. A. TM tasseled cap equivalent transformation for reflectance factor data. *Remote sensing of Environment,* v. 17, n. 3, p. 301-306, 1985.

DINIZ, C.; CORTINHAS, L.; NERINO, G.; RODRIGUES, J.; SADECK, L.; ADAMI, M.; SOUZA-FILHO, W.P. Brazilian Mangrove Status: Three Decades of Satellite Data Analysis. *Remote Sensing* 11, 2019. http://dx.doi.org/10.3390/rs11070808.

DINIZ, C.; CORTINHAS, L.; PINHEIRO, M.L.; SADECK, L.; FERNANDES FILHO, A.; BAUMANN, L.R.F.; ADAMI, M.; SOUZA-FILHO, P.W.M. A Large-Scale Deep-Learning Approach for Multi-Temporal Aqua and Salt-Culture Mapping Remote Sensing, *Remote Sensing,* v. 13, n. 8, 1415, 2021. http://dx.doi.org/10.3390/rs13081415

DONCHYTS, G.; WINSEMIUS, H.; SCHELLEKENS, J.; ERICKSON, T.; GAO, H.; SAVENIJE, H.; VAN DE GIESEN, N. "Global 30m Height Above the Nearest Drainage (HAND)", *Geophysical Research Abstracts*, v.. 18, EGU2016-17445-3, EGU General Assembly, 2016.

ESA. Sentinel-2 User Handbook. European Space Agency, 2015. 64 p. Available at https://sentinel.esa.int/documents/247904/685211/Sentinel-2_User_Handbook.

FAO. *Manual for integrated field data collection.* Rome: FAO, 2012.

FARR, T. G. et al. The Shuttle Radar Topography Mission. *Reviews of Geophysics*, v. 45, n2, 2007. https://doi.org/10.1029/2005rg000183

FEYISA, G. L.; MEILBY, H.; FENSHOLT, R.; PROUD, S. R. Automated Water Extraction Index: A new technique for surface water mapping using Landsat imagery. *Remote sensing of environment*, v. 140, p. 23-35, 2014.

FISHER, A.; FLOOD, N.; DANAHER, T. Comparing Landsat water index methods for automated water classification in eastern Australia. *Remote Sensing of Environment,* v. 175, p. 167–182. 2016.

FRIEDL, M. A. et al. MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sensing of Environment,* v. 114, p. 168–182. 2010.

FRAMPTON, W. J.; DASH, J.; WATMOUGH, G.; MILTON, E. J. Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. ISPRS *Journal of Photogrammetry and Remote Sensing*, v. 82, p. 83-92, 2013. https://doi.org/10.1016/j.isprsjprs.2013.04.007

FRIEDMAN, J. H. 2001. Greedy Function Approximation: A Gradient Boosting Machine. *Annals of Statistics*, v. 2, n 5, p. 1189–1232.

GAMON, J. A.; SURFUS, J. S. Assessing leaf pigment content and activity with a reflectometer. *The New Phytologist,* v. 143, n1, p. 105–117, 1999. https://doi.org/10.1046/j.1469-8137.1999.00424.x

GANERØD, A. J.; BAKKESTUEN, V.; CALOVI, M.; FREDIN, O.; RØD, J. K. Where are the outcrops? Automatic delineation of bedrock from sediments using Deep-Learning techniques. *Applied Computing and Geosciences*, v. 18, 100119, 2023

GAO, B.-C. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment,* v. 58, n3, p. 257–266, 1996. https://doi.org/10.1016/s0034-4257(96)00067-3

GIRI, C. et al. Status and distribution of mangrove forests of the world using earth observation satellite data. *Global Ecology and Biogeography*, v. 20, n. 1, p. 154-159, 2011. Available at: https://doi.org/10.1111/j.1466-8238.2010.00584.x.

GITELSON, A.; MERZLYAK, M. N. Quantitative estimation of chlorophyll-a using reflectance spectra: Experiments with autumn chestnut and maple leaves. *Journal of Photochemistry and Photobiology B: Biology*, v. 22, n. 3, p. 247-252, 1994. https://doi.org/10.1016/1011-1344(93)06963-4

GITELSON, A. A.; KAUFMAN, Y. J.; STARK, R.; RUNDQUIST, D. Novel algorithms for remote estimation of vegetation fraction. *Remote Sensing of Environment*, v. 80, n. 1, p. 76-87, 2002. https://doi.org/10.1016/S0034-4257(01)00289-9

GITELSON, A. A.; GRITZ, Y.; & MERZLYAK, M. N. Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *Journal of Plant Physiology,* v. 160, n. 3, p. 271-282, 2003. https://doi.org/10.1078/0176-1617-00887

GITELSON, A. A.; VIÑA, A.; CIGANDA, V.; RUNDQUIST, D. C.; ARKEBAUER, T. J. Remote estimation of canopy chlorophyll content in crops. *Geophysical research letters*, v. *32, n.* 8, 2005.

GITELSON, A.; MERZLYAK, M. N.; CHIVKUNOVA, O.B. Optical Properties and Nondestructive Estimation of Anthocyanin Content in Plant Leaves. *Photochemistry and Photobiology*, v. 74, n. 1, p. 38-45, 2007.

GU, Y.; BROWN, J. F.; VERDIN, J. P.; WARDLOW, B. A five-year analysis of MODIS NDVI and NDWI for grassland drought assessment over the central Great Plains of the United States. *Geophysical research letters*, v. 34, n. 6, 2007.

GUIMARÃES, A. S. et al. Impact of aquaculture on mangrove areas in the northern Pernambuco Coast (Brazil) using remote sensing and geographic information system. *Aquaculture Research*, v. 41, n. 6, p. 828–838, 2010.

HABOUDANE, D.; MILLER, J. R.; TREMBLAY, N.; ZARCO-TEJADA, P. J.; DEXTRAZE, L. Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. Remote Sensing of Environment, v. 81, n. 2–3, p. 416-426, 2002. https://doi.org/10.1016/S0034-4257(02)00018-4

HARALICK, R. M.; SHANMUGAM, K.; DINSTEIN, I. H. Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics,* v. 6, p. 610-621, 1973.

HASENACK, H.; WEBER, E.; BOLDRINI, I. I. TREVISAN, R. 2010. Mapa de sistemas ecológicos das Savanas Uruguaias em escala 1:500.000 ou superior. Porto Alegre: Centro de Ecologia. Relatório técnico Projeto UFRGS/TNC. 18 p.

HASENACK, H.; CORDEIRO, J. L. P; WEBER, E. J. (Org.). Uso e cobertura vegetal do Estado do Rio Grande do Sul – situação em 2002. Porto Alegre: UFRGS IB Centro de Ecologia, 2015. Available at http://www.ecologia.ufrgs.br/labgeo

HOFMANN, G. S.; WEBER, E. J.; HASENACK, H. (Org.). *Uso e cobertura vegetal do Estado do Rio Grande do Sul – situação em 2015.* Porto Alegre: UFRGS IB Centro de Ecologia, 2018. Available at http://www.ecologia.ufrgs.br/labgeo

HUETE, A.; JUSTICE, C.; LIU, H. Development of vegetation and soil indices for MODIS-EOS. *Remote Sensing of environment,* v. 49, n 3, 224-234, 1994.

IBGE. *Mapa de biomas do Brasil (escala 1:5.000.000).* Rio de Janeiro: IBGE. Mapa e nota técnica. 2004a.

IBGE. *Mapa de biomas do Brasil (escala 1:5.000.000).* Rio de Janeiro: IBGE, 2004b.

IBGE. *Uso da terra no Estado do Rio Grande do Sul: relatório técnico.* Rio de Janeiro: IBGE. 2010.

IBGE. *Manual técnico de uso da terra.* Rio de Janeiro: IBGE. 1999.

IBGE. *Manual técnico da vegetação brasileira.* 2ed. Rio de Janeiro: IBGE. 2012.

IBGE. *Biomas e Sistema Costeiro-Marinho do Brasil 1:250.000*. Rio de Janeiro: IBGE, 2019. Available at https://www.ibge.gov.br/geociencias/informacoes-ambientais/vegetacao/15842-biomas.html .

IBGE. *Malha de Setores Censitários.* Rio de Janeiro: IBGE, 2021. Available at: https://www.ibge.gov.br/geociencias/downloads-geociencias.html?caminho=organizacao_do_territorio/malhas_territoriais/malhas_de_setores_censitarios__divisoes_intramunicipais/2021/Malha_de_setores_(shp)_Brasil

IBGE. *Banco de Dados e Informações Ambientais (BDiA) - Um instrumento para organização e preservação.* 2023 Available at https://bdiaweb.ibge.gov.br/#/home .

IBGE. *Vegetação 1:250000. Versão 2021*. Downloads/ Geociências. Available at: https://geoftp.ibge.gov.br/informacoes_ambientais/vegetacao/vetores/escala_250_mil/versao_2021/.

IBGE. *Pedologia 1:250000. Versão 2021.* Downloads/ Geociências. Available at: https://geoftp.ibge.gov.br/informacoes_ambientais/pedologia/vetores/escala_250_mil/

ICMBio. Atlas dos Manguezais do Brasil / Instituto Chico Mendes de Conservação da Biodiversidade. – Brasília: Instituto Chico Mendes de Conservação da Biodiversidade, 2018.

IF. Inventário Florestal do Estado de São Paulo. https://semil.sp.gov.br/2020/08/novo-inventario-florestal-do-esp-aponta-crescimento-de-214-mil-hectares-de-vegetacao-nativa-no-territorio-paulista/ . 2020.

JORDAN, C. F. Derivation of leaf-area index from quality of light on the forest floor. *Ecology***,** v. 50, n.4, p.663-666, 1969.

JUSTINIANO, E. F.; DOS SANTOS JUNIOR, E. R.; DE MELO, B. M.; SIQUEIRA, J. V. N;, MORATO, R. G.; FANTIN, M.; PEDRASSOLI, J. C.; MARTINES, M. R.; KAWAKUBO, F. S. Proposal for an index of roads and structures for the mapping of non-vegetated urban surfaces using OSM and Sentinel-2 data. *International Journal of Applied Earth Observation and Geoinformation,* v. 109, 2022.

KAWAMURA, M.; JAYAMANNA, S.; TSUJIKO, Y. Quantitative evaluation of urbanization in developing countries using satellite data. *Doboku Gakkai Ronbunshu,* v. 1997, n. 580, p. 45-54, 1997. https://doi.org/10.2208/jscej.1997.580_45

LANG, N. et al. A high-resolution canopy height model of the Earth. *Nature Ecology & Evolution*, v. 7, p. 1778–1789, 2023. https://doi.org/10.1038/s41559-023-02206-6

LIANG, S. Narrowband to broadband conversions of land surface albedo I: Algorithms. *Remote sensing of environment*, v. 76, n. 2, 213-238, 2001.

LAIPELT, L.; COMINI DE ANDRADE, B.; COLLISCHONN, W.; DE AMORIM TEIXEIRA, A.; PAIVA, R.C.D.D.; RUHOFF, A. ANADEM: A Digital Terrain Model for South America. *Remote Sensing.* v. 16, 2321, 2024. https://doi.org/10.3390/rs16132321

LIMA, M. I. C. *Projeto Radam: uma saga na Amazônia*. Belém: Paka-tatu, 2009.

LOI, D. T.; CHOU, T. Y.; FANG, Y. M. Integration of GIS and remote sensing for evaluating forest canopy density index in Thai Nguyen Province, Vietnam. *International journal of environmental science and development,* v. 8, n. 8, p. 539-542, 2017.

LOURENÇO, P. Biomass estimation using satellite-based data. In: Forest Biomass-From Trees to Energy. Gonçalves, A. C.; Souza, A.; Malico, I. (Eds). London: IntechOpen, 2021. Available at https://www.intechopen.com/chapters/73176 .

MARSETT, R. C.; QI, J.; HEILMAN, P;, BIEDENBENDER, S. H.; CAROLYN WATSON, M.; AMER, S.; WELTZ, M.; GOODRICH, D.; MARSETT, R. Remote sensing for grassland management in the arid southwest. *Rangeland Ecology & Management,* v. 59, n. 5, p. 530–540, 2006. https://doi.org/10.2111/05-201r.1

MATHIEU, R.; POUGET, M.; CERVELLE, B.; ESCADAFAL, R. Relationships between satellite-based radiometric indices simulated using laboratory reflectance data and typic soil color of an arid environment. *Remote sensing of environment*, v. 66, n. 1, p. 17-28, 1998.

MCFEETERS, S. K. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International journal of remote sensing,* v. 17, n. 7, p. 1425-1432, 1996.

MMA. Gerência de Biodiversidade Aquática e Recursos Pesqueiros. Panorama da conservação dos ecossistemas costeiros e marinhos no Brasil. Brasília: MMA/SBF/GBA, 2010. 148 p.

MORELLATO L. P. C.; HADDAD C. F. B. Introduction: The Brazilian Atlantic Forest. *Biotropica* v. 32, p. 786–792, 2000. doi:10.1111/j.1744-7429.2000.tb00618.x.

MERZLYAK, M. N.; GITELSON, A. A.; CHIVKUNOVA, O. B.; RAKITIN, V. Y. U. Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening. *Physiologia Plantarum,* v. 106, n. 1, p. 135–141, 1999. https://doi.org/10.1034/j.1399-3054.1999.106119.x

NASA JPL. NASADEM Merged DEM Global 1 arc second V001. NASA EOSDIS Land Processes DAAC, 2020. Available at doi:10.5067/MEaSUREs/NASADEM/NASADEM_HGT.001.

NAGLER, P. L.; DAUGHTRY, C. S. T.; GOWARD, S. N. Plant litter and soil reflectance. *Remote Sensing of Environment,* v. 71, n. 2, p. 207–215. 2000. https://doi.org/10.1016/s0034-4257(99)00082-6.

NATEKIN, A.; KNOLL, A. Gradient boosting machines, a tutorial. *Frontiers in neurorobotics*, v. *7*, n. 21, 2013.

DE OLIVEIRA G. et al. Conserving the Brazilian semiarid (Caatinga) biome under climate change. *Biodiversity and Conservation*, v. 21, p. 2913–2926, 2012. doi:10.1007/s10531-012-0346-7.

NGUYEN, C. T.; CHIDTHAISONG, A.; KIEU DIEM, P.; HUO, L. Z. A modified bare soil index to identify bare land features during agricultural fallow-period in southeast Asia using Landsat 8. *Land,* v. 10, n. 3, 231, 2021.

OLOFSSON P. et al. Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment,* v. 148, p. 42-57, 2014.

OU, J., TIAN, Y., ZHANG, Q., XIE, X., ZHANG, Y., TAO, J., & LIN, J. . Coupling UAV Hyperspectral and LiDAR Data for Mangrove Classification Using XGBoost in China's Pinglu Canal Estuary. *Forests,* v. 14, n. 9, 1838. 2023.

PARENTE, L.L.; MESQUITA, V.V.; MIZIARA, F.; BAUMANN, L.F.; FERREIRA, L.G. Assessing the pasturelands and livestock dynamics in Brazil, from 1985 to 2017: A novel approach based on high spatial resolution imagery and Google Earth Engine cloud computing. *Remote Sensing of Environment,* v. 232, p. 111301, 2019. https://doi.org/10.1016/j.rse.2019.111301

PEARSON, R. L.; L. D. MILLER. Remote mapping of standing crop biomass for estimation of the productivity of the shortgrass prairie. *In* 8th International Symposium on Remote Sensing of Environment, 1972, University of Michigan, Ann Arbor .

PINTY, B.; VERSTRAETE, M. M. GEMI: a non-linear index to monitor global vegetation from satellites. *Vegetation,* 101, 15-20, 1992.

PONTIUS, R. G. *Metrics That Make a Difference.* Berlin/Heidelberg: Springer, 2022. p. 1–130.

POTAPOV, P.; LI X.; HERNANDEZ-SERNA, A.; TYUKAVINA, A.; HANSEN, M. C.; KOMMAREDDY, A.; PICKENS, A.; TURUBANOVA, S.; TANG, H.; SILVA, C.E.; ARMSTON, J.; DUBAYAH, R.; BLAIR J.B.; HOFTON, M. Mapping global forest canopy height through integration of GEDI and Landsat data. *Remote Sens. Environ.*, v. 253, 112165, 2021. https://doi.org/10.1016/j.rse.2020.112165 .

PRATES, A. P. L. Atlas dos recifes de coral nas unidades de conservação brasileiras. 2. ed. Brasília: MMA, 2006. 232 p.

PRATES, A. P. L.; GONÇALVES, M. A.; ROSA, M. R. *Panorama da conservação dos ecossistemas costeiros e marinhos no Brasil.* Brasília: MMA/SBF/GBA, 2010.

QUEIROZ, L. et al. Shrimp aquaculture in the federal state of Ceará, 1970–2012: Trends after mangrove forest privatization in Brazil. *Ocean and Coastal Management,* v. 73, p. 54-62, 2013.

RIKIMARU, A.; ROY, P. S.; MIYATAKE, S. Tropical forest cover density mapping. *Tropical ecology,* v. 43, n. 1, p. 39-47, 2002.

ROESCH L. F. W. *et al.* The Brazilian Pampa: A fragile biome. *Diversity,* v.1, p. 182–198, 2009. doi:10.3390/d1020182.

RONDEAUX, G.; STEVEN, M.; BARET, F. Optimization of soil-adjusted vegetation indices. *Remote sensing of environment,* v. 55, n. 2, p. 95-107.

ROUSE, J. W. Jr.; HAAS, R. H. ; SCELL, J. A.; DEERING, D. W.; J. C. HARLAN. Monitoring the Vernal Advancement of Retrogradiation of Natural Vegetation, NASA/GSFC, 371, 1974.

SANO, E. E.; RODRIGUES, A. A.; MARTINS, E. S.; BETTIOL, G. M.; BUSTAMANTE, M. M. C.; BEZERRA, A. S.; COUTO, A. F.; VASCONCELOS, V.; SCHÜLER, J.; BOLFE, E. L. Cerrado Ecoregions: A spatial framework to assess and prioritize Brazilian savanna environmental diversity for conservation. *Journal of Environmental Management*, v. 232, p. 818–828, 2019.

SILVA, J.; ABDON, M. Delimitação do Pantanal Brasileiro e suas sub-regiões. *Pesq. Agropec. Bras.,* v. 33, 2022, pp. 1–130, 1998.

SGB/CPRM. Dados, informações e produtos do Serviço Geológico do Brasil . 2024. Available at https://geosgb.sgb.gov.br/.

SOS MATA ATLÂNTICA. *Atlas dos remanescentes florestais da Mata Atlântica, período 2019-2020.* São Paulo: *Fundação SOS Mata Atlantica.* Instituto Nacional das Pesquisas Espaciais, 2020.

SOUZA Jr. C. M, ROBERTS, D. A.; COCHRANE, M. A. Combining spectral and spatial information to map canopy damage from selective logging and forest fires. *Remote Sens. Environ.,* v. 98, p. 329–343, 2005. doi: 10.1016/j.rse.2005.07.013

SOUZA, C.M., JR.; Z. SHIMBO, J.; ROSA, M. R.; PARENTE, L. L.; ALENCAR, A.; RUDORFF, B. F. T.; HASENACK, H.; MATSUMOTO, M.; G. FERREIRA, L.; SOUZA-FILHO, P. W. M.; DE OLIVEIRA, S. W.; ROCHA, W. F.; FONSECA, A. V.; MARQUES, C. B.; DINIZ, C. G.; COSTA, D.; MONTEIRO, D.; ROSA, E. R.; VÉLEZ-MARTIN, E.; WEBER, E. J.; LENTI, F. E. B.; PATERNOST, F. F.; PAREYN, F. G. C.; SIQUEIRA, J. V.; VIERA, J. L.; NETO, L. C. F.; SARAIVA, M. M.; SALES, M. H.; SALGADO, M. P. G.; VASCONCELOS, R.; GALANO, S.; MESQUITA, V. V.; AZEVEDO, T. Reconstructing Three Decades of Land Use and Land Cover Changes in Brazilian Biomes with Landsat Archive and Earth Engine. *Remote Sensing,* v. 12, 2735, 2020. https://doi.org/10.3390/rs12172735.

STEHMAN, S. V. Sampling designs for accuracy assessment of land cover. *International Journal of Remote Sensing,* v. 30, p. 5243-5272 , 2019. doi:10.1080/01431160903131000.

STEHMAN, S. V. Estimating area and map accuracy for stratified random sampling when the strata are different from the map classes. *International Journal of Remote Sensing,* v. 34, p. 4923-4939, 2014. doi:10.1080/01431161.2014.930207

TENÓRIO, G. S. et al. Mangrove shrimp farm mapping and productivity on the Brazilian Amazon coast: Environmental and economic reasons for coastal conservation. *Ocean & Coastal Management*, v. 104, p. 65–77, 2015.

THOMAS, N. et al. Distribution and drivers of global mangrove forest change, 1996–2010. *PLOS ONE*, v. 12, n. 6, p. e0179302, 8 jun. 2017.

TORCHVISION CONTRIBUTORS. DeepLabV3 implementation in torchvision. 2023. Available at https://github.com/pytorch/vision/blob/main/torchvision/models/segmentation/deeplabv3.py.

TUCKER, C. J.; ELGIN JR, J. H.; MCMURTREY III, J. E.; FAN, C. J. Monitoring corn and soybean crop development with hand-held radiometer spectral data. *Remote Sensing of Environment,* v. 8, n.3, p. 237–248, 1979. https://doi.org/10.1016/0034-4257(79)90004-x.

UNEP-WCMC, WorldFish Centre, WRI, TNC. Global distribution of warm-water coral reefs, compiled from multiple sources including the Millennium Coral Reef Mapping Project. Version 4.1. Includes contributions from IMaRS-USF and IRD (2005), IMaRS-USF (2005) and Spalding et al. (2001). Cambridge (UK): UN Environment World Conservation Monitoring Centre, 2021. Data DOI: https://doi.org/10.34892/t2wk-5t34

USGS. Landsat Collection 1 Level 1 Product Definition. 2017. Available online: https://landsat.usgs.gov/sites/default/files/documents/LSDS-1656_Landsat_Level-1_Product_Collection_Definition.pdf.

VINCINI, M.; FRAZZI, E. R. M. E. S.; D'ALESSIO, P. A. O. L. O. . A broad-band leaf chlorophyll vegetation index at the canopy scale. *Precision Agriculture*, 9, 303-319, 2008.

WEBER, E. J.; HOFFMANN, G .S.; OLIVEIRA, C. V.; HASENACK, H. (Orgs.). *Uso e cobertura vegetal do Estado do Rio Grande do Sul – situação em 2009.* Porto Alegre: UFRGS IB Centro de Ecologia, 2016. 1a ed. ISBN 978-85-63843-20-3. Available at http://www.ecologia.ufrgs.br/labgeo.

YAMAZAKI, D.; IKESHIMA, D.; TAWATARI, R.; YAMAGUCHI, T.; O'LOUGHLIN, F.; NEAL, J. C.; SAMPSON, C. C.; KANAE, S.; BATES, P. D. A high accuracy map of global terrain elevations. *Geophysical Research Letters,* v. 44, p. 5844-5853, 2017. https://doi.org/10.1002/2017GL072874

XIAO, X., et al. 2005. Mapping paddy rice agriculture in southern China using multi-temporal MODIS images. *Remote sensing of environment*, v. 95, n. 4, 480-492.

ANNEXES

Annex I: MapBiomas Network

MapBiomas is an initiative of the Climate Observatory, involving a collaborative network of NGOs, universities and technology companies organized by biomes and cross-cutting themes.

Biomes Coordination:

- Amazon – Institute of People and Environment of the Amazon (IMAZON)
- Caatinga – State University of Feira de Santana (UEFS), Northeast Plants Association (APNE), and Geodatin
- Cerrado – Amazon Environmental Research Institute (IPAM)
- Atlantic Forest – Foundation SOS Atlantic Forest and ArcPlan
- Pampa – Federal University of Rio Grande do Sul (UFRGS) and GeoKarten
- Pantanal – Institute SOS Pantanal and ArcPlan

Cross-cutting Themes Coordination:

- Pasture – Federal University of Goias (LAPIG/UFG)
- Agriculture – Agrosatelite until collection 8. Remap in collection 9.
- Coastal Zone and Mining – Vale Technological Institute (ITV) and Solved
- Urban Area – University of São Paulo (USP - QUAPÁ-FAU and YBY), Federal University of Bahia (UFBA) and Federal University of São Carlos (UFSCar - NEEPC)

Technology Partners:

- Google
- EcoStage
- Ecode

Financing:

- Amazon Fund
- Arapyaú Institute
- Children’s Investment Fund Foundation (CIFF)
- Climate and Land Use Alliance (CLUA)
- Good Energies Foundation
- Gordon & Betty Moore Foundation
- Humanize Institute
- Institute for Climate and Society (iCS)
- Montpelier Foundation

- Mulago Foundation
- Norway’s International Climate and Forest Initiative (NICFI)
- Global Wildlife Conservation (GWC)
- OAK Foundation
- Quadrature Climate Foundation (QCF)
- Sequoia Foundation
- Skoll Foundation
- Walmart Foundation
- Woods & Wayside International

Institutional Partners:

- Arapyaú Institute
- MapBiomas Support Institute (IAMap)
- WRI Brasil
- AVINA Foundation

General Coordination: Tasso Azevedo (SEEG/OC)

Technical Coordination: Marcos Rosa (ArcPlan)

Scientific Coordination: Julia Shimbo (IPAM)

The project counts on an Independent Committee of Scientific Advice composed by renowned specialists:

- Dr. Alexandre Camargo Coutinho (Embrapa)
- Dr. Edson Eygi Sano (IBAMA)
- Dr. Gerd Sparovek (University of São Paulo)
- Dra. Leila Maria Garcia Fonseca (INPE)
- Dra. Liana Oighenstein Anderson (CEMADEN)
- Dra. Marina Hirota (Federal University of Santa Catarina)

Former members:

- Dr. Gilberto Camara Neto (INPE)
- Dr. Joberto Veloso de Freitas (Federal University of Amazonas)
- Dr. Matthew C. Hansen (Maryland University)
- Dr. Mercedes Bustamante (University of Brasília)
- Dr. Timothy Boucher (TNC)
- Dr. Robert Gilmore Pontius Jr (Clark University)

Technical Partners:

- Institute of Agricultural and Forest Management and Certification - Imaflora (IMAFLORA)

- Energy and Environment Institute (IEMA)
- Socioambiental Institute (ISA)
- Institute for Democracy and Sustainability (IDS)
- The Nature Conservancy (TNC)
- Life Center Institute (ICV)
- WWF Brasil
- Brasil I.O

Annex II

MapBiomas 10 m Collection 2 (beta) prevalence rules, given by the “Prevalence ID” (from the most to the less prevalent class), used for integrating biomes and cross-cutting themes maps. Since some classes are mapped both as cross-cutting themes and in the biomes, the “Source” column indicates the source of information for that specific rule. “Exceptions” are classes that are prevalent over the listed one in that region.

CLASS ID	CLASS NAME	SOURCE	PREVALENCE ID	EXCEPTION
30	4.3. Mining	Mining	1	In São Paulo and Mato Grosso states, the cross-cutting class 24 is prevalent
23	4.1. Beach, Dune, and Sand Spot	Coastal Zone	2	
23	4.1. Beach, Dune, and Sand Spot	Biomes	3	
5	1.3. Mangrove	Coastal Zone	4	
31	5.2. Aquaculture	Coastal Zone	5	
32	2.3. Hypersaline Tidal Flat	Coastal Zone	6	
24	4.2. Urban Area	Urban Area	7	
9	3.3. Forest Plantation	Agriculture	8	At Lagoa dos Peixes (Pampa), the classes 3, 11, 12, 29, 33, 49, 50 are prevalent
9	3.3. Forest Plantation	Biomes	9	
29	2.4. Rocky Outcrop	Biomes	10	
20	3.2.1. Temporary crop	Agriculture	11	Where both biomes and MapBiomas Water indicates class 33, class 33 is prevalent
50	2.5. Herbaceous Sandbank Vegetation	Biomes	12	
25	4.4. Other Non Vegetated Areas	Biomes	13	

33	5.1. River, Lake and Ocean	Water	14	
33	5.1. River, Lake and Ocean	Biomes	15	
3	1.1. Forest Formation	Biomes	16	
4	1.2. Savanna Formation	Biomes	17	Outside protected areas in Cerrado the class 15 is prevalent
49	1.5. Wooded Sandbank Vegetation	Biomes	18	
6	1.4. Floodable Forest	Biomes	19	
11	2.1 Wetland	Biomes	20	Outside protected areas in Cerrado the class 15 is prevalent
12	2.2. Grassland Formation	Biomes	21	Outside protected areas in Cerrado the class 15 is prevalent
15	3.1. Pasture	Pasture	22	n Pantanal class 21 is converted to 15. In Pampa, class 21 is prevalent. Within protected areas in Cerrado natural classes are prevalent.
15	3.1. Pasture	Biomes	23	
21	3.4. Mosaic of Uses	Biomes	24	
36	3.2.2. Perennial crop	Agriculture	25	
25	4.4 Other non vegetated areas	Biomes	26	