Cerrado - Appendix

Collection 11

Version 1

General coordinator

Ane A. Alencar

Team

Ana G. P. de Souza

Bárbara C. da Silva

Dhemerson E. Conciani

Joaquim J. S. P. Pereira

Julia Z. Shimbo

Vera L. S. Arruda

Wallace V. da Silva

August, 2026

1. OVERVIEW OF THE CERRADO CLASSIFICATION METHOD

The land use and land cover (LULC) classification methodology developed for the Cerrado biome under the MapBiomas project relies primarily on a tree-based ensemble machine-learning algorithm to generate annual maps of the natural land-cover classes. These classes are organized into six categories: forest formation, savanna formation, wetland, grassland formation, herbaceous sandbank vegetation and rocky outcrop. Since the first MapBiomas collection, the methodology has undergone continuous refinement, leading to substantial improvements in classification accuracy, spatial consistency, and temporal stability. Although recent collections have introduced important methodological advances, the workflow has also sought to preserve and consolidate the improvements achieved in previous versions. The Cerrado classification workflow comprises several key stages. First, identify the optimal time window to generate annual Landsat mosaics. Then, remote sensing metrics were defined as potential predictors. Reference training samples were generated per year and region to train the classification algorithm. Post-classification treatments were applied to remove noise and produce a consistent time series. Finally, the resulting maps were integrated with other cross-cutting themes. The final results are evaluated through visual inspection and accuracy assessment. A detailed overview of the methodological evolution of the Cerrado classification is in Table 1. This document is intended as a technical appendix to the general MapBiomas ATBD and provides a detailed description of the methodological procedures adopted for LULC mapping in Cerrado.

Table 1. Summary of methodological evolution across MapBiomas Cerrado collections. EDT: Empirical Decision Tree; RF: Random Forest; SR: Surface Reflectance; NV: Native Vegetation.

Collection	Year Range	Method	Mapped classes	Key Improvements	Global Accuracy*
1.0	2008 – 2015	EDT	Forest	First version of the Cerrado collection.	- -
2.0	2000 – 2016	EDT	Forest, Savanna,Grassland	Inclusion of new NV classes (savanna formation and grassland formation).	- -
2.3	2000 – 2016	RF	Forest, Savanna, Grassland, Mosaic of Agriculture and Pasture, Other Non-Vegetated Area, Water	Introduction of Random Forest (RF); Auxiliary classes (mosaic, non-vegetated areas, and water); Training samples from stable areas.	- -
3.0	1985 – 2017	RF	Same as Collection 2.3	Full Landsat series (since 1985); Improved training sample quality with outlier removal.	- -
3.1	1985 – 2017	RF	Same as Collection 3.0	The classification by ecoregions (38) was used in the substitution of regular tiles.	74.4%
4.0	1985 – 2018	RF	Same as Collection 3.1	Improvement in training samples quality by using new reference maps.	73.5%

Collection	Year Range	Method	Mapped classes	Key Improvements	Global Accuracy*
4.1	1985 – 2018	RF	Forest, Savanna, Grassland, Pasture, Agriculture, Other Non-Vegetated Area; Water	Variable importance for feature selection; Enhanced post-processing temporal filter; Significant gains in the accuracy of NV mapping.	74.8%
5.0	1985 – 2019	RF	Same as Collection 4.1	New reference maps; Improvements in spatial continuity between classification regions New vegetation dynamic product (deforestation and secondary vegetation).	75.0%
6.0	1985 – 2020	RF	Forest, Savanna, Wetland, Grassland, Mosaic of Agriculture and Pasture, Other Non-Vegetated Area, Water	New NV class (wetland); New reference maps; Landsat SR mosaics; Feature space refinement.	76.0%
7.0	1985 – 2021	RF	Forest, Savanna, Wetland, Grassland, Rocky Outcrop, Mosaic of Uses, Agriculture and Pasture, Other Non-Vegetated Area, Water	New NV class (rocky outcrop); New reference maps; GEDI for outlier training samples filtering; Revision of classification regions (38); Significative accuracy gains in NV mapping.	77.8%
7.1	1985 – 2021	RF	Same as Collection 7.0	Improvement of temporal filter rules in the last year (2021).	77.8%
8.0	1985 – 2022	RF	Forest, Savanna, Wetland, Grassland, Rocky Outcrop, Mosaic of Uses, Other Non-Vegetated Area, Water	New reference maps; Regionalization of hyperparameters and classification; Revision of the temporal filtering rules; Territorial expansion of rocky outcrop classification.	78.0%
9.0	1985 – 2023	RF	Same as Collection 8.0	New reference maps; Multiprobability approach; Post-processing filters revision; New false regrowth filter; New rocky outcrop processing workflow.	79.1%
10.0	1985-2024	RF	Forest, Savanna, Wetland, Grassland, Rocky Outcrop, Herbaceous Sandbank Vegetation, Mosaic of Uses, Other Non-Vegetated Area, Water	New reference maps; New NV Class (herbaceous sandbank vegetation); Monthly Landsat SR mosaics; Geomorphometric filter; Redefined rocky outcrop class concept.	79.4%
10.1	1985-2024	RF	Same as Collection 10.0	Correction in the Amazon and Pampa biomes.	79.4%
11.0	1985-2025	RF	Same as Collection 10.0	New reference maps; Review of classification regions; Temporal Training Window, Spatial-context metrics, New post-processing sequence; Review of rocky outcrop concept.	79.9%

* Overall accuracy for level 2 of the legend. Collections 1 through 3.0 do not have accuracy metrics.

In Collections 1.0 and 2.0, empirical decision trees were applied, with rule-based thresholds informed by expert knowledge of the spectral characteristics of each class. Collection 1.0 covered the period 2008–2015 and was released in 2016. Collections 2.0 and 2.3 extended the temporal range to 2000–2016, with Collection 2.3 (a revised version of Collection 2.0) introducing the use of the Random Forest (RF) algorithm. Empirical decision trees were used to generate stable training samples for the 2000–2016 period, which in turn were used to train the RF classifier across the full time-series. Collections 3.0 and 3.1 expanded the historical coverage to include the years 1985–2017, and a methodological paper was published (Alencar et al., 2020). Collections 4.0 and 4.1 (1985-2018) demonstrated a notable enhancement in mapping precision compared to their predecessors. These improvements were largely driven by the refinement of the feature space, which was redesigned based on a variable importance analysis using the RF algorithm. Additionally, new post-classification filters were implemented to better account for temporal consistency and reduce noise in the time series, resulting in more accurate detection of land cover transitions. From Collection 2.3 onward, training samples were no longer generated using empirical trees but were instead based on stable samples from previous collections, contributing to improved classification accuracy.

Beginning with Collection 5.0 (1985-2019), official reference maps (PRODES Cerrado and PROBIO) were incorporated to spatially constrain training sample collection for NV classes, minimizing bias. Collection 6.0 (1985-2020) introduced several important methodological updates: the classified time series was extended to encompass the period from 1985 to 2020, a new NV class (wetland) was introduced, the surface reflectance mosaic was implemented, and the feature space was refined. In addition to the reference maps already used in Collection 5.0, the *"Inventário Florestal do Estado de São Paulo"* was incorporated to support training sample filtering for NV classes. In Collection 7.0 (1985–2021), a new class (rocky outcrop) was introduced. The training dataset was refined through the application of an outlier detection filter based on GEDI Global Forest Height data (Lang et al., 2022), and the incorporation of the *"Base Temática Digital do Estado do Tocantins"* as an additional reference map. These improvements enhanced the quality and reliability of training samples. The reference maps adopted in Collection 6.0 were maintained. Unlike in Collection 6.0, where the wetland class was treated as a pseudo cross-cutting theme, it was directly integrated into the main classification process in Collection 7.0. The Collection 7.1 was released as a reprocessed version of Collection 7.0 to correct temporal filtering issues in the final year of the series. It addressed the overestimation of deforestation peaks in 2021 caused by misclassified transitions.

Collection 8.0 (1985–2022) incorporated substantial methodological advances. The mask used for training sample selection was improved through the inclusion of deforestation alerts from MapBiomas Alerta and *Sistema de Alerta de Desmatamento* (SAD) Cerrado (2019–2022), enhancing the filtering process for anthropogenic classes.

Additionally, the classification approach was strengthened by the regionalization of both the hyperparameters and the classification process itself, allowing for better adaptation to local biophysical conditions. The classification of the rocky outcrop class was also expanded territorially, ensuring its representation across the entire Cerrado using a more robust and spatially consistent methodology. Collection 9.0 (1985–2023) continued this strategy, updating training sample filtering with alerts from MapBiomas Alerta and SAD Cerrado (2019–2023). New reference datasets were introduced, such as the *"Mapa de Uso e Cobertura da Terra do Distrito Federal"* and the *"Mapeamento dos Remanescentes de Campos de Murundus do Estado de Goiás"*. In this collection, the regionalization of hyperparameters was discontinued, while the regional classification scheme was maintained. Moreover, the classification process adopted a multiprobability approach, which considers the probability distribution of all classes instead of relying on majority voting, thereby improving the reporting of classification uncertainties. Post-processing routines were revised, and a new false regrowth filter was introduced to prevent spurious transitions in recent years. The rocky outcrop class was also expanded using a revised classification approach, updated training samples, and new predictor variables.

In Collection 10.0 (1985–2024), training-sample filtering was updated to incorporate MapBiomas Alerta deforestation alerts for 2019–2024, while SAD Cerrado data used in previous collections were discontinued. The training masks were also refined through the inclusion of the "*Mapa de Veredas e Áreas Úmidas do Sudoeste do Tocantins*" and the "*Cobertura e Uso da Terra do Sudoeste do Tocantins*". Monthly Landsat surface reflectance mosaics were adopted, while the Random Forest classification continued to use the multiprobability approach. Post-processing improvements included a specific procedure for mapping herbaceous sandbank vegetation and a geomorphometric filter designed to remove wetland misclassifications associated with terrain shadows and relief. The rocky outcrop concept was also revised to include only exposed rock, excluding rupestrian vegetation, unlike the broader definition applied in Collections 7.0 to 9.0. Collection 10.1 was subsequently processed to incorporate adjustments affecting the Amazon and Pampa biomes, without methodological or mapping changes for the Cerrado.

In the current Collection 11.0 (1985–2025), the filtering of training samples was updated to incorporate deforestation alerts from MapBiomas Alerta (2019–2025). The same reference maps used in the previous collection continued to support the construction and refinement of the training masks, with the addition of the "*Mapa de Vegetação do Parque Nacional da Chapada das Mesas*", the "*Mapa de Veredas do Mato Grosso*", and the "*Mapa de Vegetação da Fazenda Água Limpa - UnB*". A major innovation in the training-sample procedure was the adoption of a temporal-window-based stability approach. Monthly Landsat surface reflectance mosaics were retained, as was the Random Forest classifier operating in multiprobability mode. The main methodological change introduced in Collection 11.0 was a comprehensive review and reorganization of the post-processing filters, with existing spatial and temporal filters reorganized and new filters incorporated to address specific classification inconsistencies. The filtering workflow, processing sequence, class-specific rules, and interactions among filters were revised and standardized with the same logical sequence to that adopted in MapBiomas 10 m Collection 4.0. The conceptual definition of the rocky outcrop class was also revised. Areas associated with rupestrian vegetation were reincorporated into this class, reversing the narrower definition adopted in Collection 10.0. Collection 11.0 therefore restores the broader rocky outcrop concept applied in Collections 7.0 to 9.0. The classification and post-processing scripts employed in this collection are available at https://github.com/mapbiomas/brazil-cerrado.

2. METHODOLOGICAL DESCRIPTION

In MapBiomas Collection 11.0, the classification of Landsat mosaics for the Cerrado biome initially comprised ten land use and land cover classes, as defined in the official MapBiomas legend and summarized in Table 2. These classes included native vegetation (NV) types, water bodies, and anthropogenic land uses. Agriculture and pasture were classified separately during the initial stages of the workflow, although their individual mapping was not the primary objective of the Cerrado-specific methodology, as these land-use classes are subsequently produced and refined by specialized cross-cutting teams. Their inclusion in the initial classification was intended to improve the discrimination between native vegetation and anthropogenic areas, thereby reducing omission and commission errors and minimizing spectral confusion between natural and managed land-cover types. Mining areas, urbanized areas, and beaches and dunes were grouped into the other non-vegetated areas class. Then, during post-processing, the agriculture and pasture classes were merged into the mosaic of uses class (21). The subsequent sections of this document provide a detailed description of each major component of Collection 11.0 classification methodology. An overview of the workflow is in Figure 1. The core methodological steps can be summarized as follows:

- **Landsat Mosaic:** Annual multi-temporal surface reflectance mosaics were generated from harmonized monthly Landsat imagery to ensure optimal spectral quality, minimize atmospheric noise, and maintain temporal consistency.

- **Training Samples:** Stable-pixel masks are constructed for four overlapping temporal windows based on Collection 8.0 (1985–2022), Collection 9.0 (1985–2023), and Collection 10.1 (1985–2024), complemented by thematic reference data and GEDI-derived information. A stratified sampling design ensured proportional class representation across regions and years.

- **General Classification:** Land cover and land use (LULC) classification was performed using the Random Forest (RF) algorithm in Google Earth Engine (GEE), applying a multiprobability approach to capture class uncertainty. Post-classification filters were applied to correct temporal inconsistencies and remove spatial artifacts.

- **Rocky Outcrop Classification:** A dedicated classification workflow was developed to map rocky outcrop. Training samples were visually validated, and classification was performed using RF. Temporal and spatial filters refined the final output.

- **Internal Integration and Paraguay River Basin harmonization:** Two outputs were produced: (i) the general LULC map and (ii) the rocky outcrop map. During the final integration stage, the rocky outcrop classification was overlaid on the general map. In addition, specific spatial corrections were applied within the Paraguay River Basin (BAP) to harmonize the annual maps along the Pantanal biome.

- **Integration with Cross-Cutting Themes:** Annual biome maps were harmonized with thematic layers following predefined MapBiomas prevalence rules to ensure consistency and avoid overlaps between classes.

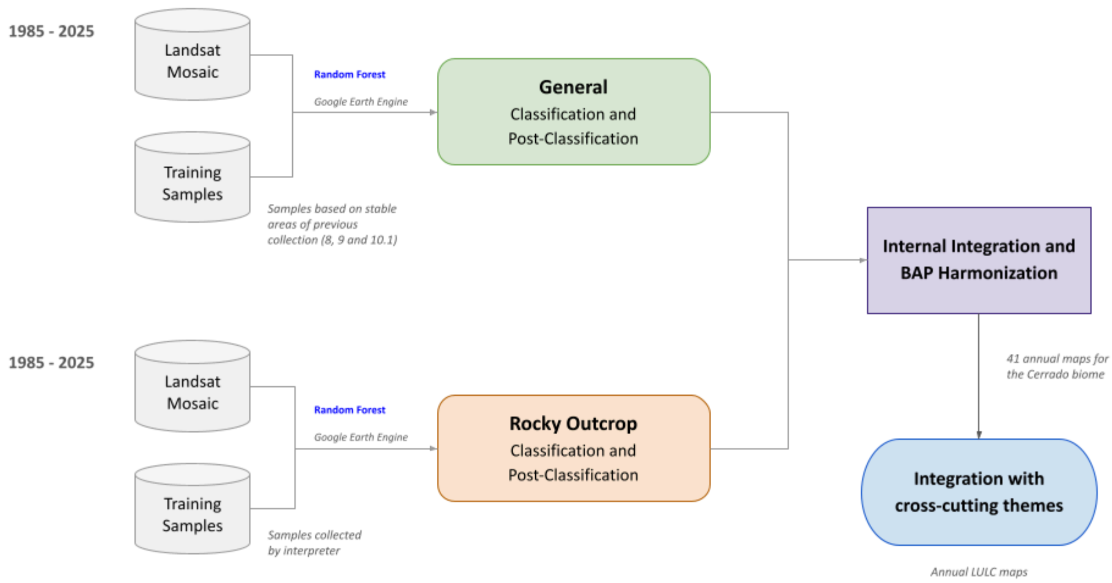
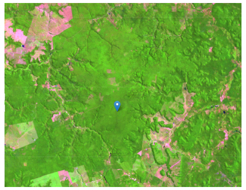
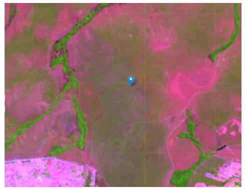
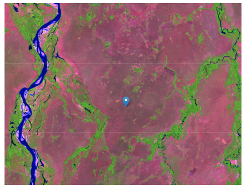
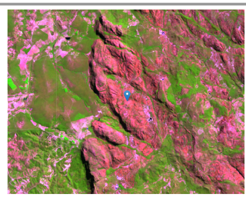
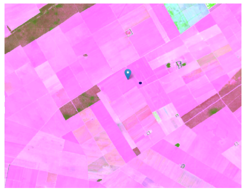
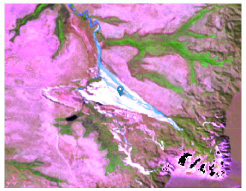
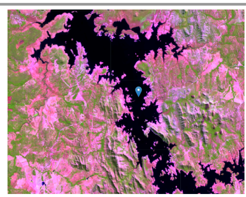


Figure 1. Workflow for the LULC classification of the Cerrado biome. Two parallel modules were applied: (i) the general classification, and (ii) a specific classification of rocky outcrop. Outputs were integrated into two stages: first, the internal integration combined the rocky outcrop with the general map; second, the maps were integrated with cross-cutting themes.

Table 2. Land use and land cover categories used for the Landsat classification for the Cerrado biome in MapBiomas Collection 11.0.

Classes Level 1	Classes Level 2	ID	Legend Color	RGB composite (SWIR1-NIR-Red)	Description
Forest	Forest Formation	3			Vegetation types characterized by the predominance of tree species forming a continuous canopy. This includes Riparian Forests, Gallery Forests, Dry Forests, and Forested Savannas (Ribeiro and Walter, 2008), as well as Semi-deciduous Seasonal Forests.
	Savanna Formation	4			Vegetation types with a distinct stratification of tree, shrub, and herbaceous stratum. This includes different physiognomies of Cerrado *sensu stricto* (Dense, Typical, Sparse, and Rupestrian Savanna) (Ribeiro and Walter, 2008). The class may also include native Cerrado *sensu stricto* areas suitable for grazing, if they maintain their natural structure and composition.
Herbaceous and Shrubby Vegetation	Wetland	11			Ecosystems dominated by herbaceous vegetation subject to seasonal flooding or constant fluvial/lacustrine influence. Examples include *Campo Úmido* and *Brejo*. In some areas, the herbaceous matrix is interspersed with savanna tree species (e.g., *Parque de Cerrado*) or palm species, such as in *Veredas* and *Palmeirais* (Ribeiro and Walter, 2008).
	Grassland	12			Open vegetation dominated by herbaceous species, with minimal or no tree cover. Includes the Dirty, Clean and Rupestrian Grassland phytophysiognomies (Ribeiro and Walter, 2008).
	Rocky Outcrop	29			Naturally exposed rocky surfaces, including monoliths, bedrock, and slabs with little or no soil cover, these features are typically associated with stable geological formations of sedimentary, igneous, or metamorphic origin. Includes areas of rupestrian vegetation. This class also includes areas of rupestrian vegetation associated with these environments, such as *Cerrado Rupestre* and *Campo Rupestre* (Ribeiro and Walter, 2008).

Classes Level 1	Classes Level 2	ID	Legend Color	RGB composite (SWIR1-NIR-Red)	Description
	Herbaceous Sandbank Vegetation*	50			Coastal sandy plain ecosystems characterized by predominantly herbaceous and shrubby vegetation, with sparse shrub distribution.
Farming	Pasture**	15			Pasture area, predominantly planted, linked to cattle ranching activities.
	Agriculture**	18			Areas occupied with short to long vegetative cycle agricultural crops. This encompasses both perennial and temporary crops.
Non-Vegetated Area	Other Non-Vegetated Areas	25			Includes impermeable surfaces (e.g., roads, buildings, mining infrastructure), exposed soil in natural settings (e.g., erosion features, gullies, landslides), and croplands in the off-season.
Water	River, Lake, and Ocean	33			Rivers, lakes, dams, reservoir, and other water bodies.

*Herbaceous Sandbank Vegetation is added via post-classification method; ** Agriculture and Pasture are merged into Mosaic of Uses during the first spatial filter.

3. LANDSAT IMAGE MOSAICS

The first step in the classification of native vegetation (NV) in the Cerrado biome consists of generating Landsat-annual mosaics that serve as the primary input for the LULC classification process. In earlier MapBiomas collections, up to Collection 5.0, the annual mosaics were generated using top-of-atmosphere (TOA) reflectance data from Landsat 5 (TM), Landsat 7 (ETM+), and Landsat 8 (OLI). Starting with Collection 6.0, this approach was replaced with surface reflectance (SR) data from Landsat Collection 2, providing atmospherically corrected and radiometrically consistent observations across sensors and years. In these collections, to ensure temporal continuity and consistency from 1985 onward, a sensor-prioritization strategy was adopted. Landsat 5 (TM) SR data was prioritized for the period 1985 to 2010, except in 2001 and 2002, when TM sensor issues led to the use of Landsat 7 (ETM+). Landsat 7 data remained the primary source for 2011 and 2012, followed by Landsat 8 (OLI) from 2013 onward. Up to Collection 9.0, annual mosaics were constructed using Level-2 surface reflectance images from Landsat Collection 2. Starting with Collection 10.0, the workflow adopted the **LANDSAT/COMPOSITES/C02/T1_L2_32DAY** dataset, which provides 32-day composites generated from orthorectified Landsat Collection 2, Tier 1, Level-2 observations. The product improved temporal continuity and image availability, particularly in the northern region of the biome in the early years. This data source and mosaic-construction procedure were maintained in Collection 11.0. For more information about this product, refer to the Earth Engine Data Catalog.

Since the first collection, the optimal temporal window for Landsat-mosaic generation has been systematically evaluated, considering the strong seasonality of Cerrado vegetation. Given that vegetation spectral responses differ significantly between the rainy and dry seasons, tests were conducted to compare image compositions from both the rainy and dry seasons (Figure 2). Imagery from the end of the rainy season (April to June), when vegetation remains vigorous and cloud cover tends to decrease, produced greener mosaics but also increased commission errors in forest formation class mapping. Conversely, mosaics generated from images acquired in the final months of the dry season (July to September) resulted in drier mosaics, which led to underestimation of forest coverage, primarily due to the reduced potential to map deciduous forests. The use of all available images throughout the year combines markedly different seasonal and phenological conditions within the same annual mosaic. This may increase spectral variability and mix observations representing senescence, seasonal flooding, fire occurrence, agricultural cycles, and exposed soil, producing annual metrics that are less representative of a consistent vegetation condition.

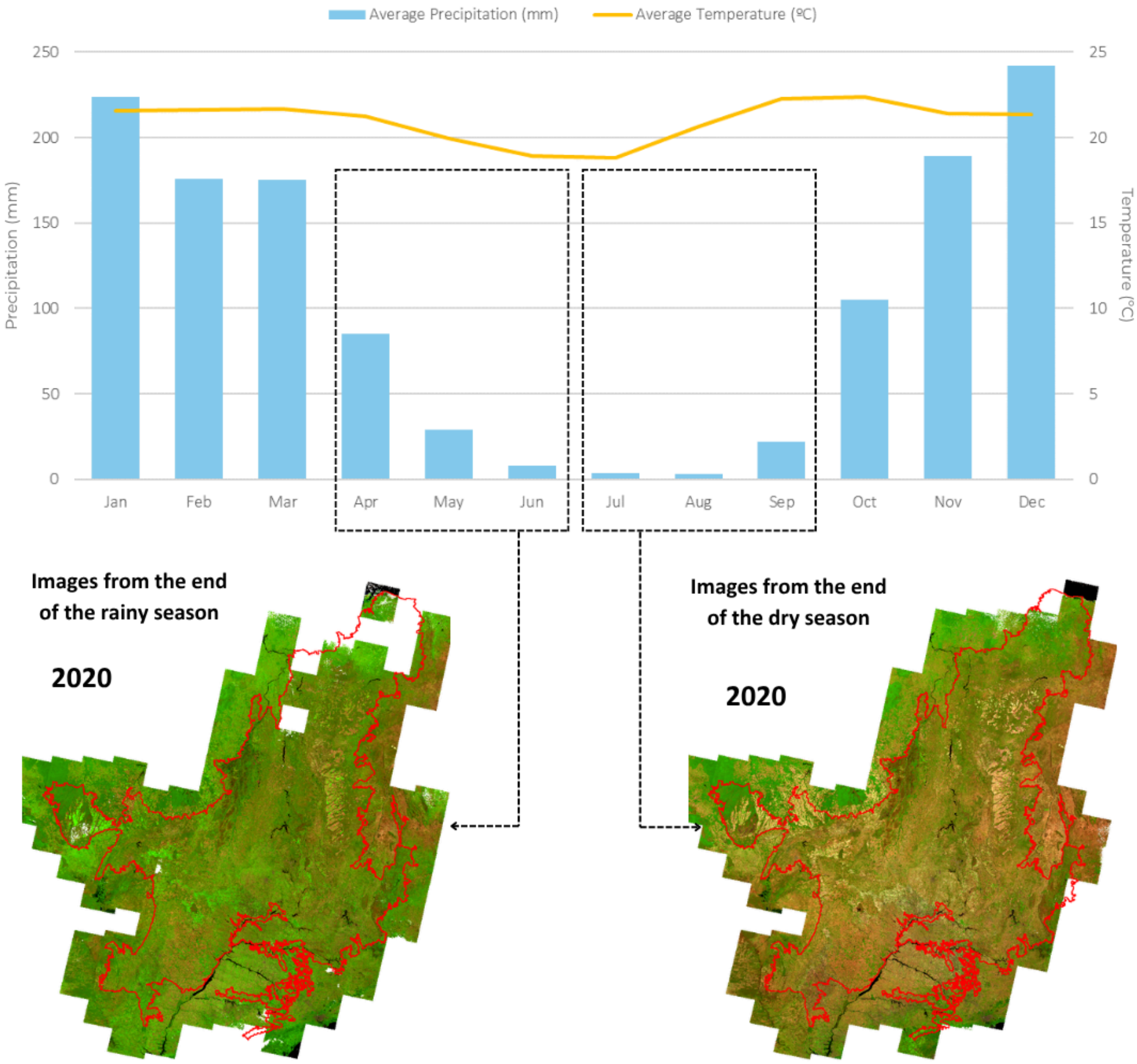


Figure 2. False-color composite of Landsat 8 (SWIR1-NIR-Red) mosaics illustrating spectral differences between the end of the rainy season and the end of the dry season in the Cerrado biome. Precipitation data represent monthly averages for the Cerrado region (Macena et al., 2008), while temperature data correspond to monthly averages for the Federal District (INMET, 2024).

Based on these findings, a standardized six-month compositing window (April to September) was adopted for all 38 classification regions and all years of the historical series. This broader window improved regional consistency and addressed limitations identified in narrower-period tests (Figure 3). The annual mosaics are generated by combining all valid Landsat observations available within the April–September window. For each pixel, reduction metrics such as median, amplitude, median deviation, and minimum are calculated and aggregated annually, resulting in a multi-band image used as input for the classification models. The final mosaics undergo visual inspection to ensure quality and suitability for subsequent classification steps.

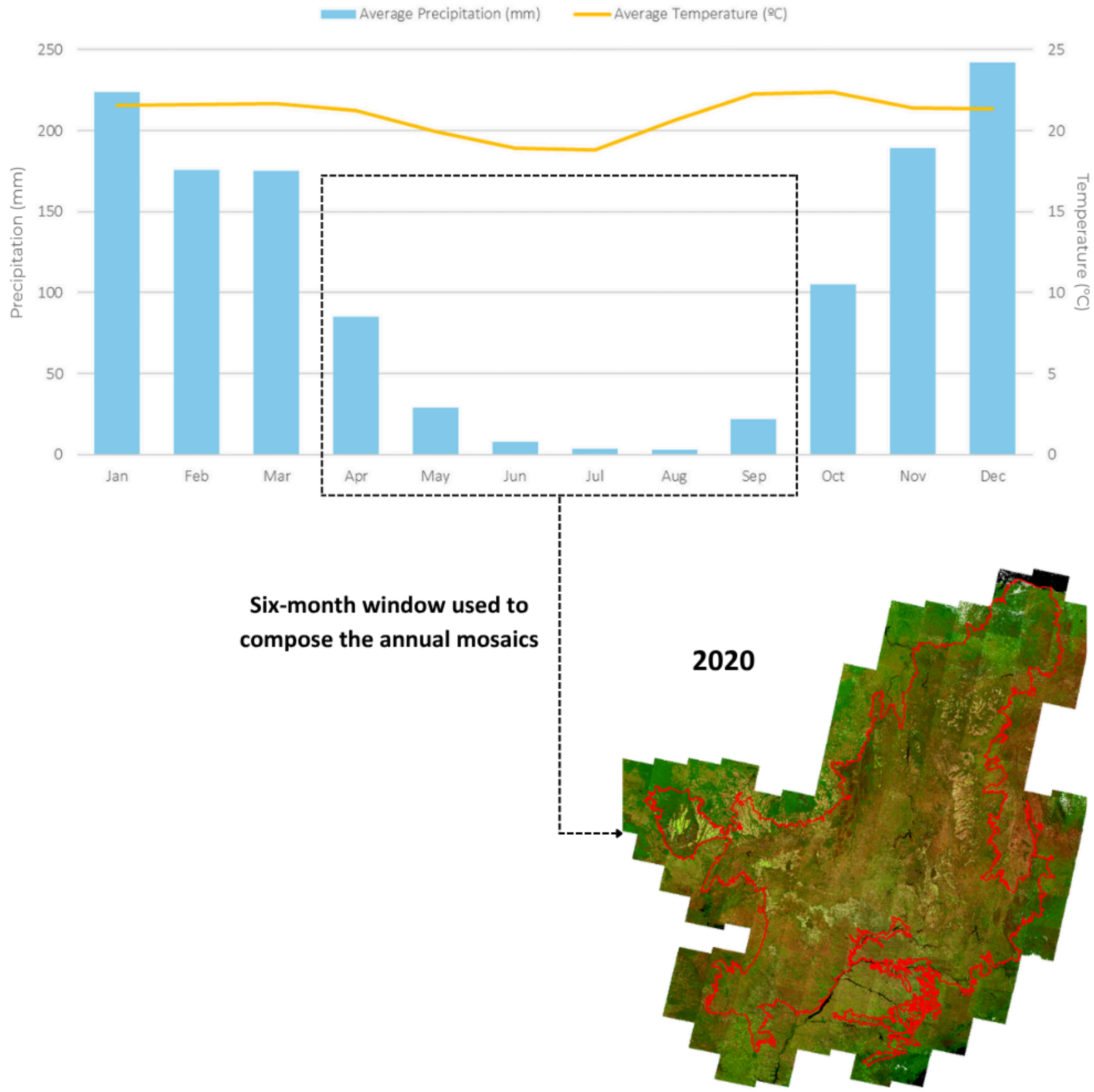


Figure 3. Time window adopted for the construction of annual classification mosaics in the Cerrado biome. Precipitation data represent monthly averages for the Cerrado region (Macena et al., 2008), and temperature data correspond to monthly averages for the Federal District (INMET, 2024).

4. GENERAL MAP CLASSIFICATION

The complete workflow for the general land use and land cover classification is illustrated in Figure 4. This workflow integrates multiple processing stages to ensure accurate and temporally consistent mapping across the Cerrado biome. The methodological steps are described in detail in the subsequent sections, starting with the delineation of classification regions (4.1), followed by the construction of the feature space used in model training (4.2). The sampling strategy and classification procedure are presented in section 4.3, which includes the generation of a training mask based on stable areas, the application of a stratified sampling design, and the classification using the Random Forest algorithm. Finally, the post-classification filtering steps applied to improve temporal and spatial consistency are detailed in section 5.

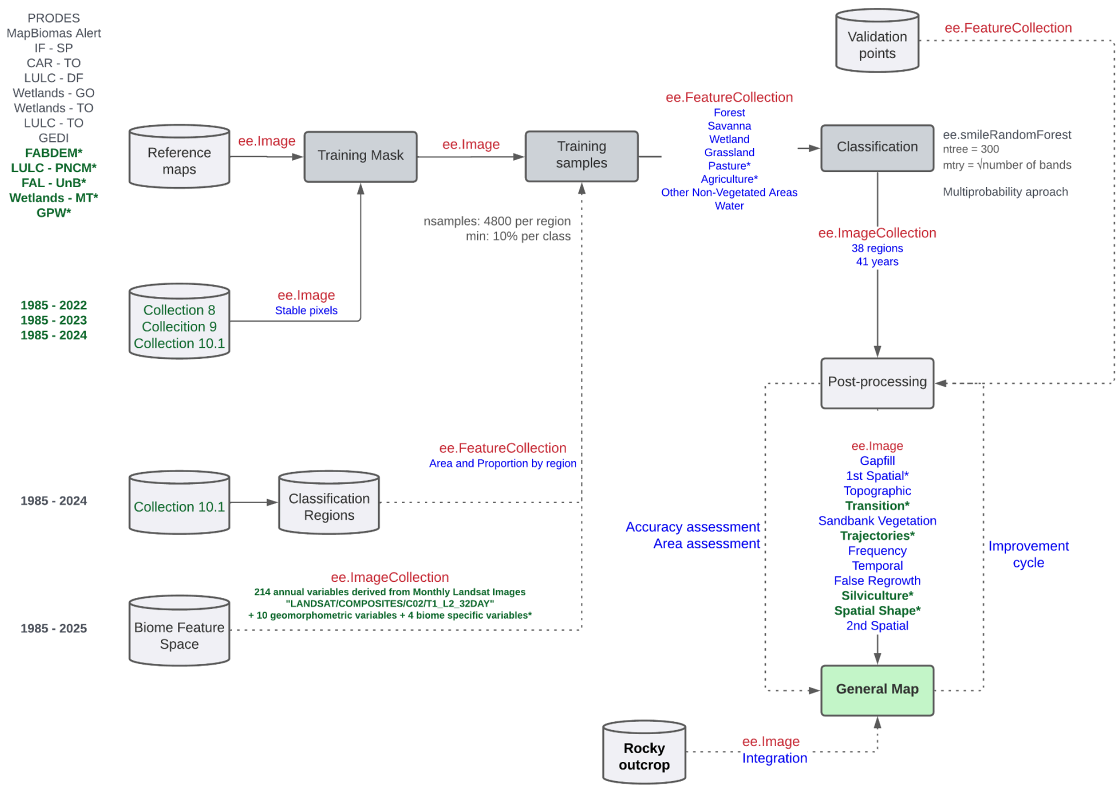


Figure 4. Each gray geometry (cylinders for databases and rectangles for processes) represents a key step in the classification schema, with the respective name inside. The gray text near databases and processes offers a description of the step, while the green text highlights the main innovations in Collection 11.0. 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 concise description of the asset content.

4.1. Classification regions

In the initial MapBiomas Cerrado collections, up to Collection 3.0, the classification process was organized using a regular grid based on a 1:250,000 scale, consisting of 172 classification tiles. Each tile was processed independently by the classification algorithm.

While this grid-based approach was operationally simple, it often led to discontinuities and inconsistencies along tile boundaries, reducing the spatial coherence of the resulting maps. Starting with Collection 3.1, a new regionalization strategy was implemented to better represent the ecological and biophysical heterogeneity of the Cerrado biome. The classification units were defined based on the ecoregions proposed by Sano et al. (2019) and further refined using Brazil's major hydrographic basins and the spatial distribution of land use and land cover classes in Collection 3.0 (2017). This integration resulted in the delineation of 38 classification regions, replacing the previous tile-based system and enabling a more ecologically meaningful and spectrally consistent classification.

In Collection 7.0, these classification regions were refined to incorporate phenological variability, particularly related to native vegetation seasonality (Figure 5). This refinement was based on an empirical analysis of Sentinel-2 surface reflectance (SR) data from 2017 to 2020, in which the Normalized Difference Vegetation Index (NDVI) was calculated for each pixel. The difference between the 90th and 10th percentiles (p90–p10) of NDVI values was used to identify areas with high seasonal variation. These patterns were then used to adjust the region boundaries, ensuring that areas with distinct spectral and phenological behavior were not grouped within the same classification region. The revised set of 38 classification regions remained unchanged from Collection 7.0 through Collection 10.0. In Collection 11.0, however, the updated Cerrado biome boundaries proposed by the *Instituto Brasileiro de Geografia e Estatística* (IBGE) required a revision of region 38 (São Paulo and Minas Gerais states). Its boundary was redesigned to incorporate the newly included portion of the biome, increasing the region's area by approximately 15,000 hectares. The remaining classification regions were maintained without changes. Overall, this regionalization framework continues to provide an effective basis for representing the spatial variability of Cerrado vegetation and improving the consistency of classification results across contrasting environmental conditions.

4.2. Feature space

The feature space used for land use and land cover classification in Collection 11.0 comprised a set of 228 predictor variables, designed to capture the spectral, temporal, and environmental complexity of the Cerrado biome. These variables encompassed both dynamic (annual and intra-annual) and static (non-annual) components. The predictors comprised annual spectral metrics, spatial-context variables, three-year temporal metrics, geographic coordinates, terrain attributes, and fire-history information. For each of these variables, a set of statistical reduction metrics was calculated within the April–September compositing period. Up to Collection 9.0, the minimum and maximum metrics corresponded to the absolute lowest and highest values available within the image set for each year. Although these metrics captured the full observed range, they were sensitive to residual clouds, atmospheric contamination, anomalous observations, etc. In Collection 10.0, the absolute minimum and maximum were replaced by the 5th and 95th percentiles, respectively, providing more robust estimates of the lower and upper limits of the annual spectral distribution. In Collection 11.0, the temporal variability metrics were further revised. Standard deviation was replaced by the annual and seasonal median absolute deviation, which is less sensitive to extreme observations and provides a more robust measure of dispersion around the median, improving the representation of seasonal variability while reducing the influence of anomalous observations. Median, dry-season median, wet-season median, amplitude, 5th percentile, and 95th percentile metrics were retained where applicable.

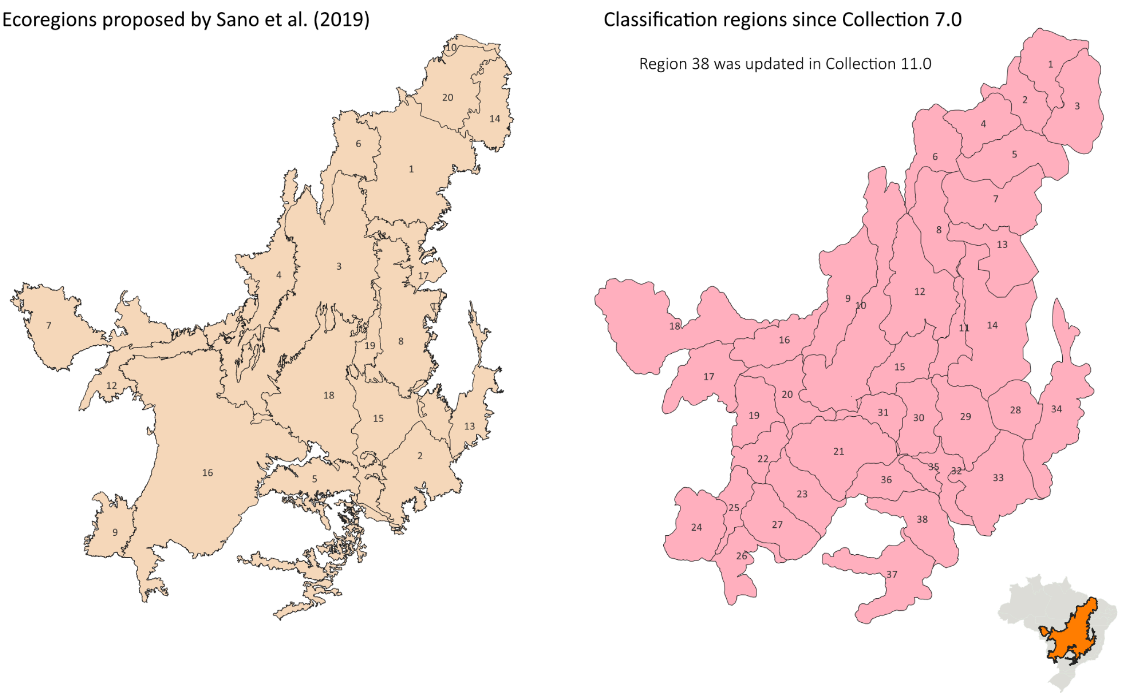


Figure 5. Cerrado biome classification regions, modified from Sano et al., 2019. Highlighted in orange is the location of the Cerrado biome in Brazilian territory.

The representation of local spatial structure was also revised in Collection 11.0. Up to Collection 10.0, image texture was characterized using the entropy metric implemented in Google Earth Engine, calculated from the median green band. In Collection 11.0, this entropy-based variable was replaced by a reduced set of neighborhood statistics designed to capture local structural context using spectrally and ecologically informative variables, helping the classifier distinguish classes with similar pixel-level responses but different spatial structures, such as homogeneous vegetation, fragmented cover, exposed surfaces, and transition zones. These metrics were calculated within a square kernel with a radius of three pixels, corresponding to a 7 × 7-pixel neighborhood in GCVI and NDFI bands. The neighborhood statistics represent the prevailing spectral conditions surrounding each pixel, describing the local spatial heterogeneity. The complete set of annual predictors and statistical metrics used in Collection 11.0 is presented in Table 3.

Table 3. Feature space used in the classification of the Cerrado biome in MapBiomas Collection 11.0. The column "Statistics" refers to the set of per-pixel statistical reducers applied to each variable within the annual temporal window (April–September): a) Amplitude – range of pixel values; b) Median – median value within the annual temporal window; c) Median_dry – seasonal median for dates below the first quartile of NDVI values (dry period); d) Median_wet – seasonal median for dates above the first quartile of NDVI values (wet period); e) Median Absolute Deviation (MAD) – median of the absolute deviations from the annual median; f) Minimum – 5th percentile of pixel values within the temporal window; and g) Maximum – 95th percentile of pixel values within the temporal window.

Type	Name	Formula / Description	Statistics	Reference
Landsat Band	Blue, Green, Red, NIR, SWIR1, SWIR2	Original reflectance bands	Median, Median_dry, Median_wet, Minimum (p5), Maximum (p95), MAD, Amplitude	USGS
Spectral Index	NDVI Normalized Difference Vegetation Index	(NIR - Red) / (NIR + Red)	Median, Median_dry, Median_wet, Minimum (p5), Maximum (p95), MAD, Amplitude, p25, p75	Rouse et al., 1974
	EVI2 Enhanced Vegetation Index 2	2.5 × (NIR - Red) / (NIR + 2.4 × Red + 1)	Median, Median_dry, Median_wet, Minimum (p5), Maximum (p95), MAD, Amplitude	Parente et al., 2018
	GCVI Green Chlorophyll Vegetation Index	(NIR / Green - 1)	Median, Median_dry, Median_wet, Minimum (p5), Maximum (p95), MAD, Amplitude	Burke et al., 2017
	PRI Photochemical Reflectance Index	(Blue - Green) / (Blue + Green)	Median, Median_dry, Median_wet, Minimum (p5), Maximum (p95), MAD, Amplitude	Gamon et al., 1992
	MNDWI Modified Normalized Difference Water Index	(Green − SWIR1) / (Green + SWIR1)	Median, Median_dry, Median_wet, Minimum (p5), Maximum (p95), MAD, Amplitude	Xu et al., 2006
	NBR Normalized Burn Ratio	(NIR - SWIR2) / (NIR + SWIR2)	Median, Median_dry, Median_wet, Minimum (p5), Maximum (p95), MAD, Amplitude	Key and Benson, 2006

Type	Name	Formula / Description	Statistics	Reference
	CAI Cellulose Absorption Index	SWIR2 / SWIR1	Median, Median_dry, Median_wet, Minimum (p5), Maximum (p95), MAD, Amplitude	Nagler et al., 2003
	MSAVI Modified Soil Adjusted Vegetation Index	0.5 × [2 × NIR + 1 − √(2 × NIR + 1)² − 8 × (NIR − Red)]	Median, Median_dry, Median_wet, Minimum (p5), Maximum (p95), MAD, Amplitude	Qi et al., 1994
	GRND Green Normalized Difference Vegetation Index	(Green − Red) / (Green + Red)	Median, Median_dry, Median_wet, Minimum (p5), Maximum (p95), MAD, Amplitude	Gitelson et al., 1996
	MSI Moisture Stress Index	(SWIR1 / NIR)	Median, Median_dry, Median_wet, Minimum (p5), Maximum (p95), MAD, Amplitude	Zarco-Tejada et al., 2003
	GARI Green Atmospherically Resistant Vegetation Index	(NIR − (Green − (Blue − Red))) / (NIR + (Green − (Blue − Red)))	Median, Median_dry, Median_wet, Minimum (p5), Maximum (p95), MAD, Amplitude	Gitelson et al., 2003
	GNDVI Green Normalized Difference Vegetation Index	(NIR − Green) / (NIR + Green)	Median, Median_dry, Median_wet, Minimum (p5), Maximum (p95), MAD, Amplitude	Gitelson et al., 1996
Vegetation Structure	Hall Height	(−0.039 × Red − 0.011 × NIR − 0.026 × SWIR1 + 4.13)	Median, MAD	Hall et al., 2006
	Hall Cover	(- Red × 0.017 - NIR × 0.007 - SWIR2 × 0.079 + 5.22)	Median, MAD	Hall et al., 2006
Fraction	GV Green Vegetation Fraction	Fractional abundance of green vegetation within the pixel	Median, Median_dry, Median_wet, Minimum (p5), Maximum (p95), MAD, Amplitude	Souza et al., 2005
	GVS Green Vegetation Shade Fraction	GV / (GV + NPV + Soil + Cloud)	Median, Median_dry, Median_wet, Minimum (p5), Maximum (p95), MAD, Amplitude	Housman et al., 2018
	FNS Forest/Non-Forest Index	100 *((GVS - Soil) / (GVS+ Soil)) + 100	Median, Median_dry, Median_wet, Minimum (p5), Maximum (p95), MAD, Amplitude	Souza et al., 2005
	NDFI Normalized Difference Fraction Index	(GVS - (NPV + Soil)) / (GVS + (NPV + Soil))	Median, Median_dry, Median_wet, Minimum (p5), Maximum (p95), MAD, Amplitude	Souza et al., 2005
	NPV Non-photosynthetic Vegetation Fraction	Fractional abundance of non-photosynthetic vegetation within the pixel	Median, Median_dry, Median_wet, Minimum (p5), Maximum (p95), MAD, Amplitude	Souza et al., 2005
	SEFI Savanna Ecosystem Fraction Index	(GV + NPV_S - Soil) / (GV + NPV_S + Soil)	Median, Median_dry, Median_wet, Minimum (p5), Maximum (p95), MAD, Amplitude	Alencar et al., 2020
	Shade Fraction	100 - (GV + NPV + Soil + Cloud)	Median, MAD	Housman et al., 2018
	Soil Fraction	Fractional abundance of soil within the pixel	Median, Median_dry, Median_wet, Minimum (p5), Maximum (p95), MAD, Amplitude	Souza et al., 2005

Type	Name	Formula / Description	Statistics	Reference
	WEFI wetland Ecosystem Fraction Index	((GV + NPV) − (Soil + Shade)) / ((GV + NPV) + (Soil + Shade))	Median, Median_dry, Median_wet, Minimum (p5), Maximum (p95), MAD, Amplitude	Rosa, 2020
Spatial context	GCVI Median; Dry-Season GCVI and Dry-Season NDFI	Local mean of the bands, calculated within a 7 × 7-pixel square neighborhood (3 pixels)	Neighborhood mean	This collection

In addition to these annual variables, we included since Collection 7.0 the fire history derived from MapBiomas Fire Collection 5 and a set of static predictors has been incorporated to enhance the environmental and spatial context to the classification model. This set includes HAND (Height Above Nearest Drainage) and spatial coordinates (latitude and longitude) to reduce the influence of geographically distant training samples with similar spectral characteristics. Collection 10.0 introduced additional geomorphometric variables derived from digital elevation models, including elevation, slope, aspect, curvature, roughness, and other terrain metrics. These variables support the discrimination of native vegetation formations in areas with complex relief and were maintained in Collection 11.0. The complete list of static predictors is provided in Table 4.

Table 4. Static (non-annual) variables used in the classification process of MapBiomas Collection 11.0 for the Cerrado biome. "Identity" in the statistics column indicates that the variable is used directly, without temporal reduction.

Name	Formula / Description	Statistics	Reference
Latitude	Extracted from pixel latitude (ee.Image.pixelLonLat())	Identity	Geolocation function
Cosine of Longitude	cos(longitude × π / 180)	Identity	Geolocation function
Sine of Longitude	sin(longitude × π / 180)	Identity	Geolocation function
Time Since the Last Fire	Current year - Year of the last fire	Identity	Alencar et al., 2022
Height Above the Nearest Drainage	HAND Global 30m	Identity	Donchyts et al., 2016
Elevation (DEM)	FABDEM elevation (in meters)	Identity	Hawker et al., 2022
Aspect	Terrain aspect	Identity	Geomorpho 90m Amatulli et al., 2020
Slope	Terrain slope	Identity	Geomorpho 90m Amatulli et al., 2020

Name	Formula / Description	Statistics	Reference
DXX East–West Second Directional Derivative	Second-order partial derivative of elevation in the East–West direction, expressed as $\partial^2 z/\partial x^2$.	Identity	Geomorpho 90m Amatulli et al., 2020
Convergence Index	Terrain convergence	Identity	Geomorpho 90m Amatulli et al., 2020
Roughness	Surface roughness index	Identity	Geomorpho 90m Amatulli et al., 2020
Eastness	Derived from aspect to represent east-facing slopes	Identity	Geomorpho 90m Amatulli et al., 2020
Northness	Derived from aspect to represent north-facing slopes	Identity	Geomorpho 90m Amatulli et al., 2020
CTI Compound Topographic Index	Wetness index combining slope and upstream area	Identity	Geomorpho 90m Amatulli et al., 2020

4.3. Training mask, stratified sampling, and classification approach

Up to Collection 10.0, the training masks used for the Cerrado classification were primarily derived from stable pixels identified in the immediately preceding MapBiomas collection. Stable pixels were defined as those that remained in the same land use and land cover class throughout the available historical series and were subsequently reclassified according to the Cerrado classification legend. Although this approach provided high temporal consistency, requiring stability across nearly four decades restricted the training samples to areas characterized by persistent natural vegetation or long-established land use. Thus, the use of the full time-series to generate stable samples was shown to be insufficient in regions where most LULC changes occurred in the early or late time-series. In Collection 11.0, the training-mask construction was revised by dividing the historical series into four overlapping temporal windows: 1985–1996, 1994–2005, 2003–2014, and 2012–2024. Stable pixels were identified independently within each window rather than across the complete time series. The overlap between consecutive windows was introduced to ensure a gradual transition between training periods and to reduce abrupt classification changes at the boundaries between temporal windows. This strategy increased the temporal representativeness of the training datasets.

For each temporal window, stable pixels were identified using MapBiomas Collections 8.0, 9.0, and 10.1, according to the temporal coverage available in each collection. Native-vegetation samples were selected from pixels that remained in the same class in Collection 10.1 throughout the corresponding window. A stricter agreement criterion was applied to anthropogenic classes. For these classes, candidate pixels were retained only when the stable class identified in Collection 10.1 agreed with the stable classifications derived from Collections 9.0 and 8.0, whenever those collections covered the period under evaluation. This cross-collection agreement criterion was empirically adopted based on previous evaluations to reduce the propagation of historical classification errors in dynamic land-use classes. Classes with high uncertainty or transitional behavior, including other non-vegetated areas (25) and mosaic of uses (21), were excluded from the stable-sample selection procedure. Visual assessments indicated a better representation of anthropogenic areas, particularly during the initial years.

The stable areas identified for each temporal window were subsequently refined using a set of ancillary reference datasets to improve the quality and reliability of the stable training masks. The refinement incorporated deforestation alerts from PRODES (2000-2025) and MapBiomas Alerta (2019-2025), FABDEM terrain information, and regional LULC reference maps produced by state institutions for São Paulo, Tocantins, Goiás, Mato Grosso and Distrito Federal, as well as zoning data from ICMBio (*Instituto Chico Mendes de Conservação da Biodiversidade*). Ancillary reference datasets have been incorporated into the Cerrado training-sample workflow across multiple MapBiomas collections, as summarized in Table 5. This reference database is updated for each collection, with new datasets incorporated when they become available. In some cases, others may be discontinued or replaced according to Cerrado LULC collections updates.

Table 5. Ancillary reference datasets used to refine stable training masks for the Cerrado biome across MapBiomas collections.

Reference dataset	Source	Main information	Use history
PROBIO	INPE	Cerrado native vegetation mapping	Collections 5.0 to 8.0; discontinued from Collection 9.0
PRODES Cerrado	INPE	Annual Cerrado native vegetation loss	Since collection 5.0; updated for each collection
SAD Cerrado	IPAM	Cerrado deforestation alerts	Collections 8.0 and 9.0; discontinued from Collection 10.0
MapBiomas Alerta	MapBiomas	Validated deforestation alerts	Since collection 8.0; updated for each collection
Global Canopy Height	Lang et al. (2023)	Canopy-height information	Since collection 7.0
São Paulo State Forest Inventory Mapping	SEMIL/SP	High-resolution vegetation and land-cover mapping for São Paulo	Since collection 6.0
Tocantins CAR Thematic Digital Vector Database	SEMARH/TO	Detailed thematic reference data for land-cover interpretation in Tocantins	Since collection 7.0

Reference dataset	Source	Main information	Use history
Distrito Federal Land Use Land Cover Map	SEMA/DF	Regional vegetation and anthropogenic land-cover	Since collection 9.0
Remnants of Murundus in Goiás	SEMAD/GO	Reference mapping of *campos de murundus* in Goiás	Since collection 9.0
Southeastern Tocantins Cartographic Database – Wetlands	SEPLAN/TO	Regional wetlands mapping in southeastern Tocantins	Since collection 10.0
Southeastern Tocantins Cartographic Database – Land Use and Land Cover	SEPLAN/TO	Regional land use and land cover mapping for southeastern Tocantins	Since collection 10.0
Tocantins Land Use and Land Cover Map	SEPLAN/TO	Regional land use and land cover mapping for Tocantins	Introduced in Collection 11.0
Chapada das Mesas National Park Zoning	MMA / ICMBio	Vegetation mapping for Chapada das Mesas National Park	Introduced in Collection 11.0
Mapping of Veredas in Mato Grosso	SEMA/MT	Regional mapping of *veredas*	Introduced in Collection 11.0
FAL–UnB Land Use and Land Cover Map	Leal and Cartagenes (2021)	Local land use and land cover mapping for *Fazenda Água Limpa*, Federal District	Introduced in Collection 11.0
GPW Annual Short Vegetation Height	Global Pasture Watch	Annual short-vegetation height	Introduced in Collection 11.0

Following this reference-based refinement, additional spatial and structural criteria were applied to further improve the consistency of the training masks. Only homogeneous patches equal to or larger than approximately 1 ha, corresponding to approximately 11 Landsat pixels, were retained in the final training masks. Tree canopy height-based filters were also applied to remove candidate samples with values inconsistent with the expected structural characteristics of their respective classes. The procedure retained the general logic introduced in Collection 7.0, for which an increase of 0.9 percentage points in overall accuracy had previously been documented. In Collection 11.0, the procedure was expanded by incorporating short-vegetation height data from Global Pasture Watch (GPW), complementing the tree canopy-height information derived from GEDI (Lang et al., 2022). The additional information supported the filtering of samples associated with grassland formation, pasture, and agriculture. Candidate samples were excluded when they met one or more of the following class-specific conditions:

- Forest Formation (3): tree canopy height < 4 meters

- Savanna Formation (4): tree canopy height < 2 meters or > 8 meters
- Wetland (11): tree canopy height > 15 meters
- Grassland Formation (12): short-vegetation height > 3 meters
- Pasture (15): short-vegetation height > 4 meters
- Agriculture (18): short-vegetation height > 3 meters
- Non-Vegetated Area (25): tree canopy height > 0 meters
- Water (33): tree canopy height > 0 meters

A stratified random sampling procedure was used to represent the proportional distribution of the LULC classes within each classification region and temporal window. For each time-window, a predefined year located at the midpoint of the period was used as the reference year (1990, 1999, 2008, 2018) to obtain per-class proportions. The area occupied by each class in that year was calculated separately for each classification region and used to determine the proportional allocation of training samples among classes.

A total of 4,800 automatically generated samples was allocated to each classification region and year. This sample size was empirically defined to provide sufficient representation of the spectral and thematic variability within each region while maintaining computational feasibility and preventing excessive memory use during processing in GEE. The proportional allocation was subject to a minimum of 480 samples per class to ensure the representation of less frequent classes. Water samples were restricted to pixels with HAND equal to zero and limited to a maximum of 240 samples per region to prevent overrepresentation. The samples were subsequently subjected to visual inspection using Landsat imagery. This review was used to identify and remove samples located in heterogeneous areas, samples with inconsistent spectral responses, and samples that did not adequately represent their assigned land use and land cover class. Complementary samples obtained through visual interpretation were incorporated when the stratified sampling did not sufficiently represent LULC classes, spectral conditions, or regional environmental contexts.

The classification was performed independently for each classification region and year using the Random Forest algorithm implemented in Google Earth Engine (GEE) through the `ee.Classifier.smileRandomForest` function. Based on empirical testing observed in previous MapBiomas collections, the number of decision trees was fixed at 300 for all regions, while the number of variables considered at each split was set to the square root of the total number of predictor bands. The classifier was operated in multiprobability mode, producing a probability distribution across all target classes for each pixel. The final class assignment was determined by selecting the class with the highest predicted probability, an approach that increases robustness to spectral confusion and better represents classification uncertainty, particularly in heterogeneous landscapes.

5. GENERAL MAP POST-CLASSIFICATION

Given the pixel-based classification methodology and the annual processing of the time series (1985 – 2025), a structured post-classification filtering framework was applied. The main objective of this stage was to correct classification artifacts and reduce spurious transitions commonly associated with per-pixel classifiers operating over long temporal windows. The post-processing framework included a sequence of filters that was designed to address specific classification limitations and collectively contributed to the overall quality and reliability of Collection 11.0.

5.1. Temporal Gap-Fill

The gap-filling filter was designed to address missing data, commonly caused by cloud and shadow contamination or gaps in image availability, by propagating valid classifications across the temporal dimension. The implemented method applied a bidirectional temporal search. First, the forward pass processed the time series chronologically, from the initial year (t0) to the final year (tn). During this pass, an unclassified pixel was filled using the most recent valid classification available from an earlier year. Subsequently, the backward pass processed the series in reverse chronological order, from (tn) to (t0). This second pass filled any remaining unclassified pixels using the nearest valid classification available from a later year. This two-step process ensured that the final classified time series had minimal missing values, resulting in more complete and temporally coherent maps. Persistent gaps remained only in exceptional cases where a pixel was unclassified throughout the entire 1985–2025 period.

5.2. First spatial filter

The first spatial filter was applied to reduce isolated pixels and small classification patches before the thematic and temporal filters. At this step, pasture (15) and agriculture (18) were reclassified as mosaic of uses (21), consequently, these anthropogenic land-use types were treated as a single class throughout the subsequent post-classification procedures. For each annual map, connected components were calculated, and patches containing approximately 1 hectare or less (equivalent to 11 Landsat pixels) were replaced by the modal class within a 9 × 9-pixel neighborhood. Eight-neighbor connectivity was used for most classes, whereas four-neighbor connectivity was adopted for mosaic of uses to avoid connecting anthropogenic patches only through diagonal contact. forest formation (3), wetland (11), and water (33) were protected from this correction to preserve small but ecologically relevant features.

5.3. Topographic

The topographic filter was designed to correct classifications that were inconsistent with terrain conditions, particularly errors associated with relief shadows and steep slopes. Slope was calculated from the FABDEM Elevation Model and expressed as a percentage. wetland (11) pixels occurring on terrain slopes equal to or greater than 12% were reclassified as forest formation (3). Water (33) pixels on slopes equal to or greater than 15% were also reclassified as forest formation (3) when the pixel had been mapped as native vegetation in at least two years of the historical series, indicating that the apparent water occurrence was probably associated with terrain shadow. Water (33) occurrences on slopes equal to or greater than 50% were corrected independently of their historical vegetation frequency. Mosaic of uses pixels located on slopes equal to or greater than 50% were replaced by the modal class of the surrounding neighborhood.

5.4. Spatial–temporal transition

The spatial–temporal transition filter was applied to remove short and spatially restricted oscillations among broad land-cover groups. The classes were grouped into native vegetation (forest formation, savanna formation, wetland, and grassland formation), anthropogenic cover (mosaic of uses and other non-vegetated areas) and other classes (water and no data). For each intermediate year, the filter evaluated a three-year window and identified A – B – A trajectories in which a pixel temporarily changed thematic group and returned to its previous group in the following year. When the temporary transition formed a connected patch of six pixels or fewer, approximately 0.5 hectares at 30-m resolution, the intermediate classification was replaced by the detailed class mapped in the preceding year, converting into an A – A – A trajectories. This procedure reduces edge effects and small transition artifacts without suppressing spatially consolidated land-cover changes. Up to Collection 10.0, this post-classification procedure used a different approach and was referred to as the "Incidence Filter."

5.5. Herbaceous Sandbank Vegetation

Herbaceous sandbank vegetation (50) is a beta class introduced for the Cerrado biome in Collection 10.0 and maintained in Collection 11.0 through a dedicated post-classification procedure. Reference polygons were manually collected to represent areas of herbaceous sandbank vegetation and non-sandbank vegetation. Random points extracted from these polygons were used to train a binary Gradient Tree Booster (GTB) classifier configured with 50 trees. Predictor variables were selected based on expert knowledge of the spectral and structural characteristics associated with herbaceous sandbank vegetation. The selected variables included median Red and SWIR1 reflectance; median and wet-season GCVI index; wet- and dry-season EVI2; NDWI amplitude; wet-season WEFI; and dry-season SEFI. In addition, HAND, canopy height, and FABDEM were included to represent hydrological, vegetation structure, and topographic conditions, respectively. The combination of the selected predictor variables and the GTB classifier resulted in a spatial mask representing areas with potential occurrence of herbaceous sandbank vegetation. The purpose was not to generate annual maps of herbaceous sandbank, but to produce a static spatial mask representing areas with potential occurrence of this vegetation type. The GTB prediction was restricted to the coastal sandy-deposit areas mapped by the *Serviço Geológico Brasileiro* (CPRM/SGB), ensuring that the potential occurrence mask was limited to environments considered compatible with this class. This restricted mask was then used to remap the annual general LULC classification from 1985 to 2025. For each year, pixels located within the potential occurrence mask and originally classified as savanna formation (4), wetland (11), or grassland formation (12) were reassigned to herbaceous sandbank vegetation (50). Pixels outside the mask, as well as pixels belonging to other LULC classes, retained their original classification. Therefore, although the potential occurrence mask remained spatially constant throughout the time series, the annual extent of herbaceous sandbank varied according to the distribution of eligible classes in each annual general map.

5.6. Trajectory

The trajectory filter corrected specific temporal patterns associated primarily with grassland formation (12), which may occur as an unstable intermediate class between native vegetation and anthropogenic land use. One-year grassland occurrences in A) native vegetation → grassland formation → mosaic of uses, B) mosaic of uses → grassland formation → native vegetation, and C) mosaic of uses → grassland formation → mosaic of uses trajectories were replaced by the class mapped in the subsequent year. One- or two-year occurrences of grassland immediately preceding a consolidated mosaic of uses period were also reclassified as mosaic of uses. The same rule was applied to isolated grassland occurrences appearing at the end of time series otherwise dominated by mosaic of uses. In addition, temporary occurrences of other non-vegetated areas (25) enclosed by native classes were replaced by the subsequent valid native classification, reducing artificial gaps within otherwise continuous native-vegetation trajectories.

5.7. Frequency

The frequency filter was designed to reduce short-term temporal oscillations among native vegetation classes while preserving persistent changes in vegetation type.

First, pixels classified as native vegetation in at least 95% of the temporal series (2017 - 2025) were identified as temporally stable native areas. Within these pixels, the historical frequency of each native class was calculated, and a reference class was assigned using class-specific thresholds: savanna formation (>40%), herbaceous sandbank vegetation (≥60%), grassland formation (>50%), wetland (≥40%), and forest formation (≥60%). When more than one threshold was satisfied, the classes were evaluated hierarchically, with forest formation having the highest priority, followed by wetland, grassland formation, herbaceous sandbank vegetation, and savanna formation. Before applying any replacement, the annual classification was evaluated for temporal persistence. A native class was considered persistent when it occurred for at least three consecutive years, considering backward, centered, or forward three-year windows. Persistent occurrences were retained, whereas non-persistent native classifications were replaced by the reference class assigned from the long-term frequency analysis. The filter therefore targets short-lived transitions among native vegetation classes without modifying persistent temporal patterns or pixels classified as non-native land cover.

5.8. Temporal

The general temporal filter is a sequential post-classification procedure designed to reduce spurious short-term transitions and reinforce the temporal consistency of LULC trajectories throughout the 1985–2025 series. The filter compares the annual classification of each pixel across multi-year temporal windows to remove isolated or short-lived changes, preserve persistent anthropogenic conversions, and stabilize the first and last years of the series. The procedure consists of the following sequential operations:

- 5-year and 4-year window: The filter evaluates the annual classification using moving windows of five and four years. The five-year window corrects trajectories in which a pixel is assigned to a given class in the initial year, changes during the three intermediate years, and returns to the initial class in the final year (C – X – X – X – C). Similarly, the four-year window corrects temporary two-year interruptions bounded by the same class (C – X – X – C). In both cases, the intermediate classifications are replaced by the class occurring at the beginning and end of the temporal window. The rules are applied in a processing order: river, lake, and ocean (33), other non-vegetated areas (25), mosaic of uses (21), grassland formation (12), wetland (11), herbaceous sandbank vegetation (50), savanna formation (4), and forest formation (3). The classes applied later take precedence in cases of overlapping corrections.
- 3-year window: This step identifies isolated one-year changes bounded by the same class (C – X – C). The intermediate classification is replaced by the class occurring in the preceding and subsequent years, following the same class order

adopted for the five-year and four-year windows. From 2023 onward, native vegetation classes are prevented from replacing pixels classified as mosaic of uses (21), thereby preserving the recent land cover conversions.

- Correction of unstable trajectories in 2024–2025: The filter corrects two-year instabilities at the end of the time series. When a pixel presents the same consolidated class in 2022 and 2023 but is classified as grassland formation (12) or other non-vegetated areas (25) in both 2024 and 2025, the classifications for these two final years are replaced by the stable class observed in 2023.
- Correction of the last year (2025): Two specific rules are applied to improve the reliability of the final year. First, when the classifications in 2023 and 2024 are identical but 2025 presents a different class, the 2025 value is replaced by the 2024 classification. Mosaic of uses (21) is excluded from this correction when it occurs in 2025, because it may represent a recent anthropogenic conversion. Second, when a pixel is classified as mosaic of uses in 2024 and this classification is supported by its occurrence in either 2022 or 2023, but is not maintained in 2025, the final-year classification is corrected to mosaic of uses. This rule avoids the mapping of unconfirmed regeneration after a consolidated anthropogenic use.
- Stabilization of the first year (1985): The filter identifies trajectories in which the 1985 classification differs from a native vegetation class consistently mapped in both 1986 and 1987 (X – A – A). In these cases, the 1985 value is replaced by the class observed in the two subsequent years. This correction is applied sequentially to grassland formation (12), wetland (11), herbaceous sandbank vegetation (50), savanna formation (4), and forest formation (3), reducing isolated artifacts at the beginning of the Landsat time series.

5.9. False-regrowth and long-term trajectory

The false-regrowth filter complements the general temporal filter by correcting class-specific and long-term trajectory inconsistencies that are not adequately addressed by the temporal moving windows. Its main objective is to suppress implausible vegetation-regrowth events, remove short interruptions within consolidated LULC trajectories, and improve the temporal consistency throughout the 1985–2025 series. The procedure comprises the following sequential operations:

- Harmonization of wetland–grassland trajectories: A stable classification as wetland throughout the entire 1985–2024 time series and ending as grassland formation (12) in 2025, all grassland occurrences within the series were reclassified as wetland. This rule reduced inconsistent transitions between these spectrally related classes and produced a more coherent long-term trajectory.

- Stabilization of long grassland trajectories: Pixels classified as mosaic of uses (21) or other non-vegetated area (25) in 1985 were reclassified as grassland formation throughout the complete series when they presented at least 13 annual occurrences of grassland formation (⅓ of the time series) and were classified as grassland formation in the 2024 reference year. The occurrences were counted across the time series and were not required to be consecutive.
- Removal of temporary anthropogenic occurrences within native vegetation: Blocks of one to three consecutive years classified as mosaic of uses (21) and enclosed by native vegetation were interpreted as temporary commission errors. This rule considered savanna formation (4), wetland (11), and grassland formation (12) as the surrounding native classes. The mosaic of uses block was replaced by the native class observed immediately after the temporary anthropogenic interval.
- Suppression of short false-regrowth intervals: Native-vegetation intervals lasting from one to three consecutive years were reclassified as mosaic of uses (21) when they occurred between two consolidated anthropogenic periods, each consisting of at least four consecutive years of mosaic of uses (21). Forest formation (3), savanna formation (4), wetland (11), and grassland formation (12) were considered native-vegetation classes in this rule.

5.10. Silviculture

The silviculture filter was developed to reduce confusion between mature forest plantations and native forest formation (3), as closed-canopy eucalyptus and pine stands may present spectral characteristics like those of native forests. The procedure evaluated the long-term land-cover trajectory of each pixel, using the recurrence and temporal dominance of mosaic of uses (21) as evidence of persistent anthropogenic occupation. Pixels classified as forest formation were identified as potential silviculture when their trajectories met at least one of the following conditions: A) at least 12 occurrences of mosaic of uses within the preceding 15 years; B) an initial sequence of eight consecutive years classified as mosaic of uses followed by forest formation; C) a predominance of mosaic of uses through 2015 followed by a persistent forest formation classification; or D) at least 15 previous occurrences of mosaic of uses over the time series. To increase the reliability of the correction, these trajectory conditions were considered only when mosaic of uses was also observed in 2024 or 2025. Forest formation pixels meeting these criteria were therefore interpreted as likely forest plantations and reclassified as mosaic of uses. The procedure was initially developed in Collection 9.0 as a “False Regrowth Filter”, with the purpose of correcting spurious transitions from long-term anthropogenic land use to native forest formation (3). In Collection 11.0, the approach was revised and implemented as another post-classification step named the “Silviculture Filter.”

5.11. Spatial-shape

The spatial-shape filter was introduced in Collection 11.0 to remove small, fragmented, or geometrically irregular patches classified as mosaic of uses (21). The procedure was applied independently to each annual map and used an object-based approach in which spatially connected class-21 pixels were grouped into individual patches using eight-neighbor connectivity. For each patch, the filter calculated its area, the proportion of its rectangular bounding box effectively occupied by classified pixels, and the presence of an internal core composed of a complete 3 × 3-pixel neighborhood. Patches with an area equal to or smaller than 3 hectares were considered for correction when they met at least one of the following conditions: a bounding-box fill ratio below 0.65, indicating an elongated or irregular shape; absence of a 3 × 3-pixel core, indicating a thin or fragmented structure; or a total size of three pixels or fewer, representing isolated speckle noise. Pixels belonging to the selected patches were replaced by the modal LULC class within a circular neighborhood of 150 m. This procedure reduces small anthropogenic artifacts and improves the spatial coherence of the annual maps without modifying larger or more spatially consolidated patches of mosaic of uses.

5.12. Second spatial filter

The second spatial filter was applied as the final post-classification step to remove residual spatial artifacts remaining after the temporal, thematic, and trajectory-based corrections. The procedure followed the same general logic adopted in the first spatial filter (section 5.2) and was applied independently to each annual map. Spatially connected pixels belonging to the same class were grouped into patches, and patches containing approximately 1 hectare or less, corresponding to 11 Landsat pixels, were identified as potential classification noise. Pixels within these small patches were replaced by the modal LULC class calculated within a 9 × 9-pixel neighborhood. Forest formation (3), wetland (11), and river, lake, and ocean (33) were excluded from this correction to preserve small but ecologically relevant features.

6. ROCKY OUTCROP CLASSIFICATION

The classification of rocky outcrops in the Cerrado biome has undergone successive conceptual and methodological refinements across MapBiomas Cerrado collections. In Collection 7.0, a beta version of the rocky outcrop class was introduced using a stepwise classification approach. This initial procedure involved a visual interpretation sample collection, followed by a two-stage classification workflow, a preliminary map was generated and subsequently used as a reference for training a second classification model. In Collection 8.0, the same general approach was maintained, while the spatial extent of the classification was expanded to improve the representation of rocky environments throughout the biome. The principal improvement in Collection 8.0 was therefore related to spatial coverage.

A major methodological advancement occurred in Collection 9.0, when the entire classification flow was redesigned. The two-stage procedure was replaced by a direct annual classification using the Random Forest algorithm, and the class was mapped across the entire Cerrado biome. This redesign simplified the classification workflow, reduced its dependence on a preliminary map, and improved the spatial consistency of the results among regions and years. The most substantial change came with Collection 10.0, which introduced both conceptual and methodological refinements. Whereas Collections 7.0 to 9.0 included both exposed rock and associated rupestrian vegetation, Collection 10.0 restricted the class to naturally exposed rocky surfaces without soil or vegetation cover. The objective of this revision was to produce a more restrictive and geologically explicit representation of exposed rocky surfaces. The feature space was also expanded through the inclusion of additional geomorphometric variables and a spectral index specifically intended to improve the discrimination of exposed rocky substrates.

In Collection 11.0, the broader conceptual definition adopted in Collections 7.0 to 9.0 was restored. The rocky outcrop class once again encompasses exposed rocky surfaces and associated rupestrian vegetation, specifically *Campo Rupestre* and *Cerrado Rupestre* (Ribeiro and Walter, 2008). This conceptual revision recognizes the close spatial and ecological association between exposed substrates and rupestrian vegetation in many Cerrado landscapes. This class typically includes stable geological formations with clear indicators of sedimentary, igneous, or metamorphic origins (Figure 6). The classification continued to be produced independently from the general Cerrado LULC map, maintaining the strategy adopted since Collection 7.0. The resulting map is subsequently incorporated into the general Cerrado LULC product during the internal integration stage. The detailed classification workflow is presented in Figure 7, and the subsequent sections describe the methodological steps in detail.

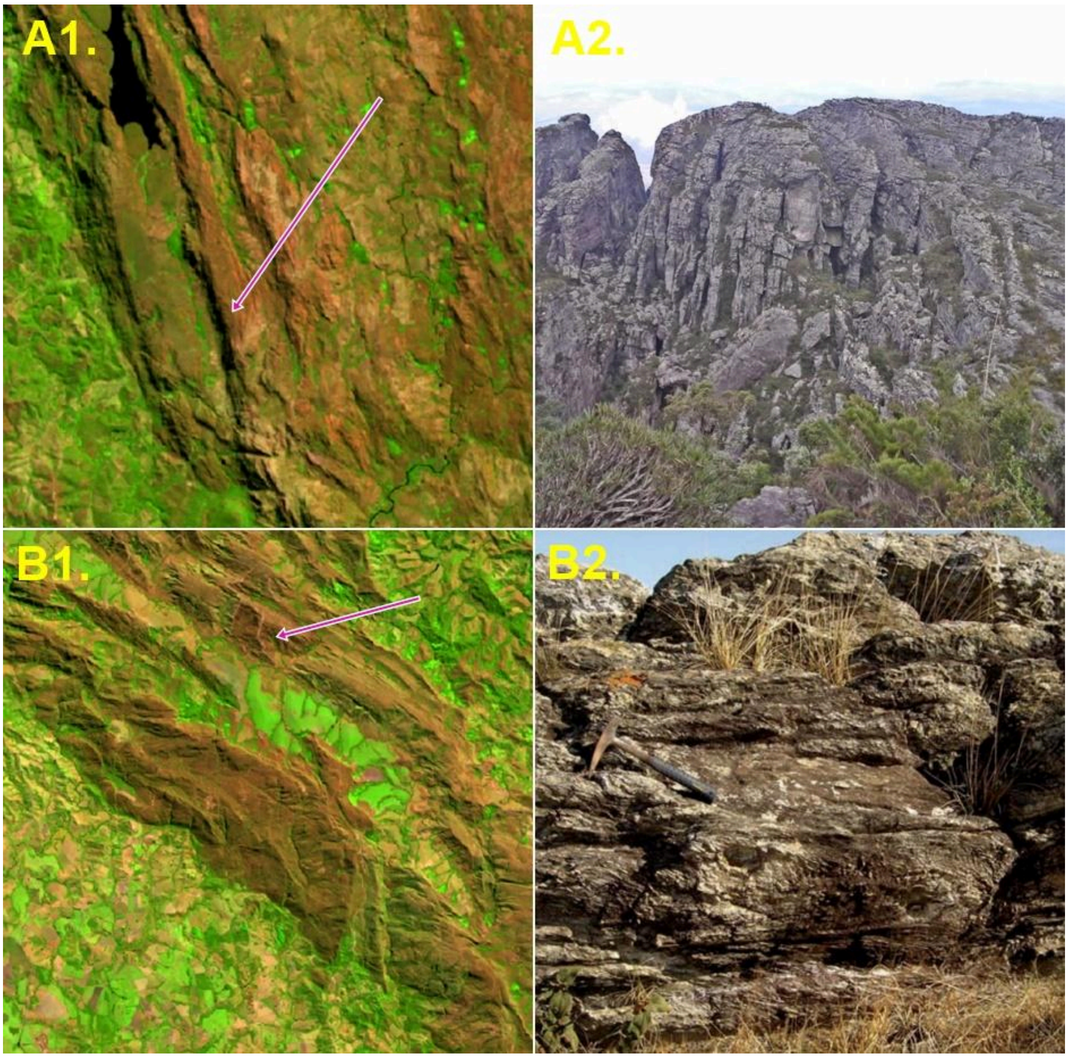


Figure 6. Example of landscapes mapped as rocky outcrop in Collection 11.0. A) “Serra do Espinhaço”: A1) Landsat false-color composition (SWIR1-NIR-Red) in 2021. A2) Field photograph (credits to TMbux). B) “Serra da Canastra”: B1) Landsat false-color composition (SWIR1-NIR-Red) in 2021.; B2) Field photograph (credits to Mario L.S.C Chaves). The red arrow in Landsat images indicates the approximate location of the field photograph.

6.1. Feature space

The feature space developed for Collection 11.0 rocky outcrop classification comprised 144 predictor bands representing spectral response, seasonal variability, interannual behavior, vegetation structure, terrain conditions, geographic position, and local spatial context. The predictor set was designed to represent the characteristics associated with exposed rock and rupestrian vegetation and to distinguish from native vegetation, agricultural areas, and other surfaces with similar spectral responses. Annual spectral information was derived from Landsat imagery using the same temporal compositing window adopted for the General Map feature space. Each spectral band was summarized using the median, dry-season median, wet-season median, amplitude, median absolute deviation, minimum, and maximum. The minimum and maximum metrics corresponded to the 5th and 95th percentiles, respectively, reducing the influence of residual clouds, shadows, atmospheric contamination, and other extreme observations. The median absolute deviation provided a robust measure of temporal variability around the annual median. The complete set of annual predictors and statistical metrics used in rocky outcrop classification is presented in Table 6.

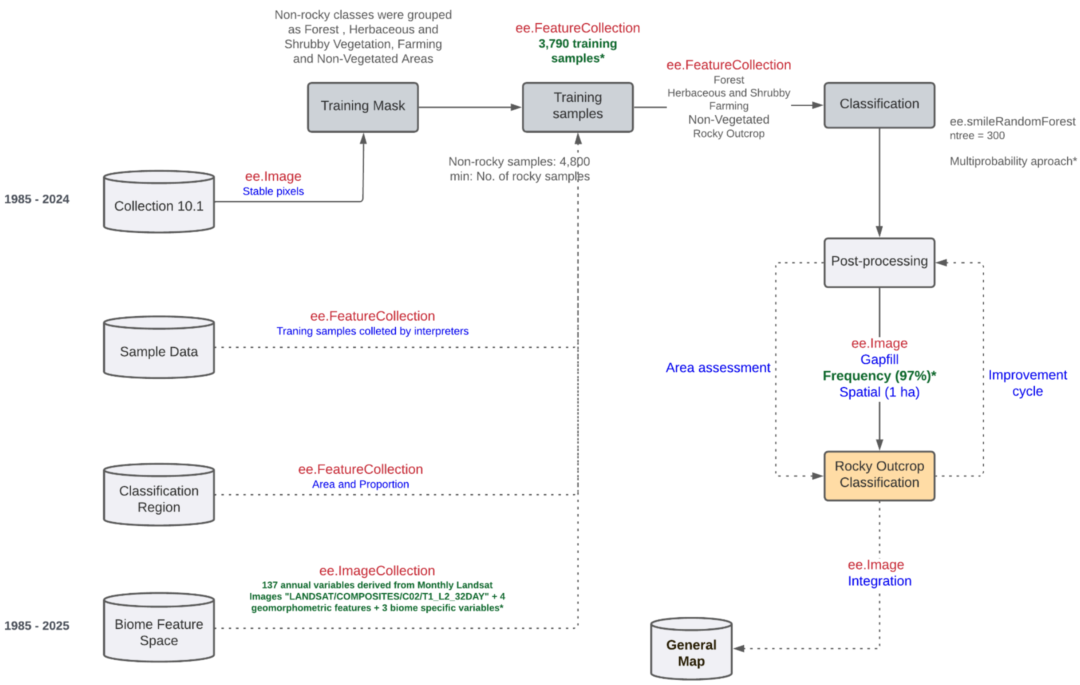


Figure 7. Each gray geometry (cylinders for databases and rectangles for processes) represents a key step in the classification schema, with the respective name inside. The gray text near databases and processes offers a description of the step, while the green text highlights the main innovations in Collection 11.0. 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 concise description of the asset content.

Two metrics calculated from three-year temporal windows were incorporated to represent short-term interannual behavior. The NDVI and BSI amplitude over the period represented the persistent rocky and rupestrian environments from temporary exposed-soil conditions, agricultural fallow, recent disturbances, and other land-cover states. Local spatial-context metrics were calculated from the annual median BSI, NDVI, and SWIR1 bands. For each of these variables, the neighborhood mean and neighborhood standard deviation were included as predictors. The neighborhood mean describes the prevailing spectral condition surrounding each pixel, whereas the neighborhood standard deviation represents local spatial heterogeneity. These metrics support the differentiation of spatially coherent rocky and rupestrian complexes from isolated bare pixels, fragmented agricultural surfaces, and spectrally heterogeneous anthropogenic areas.

The terrain component of the feature space was simplified in Collection 11.0. The extensive set of geomorphometric variables used in the Collection 10.0 was removed, reducing the dependence of the model on generalized terrain derivatives. Terrain slope and surface ruggedness were calculated directly from FABDEM because of their relevance for characterizing steep and irregular terrain associated with many rocky and rupestrian environments. Height Above Nearest Drainage (HAND) and the Topographic Position Index (TPI) were also included to represent relative drainage position and local terrain position.

Table 6. Feature space used in the rocky outcrop classification in Cerrado biome in Collection 11.0. The column "Statistic" refers to the set of per-pixel statistical reducers applied to each variable within the annual temporal window (April–September): a) Amplitude – range of pixel values; b) Median – median value within the annual temporal window; c) Median_dry – seasonal median for dates below the first quartile of NDVI values (dry period); d) Median_wet – seasonal median for dates above the first quartile of NDVI values (wet period); e) Median Absolute Deviation (MAD) – median of the absolute deviations from the annual median; f) Minimum – 5th percentile of pixel values within the temporal window; and g) Maximum – 95th percentile of pixel values within the temporal window; h) Identity – the variable is used without temporal reduction.

Type	Name	Formula	Statistics	Reference
Spectral Index	BSI Bare Soil Index	((SWIR1 + Red) − (NIR + Blue)) / ((SWIR1 + Red) + (NIR + Blue))	Median, Median_dry, Median_wet, Minimum (p5), Maximum (p95), MAD, Amplitude	Rikimaru et al., 2002
	EVI2 Enhanced Vegetation Index 2	2.5 × (NIR - Red) / (NIR + 2.4 × Red + 1)	Median, Median_dry, Median_wet, Minimum (p5), Maximum (p95), MAD, Amplitude	Parente et al., 2018
	MSI Moisture Stress Index	(SWIR1 / NIR)	Median, Median_dry, Median_wet, Minimum (p5), Maximum (p95), MAD, Amplitude	Zarco-Tejada et al., 2003
	MNDWI Modified Normalized Difference Water Index	(Green − SWIR1) / (Green + SWIR1)	Median, Median_dry, Median_wet, Minimum (p5), Maximum (p95), MAD, Amplitude	Xu et al., 2006

Type	Name	Formula	Statistics	Reference
	NBR Normalized Burn Ratio	(NIR − SWIR2) / (NIR + SWIR2)	Median, Median_dry, Median_wet, Minimum (p5), Maximum (p95), MAD, Amplitude	Key and Benson, 2006
	NDVI Normalized Difference Vegetation Index	(NIR - Red) / (NIR + Red)	Median, Median_dry, Median_wet, Minimum (p5), Maximum (p95), MAD, Amplitude, p25, p75	Rouse et al., 1974
	NDRI Normalized Difference Rock Index	(SWIR2 − NIR) / (SWIR2 + NIR)	Median, Median_dry, Median_wet, Minimum (p5), Maximum (p95), MAD, Amplitude	Huang and Cai, 2007
	TGSI Topsoil Grain Size Index	(Red − Blue) / (Red + Blue + Green)	Median, Median_dry, Median_wet, Minimum (p5), Maximum (p95), MAD, Amplitude	Xiao et al., 2006
Temporal Data	NDVI amplitude (3-year)	Difference between the maximum and minimum annual median values within the three-year window from t−2 to t	Range	Celebrezze et al., 2025
	BSI amplitude (3-year window)		Range	This study
Vegetation Structure	Hall Height	(−0.039 × Red − 0.011 × NIR − 0.026 × SWIR1 + 4.13)	Median, MAD	Hall et al., 2006
	Hall Cover	(- Red × 0.017 - NIR × 0.007 - SWIR2 × 0.079 + 5.22)	Median, MAD	Hall et al., 2006
Spatial context	BSI Median; NDVI Median and SWIR 1 Median	Local mean of the bands, calculated within a 7 × 7-pixel square neighborhood (3 pixels)	Neighborhood mean and standard deviation	This collection
Geographic position	Latitude	Extracted from pixel latitude (ee.Image.pixelLonLat())	Identity	Geolocation function
	Cosine of Longitude	cos(ee.Image.pixelLonLat() .select(['longitude']))	Identity	Geolocation function
	Sine of Longitude	sen(ee.Image.pixelLonLat() .select(['longitude']))	Identity	Geolocation function
Terrain	Height Above Nearest Drainage (HAND)	HAND Global 30m	Identity	Donchyts et al., 2016
	Slope	Terrain slope	Identity	FABDEM; calculated in this study
	Ruggedness	Surface roughness index	Identity	FABDEM; calculated in this study

Type	Name	Formula	Statistics	Reference
	TPI Topographic Position Index	Elevation − Mean (Elevation neighborhood)	Identity	Geomorpho 90m Amatulli et al., 2020

6.2. Training samples, classification algorithm, and parameters

The training dataset was constructed using two distinct sampling strategies. For all non-rocky outcrop classes, training samples were generated from a stable-pixel mask derived from MapBiomas Collection 10.1 for the 1985–2024 period. To produce this mask, the detailed MapBiomas legend was aggregated into five broad thematic groups: Forest, comprising forest and savanna formations; Herbaceous and Shrubby Vegetation, comprising grassland and herbaceous sandbank vegetation; Farming, comprising pasture, agriculture, and forest plantation; Non-Vegetated Area, comprising beach, dune and sand spot, urban area, mining, and other non-vegetated areas; and Water, comprising wetland and water bodies. Only pixels that remained assigned to the same broad thematic group throughout the entire time series were retained as stable areas. From this stable-pixel mask, 4,800 stratified random samples were generated. The rocky outcrop samples were collected manually across the entire Cerrado biome. A total of 3,790 rocky outcrop samples were identified through the visual interpretation of Landsat imagery, supported by the knowledge and experience of specialists and image interpreters. The automatically generated and manually collected samples were subsequently combined into a single training dataset. The complete sample set was then visually reviewed and refined to ensure spatial and thematic consistency with the conceptual definition of rocky outcrop adopted in MapBiomas Collection 11.0.

Unlike the general land use and land cover classification, which is performed independently for each classification region, the rocky outcrop classification follows a non-regionalized approach. A 55-km buffer was generated around these samples to delineate the classification area of interest (AOI) (Figure 8). Both the selection of training samples used in this classification and the prediction procedure were restricted to this area, thereby concentrating the modeling effort on regions with known or potential occurrences of rocky outcrop. Classification was performed for each year using the Random Forest algorithm implemented in GEE via the **ee.Classifier.smileRandomForest** function. Based on the results of previous collections, the number of decision trees was set to 300 for the entire AOI, while the number of variables considered at each split was set to the square root of the total number of predictor bands. The model operated in multiprobability mode, generating a probability distribution for each class. The final class for each pixel was assigned based on the highest probability.

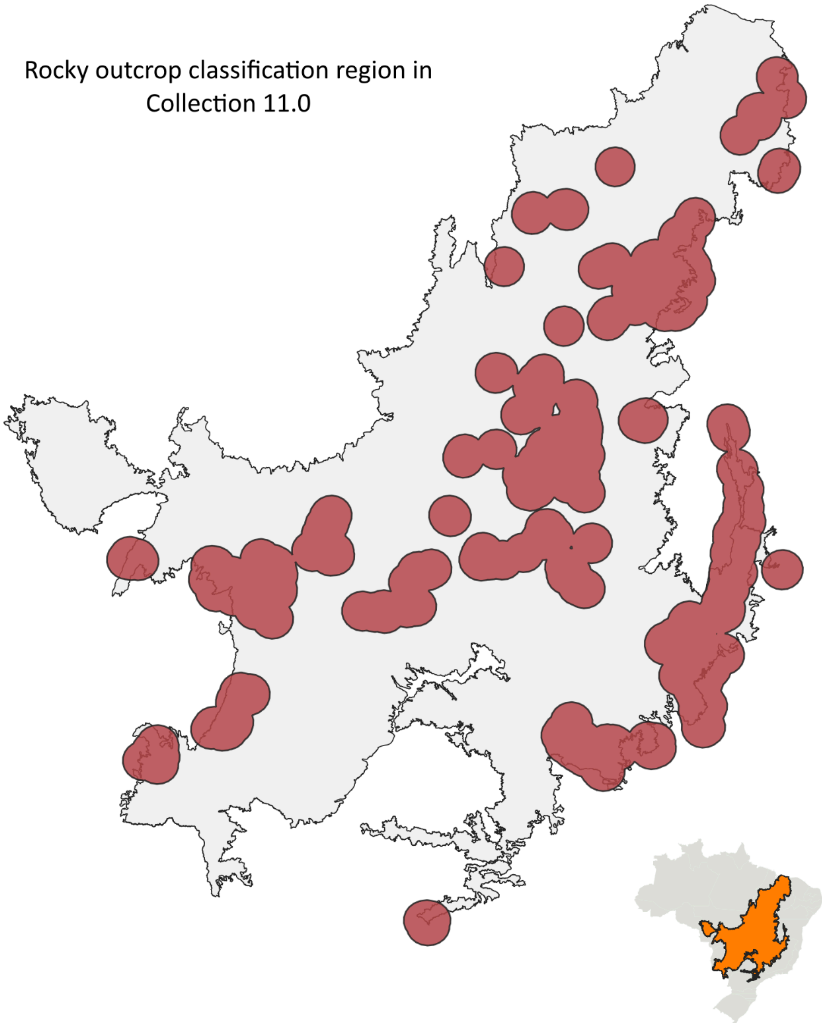


Figure 8. Area of interest used for rocky outcrop classification in MapBiomas Collection 11.0. The classification region corresponds to a 55-km buffer around the specialist-interpreted samples. The location of the Cerrado biome within Brazil is highlighted in orange.

6.3. Post-classification filters

The annual rocky outcrop classifications were refined using temporal gap filling, a frequency criterion, and spatial filtering. These procedures were applied independently from the post-classification workflow used for the general Cerrado map, reflecting the expected temporal and spatial organization of rocky and rupestrian landscapes.

- The gap-fill filter addressed inconsistencies or missing classifications across the time series. Unlike conventional unidirectional approaches, this filter operated bidirectionally (both forward and backward in time) filling undefined pixels by referencing the classification of subsequent and preceding years. This ensured temporal continuity in cases where rocky outcrop presence was stable but momentarily unclassified due to spectral noise or image limitations.

- The frequency filter was designed to reinforce the temporal stability of rocky outcrops. A pixel was retained as a rocky outcrop only if it was classified as such in at least 97% of the years within the observation period. This strict threshold was used to identify areas with persistent spectral and spatial association with rocky substrates or rupestrian vegetation and to remove isolated or temporally unstable classifications caused by confusion with exposed soil, burned areas, sparse vegetation, or other spectrally similar surfaces.
- Finally, a spatial filter was used to eliminate isolated pixels or small misclassified patches inconsistent with the typical spatial pattern of rocky outcrops. The filter removed connected components smaller than 11 connected Landsat pixels, approximately 1 hectare, based on an 8-neighbor connectivity criterion.

7. INTEGRATION

The integration stage combines the thematic outputs generated within the Cerrado workflow and harmonizes the resulting biome-level maps with the cross-cutting land use and land cover themes produced by the MapBiomas Brazil network. The process is implemented in three sequential stages described as follows:

7.1. Internal integration

The annual rocky outcrop classification was integrated into the corresponding general Cerrado map. Rocky outcrop (29) replaced the underlying general-map class within the validated rocky outcrop classification area, including pixels previously classified as native vegetation or anthropogenic land cover. Water-related restrictions were subsequently applied to prevent the insertion of rocky outcrop in persistent water bodies. After integration, isolated overlap artifacts containing six pixels or fewer were corrected using four-neighbor connectivity and the modal class within a two-pixel neighborhood. The resulting annual maps were then used in the Paraguay Basin harmonization and cross-cutting-theme integration stages.

7.2. Paraguay River Basin harmonization

A specific boundary-harmonization procedure has been applied in the *Bacia do Alto Paraguai (BAP)* since Collection 8.0. To reduce abrupt thematic discontinuities, the annual Cerrado classification is compared with the corresponding Pantanal map within the BAP boundary. The correction is restricted to pixels classified as mosaic of uses (21) in the Cerrado map. Where the Pantanal classification provides evidence of native vegetation, the Cerrado class is replaced according to the following correspondence:

- Cerrado mosaic of uses and Pantanal forest formation → savanna formation;
- Cerrado mosaic of uses and Pantanal savanna formation → savanna formation;
- Cerrado mosaic of uses and Pantanal wetland → wetland;
- Cerrado mosaic of uses and Pantanal grassland formation → grassland formation

7.3. Integration with cross-cutting themes

Each annual Cerrado map is combined with the cross-cutting thematic products generated by MapBiomas, including agriculture, pasture, mining, urban, and other land use and land cover themes. To harmonize the thematic data with the biome-level classifications, a set of predefined prevalence rules is applied. These rules define which classes take precedence when overlaps occur, ensuring a standardized decision logic across biomes and years. The prevalence rules adopted are described in Table 7. It is important to note that exceptions to this general rule apply in the context of protected areas regarding class prevalence:

- In protected areas, the native vegetation (3, 4, 11, and 12) is prevalent within the Cotton (62), Citrus (47), and Coffee (46) classes.
- In the case of pasture (15) within protected areas, native vegetation (3, 4, 11, and 12) is also preserved.
- Outside protected areas, pasture (15) is the prevailing land use class, superseding wetland and grassland classes (11 and 12). From Collection 4.0 through Collection 10.0, pasture also took precedence over savanna formation (4) in the Cerrado biome. This prevalence rule was removed in Collection 11.0.

Table 7. General prevalence rules - MapBiomas Collection 11.0.

Class	Pixel value	Prevalence order	Color
Photovoltaic Power Plant	75	1	
Wind farm (beta)	91	2	
Mining	30	3	

Class	Pixel value	Prevalence order	Color
Beach, Dune, and Sand Spot	23	4	
Mangrove	5	5	
Aquaculture	31	6	
Hypersaline Tidal Flat	32	7	
Urban Infrastructure	24	8	
Forest Plantation	9	9	
Rocky Outcrop	29	10	
Sugar Cane	20	11	
Soybean	39	12	
Rice	40	13	
Cotton	62	14	
Other Temporary Crops	41	15	
Coffee	46	16	
Citrus	47	17	
Other Perennial Crops	48	18	
Herbaceous Sandbank Vegetation	50	19	
River, Lake, and Ocean	33	20	
Forest Formation	3	21	
Savanna Formation	4	22	
Wetland	11	23	
Grassland Formation	12	24	
Pasture	15	25	
Mosaic of Uses	21	26	
Other Non-Vegetated Areas	25	27	

8. ACCURACY METRICS

The accuracy assessment of MapBiomas Collection 11.0 was conducted using a reference dataset provided by LAPIG/UFG, comprising 20,851 validation samples distributed across the Cerrado biome and covering the 1985–2025 period. The reference samples were interpreted by specialists with expertise in Cerrado LULC. Accuracy was assessed through confusion matrices comparing the reference samples with the corresponding classes in the integrated Collection 11.0 maps. The evaluation included overall and class-specific accuracy metrics, omission and commission errors, as well as quantity and allocation disagreements. All accuracy metrics are available at:

https://brasil.mapbiomas.org/en/analise-de-acuracia/. For a detailed description of the accuracy assessment methodology, including the sampling design, calculation and interpretation of accuracy metrics, see the MapBiomas Accuracy ATBD. Figures 9 to 11 show the overall accuracy, allocation error, and quantity error for Cerrado LULC maps.

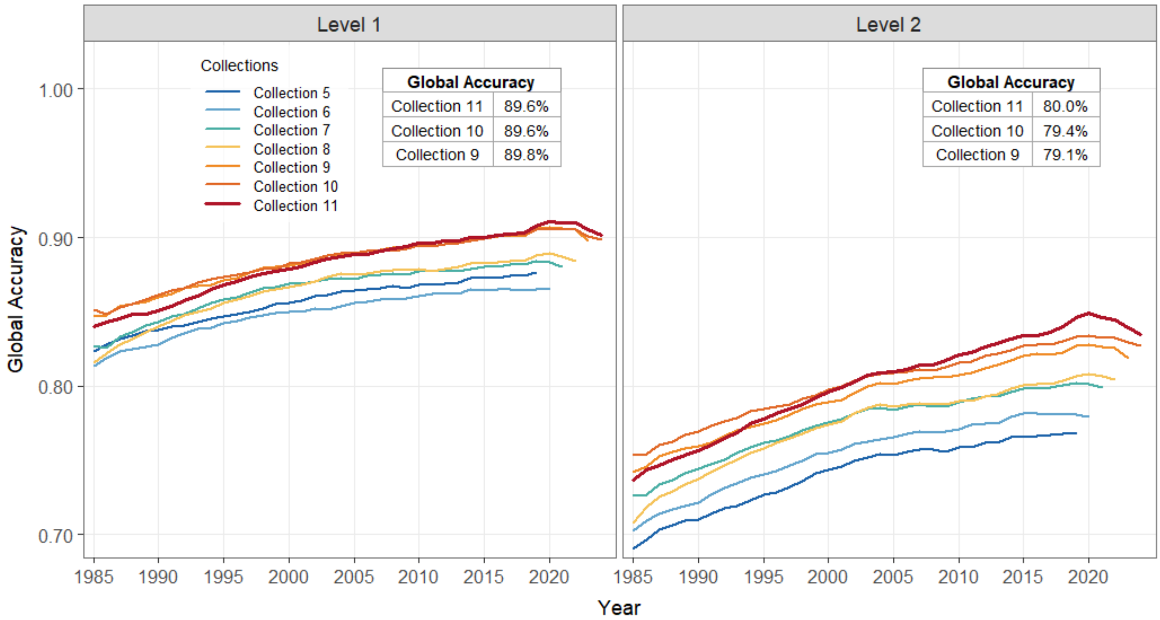


Figure 9. Global accuracy for the Cerrado biome at legend level 1 and level 2. The x-axis represents the years (from 1985 to 2025), while the y-axis represents the global accuracy value (1 = high accuracy). The colored lines indicate the accuracy per year of the current collection (11.0 - red line) and the previous collections (10.0, 9.0, 8.0, 7.0, 6.0, and 5.0 - orange to blue lines).

At Level 1, Collection 11.0 achieved an overall accuracy of 89.6%, unchanged from the 89.6% in Collection 10.0. Despite the similar overall accuracy, the composition of the disagreement changed between collections. Allocation disagreement decreased from 6.6% in Collection 10.0 to 5.9% in Collection 11.0, indicating improved spatial agreement between the mapped and reference classes. Conversely, quantity disagreement increased slightly, from 3.8% to 4.5%, indicating a modest increase in differences associated with the estimated proportions of the broad land-cover classes. At Level 2, Collection 11.0 showed a moderate improvement in overall accuracy, increasing from 79.4% in Collection 10.0 to 79.9%. This improvement was accompanied by a substantial reduction in quantity disagreement, from 11% to 8.5%, suggesting better agreement in the relative proportions of the mapped classes. Allocation disagreement, however, increased from 9.6% to 11.5%, indicating that part of the remaining classification disagreement was associated with the spatial allocation of classes rather than their overall mapped quantities.

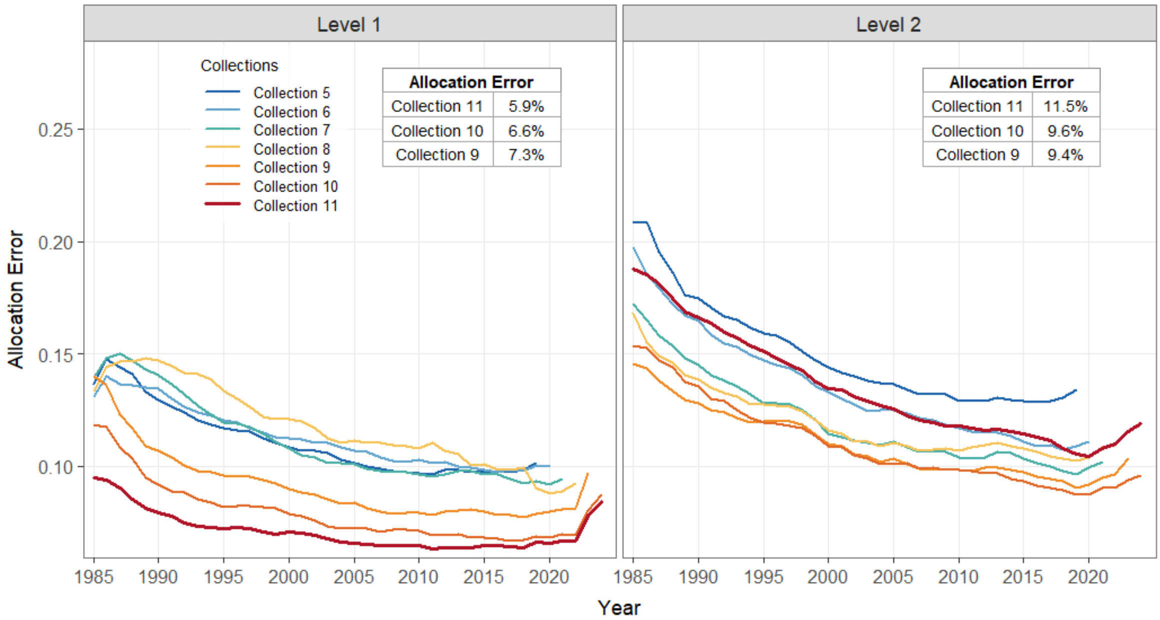


Figure 10. Allocation error for the Cerrado biome at legend level 1 and level 3. The x-axis represents the years (from 1985 to 2025), while the y-axis represents the allocation error value. The colored lines indicate the accuracy per year of the current collection (11.0 - red line) and the previous collections (10.0, 9.0, 8.0, 7.0, 6.0, and 5.0 - orange to blue lines).

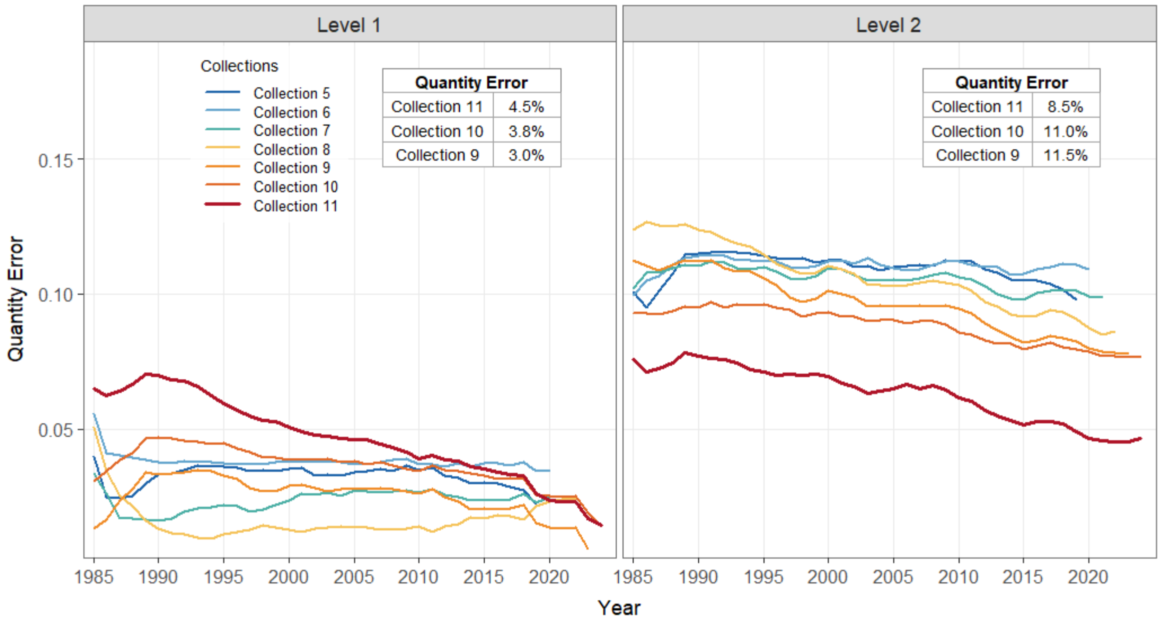


Figure 11. Quantity error for the Cerrado biome at legend level 1 and level 3. The x-axis represents the years (from 1985 to 2025), while the y-axis represents the quantity error value. The colored

lines indicate the accuracy per year of the current collection (11.0 - red line) and the previous collections (10.0, 9.0, 8.0, 7.1, 6.0, and 5.0 - orange to blue lines).

9. REFERENCES

Alencar, A., Shimbo, J. Z., Lenti, F., Marques, C. B., Zimbres, B., Rosa, M., Arruda, V., Castro, I., Ribeiro, J. P. F. M., Varela, V., Alencar, I., Piontekowski, V., Ribeiro, V., Bustamante, M. M. C., Sano, E. E., & Barroso, M. (2020). Mapping three decades of changes in the Brazilian Savanna native vegetation using Landsat data processed in the Google Earth Engine platform. Remote Sensing, 12(6), 924. https://doi.org/10.3390/rs12060924

Alencar, A. A. C., Arruda, V. L. S., Silva, W. V. da, Conciani, D. E., Costa, D. P., Crusco, N., Duverger, S. G., Ferreira, N. C., Franca-Rocha, W., Hasenack, H., Martenexen, L. F. M., Piontekowski, V. J., Ribeiro, N. V., Rosa, E. R., Rosa, M. R., Santos, S. M. B. dos, Shimbo, J. Z., & Vélez-Martin, E. (2022). Long-term Landsat-based monthly burned area dataset for the Brazilian biomes using deep learning. Remote Sensing, 14(11), 2510. https://doi.org/10.3390/rs14112510

Amatulli, G., McInerney, D., Sethi, T., Strobl, P., & Domisch, S. (2020). Geomorpho90m, empirical evaluation and accuracy assessment of global high-resolution geomorphometric layers. Scientific Data, 7, 162. https://doi.org/10.1038/s41597-020-0479-6

Burke, M., & Lobell, D. B. (2017). Satellite-based assessment of yield variation and its determinants in smallholder African systems. Proceedings of the National Academy of Sciences of the United States of America, 114(9), 2189–2194. https://doi.org/10.1073/pnas.1616919114

Celebrezze, J. V., Alegbeleye, O. M., Glavich, D. A., Shipley, L. A., & Meddens, A. J. (2025). Classifying rocky land cover using random forest modeling: Lessons learned and potential applications in Washington, USA. Remote Sensing, 17(5), 915.

Donchyts, G., Winsemius, H., Schellekens, J., Erickson, T., Gao, H., Savenije, H. H. G., & van de Giesen, N. (2016). Global 30 m Height Above the Nearest Drainage (HAND). Geophysical Research Abstracts, 18, EGU2016-17445-3

Gamon, J. A., Peñuelas, J., & Field, C. B. (1992). A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. Remote Sensing of Environment, 41(1), 35–44. https://doi.org/10.1016/0034-4257(92)90059-S

Gitelson, A. A., Kaufman, Y. J., & Merzlyak, M. N. (1996). Use of a green channel in remote sensing of global vegetation from EOS-MODIS. Remote Sensing of Environment, 58(3), 289–298. https://doi.org/10.1016/S0034-4257(96)00072-7.

Gitelson, A. A., Viña, A., Arkebauer, T. J., Rundquist, D. C., Keydan, G., & Leavitt, B. (2003). Remote estimation of leaf area index and green leaf biomass in maize canopies. Geophysical Research Letters, 30(5), 1248. https://doi.org/10.1029/2002GL016450

Hall, R. J., Skakun, R. S., Arsenault, E. J., & Case, B. S. (2006). Modeling forest stand structure attributes using Landsat ETM+ data: Application to mapping of aboveground biomass and stand volume. Forest Ecology and Management, 225(1–3), 378–390. https://doi.org/10.1016/j.foreco.2006.01.014

Hawker, L., Uhe, P., Paulo, L., Sosa, J., Savage, J. T. S., Sampson, C., & Neal, J. (2022). A 30 m global map of elevation with forests and buildings removed. Environmental Research Letters, 17(2), 024016. https://doi.org/10.1088/1748-9326/ac4d4f

Housman, I., Chastain, R., & Finco, M. (2018). An evaluation of forest health insect and disease survey data and satellite-based remote sensing forest change detection methods: Case studies in the United States. Remote Sensing, 10(8), 1184. https://doi.org/10.3390/rs10081184

Huang, Q., & Cai, Y. (2009). Mapping karst rock in Southwest China. Mountain Research and Development, 29(1), 14–20. https://doi.org/10.1659/mrd.857

INMET (Instituto Nacional de Meteorologia). (2024). Portal INMET. Brasília, DF. Accessed on: https://portal.inmet.gov.br/

Key, C. H., & Benson, N. C. (2006). Landscape assessment: Ground measure of severity, the Composite Burn Index, and remote sensing of severity, the Normalized Burn Ratio. In D. C. Lutes, N. C. Benson, M. Caratti, W. H. Key, C. H. Key, & L. J. Gangi (Eds.), FIREMON: Fire effects monitoring and inventory system (General Technical Report RMRS-GTR-164-CD, pp. LA-1–LA-55). U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station.

Lang, N., Jetz, W., Schindler, K., & Wegner, J. D. (2022). A high-resolution canopy height model of the Earth. arXiv. https://doi.org/10.48550/arXiv.2204.08322

Leal, F. M., & Cartagenes, R. C. C. (2021). Subsídios à elaboração do plano de manejo da ARIE Capetinga/Taquara na Fazenda Água Limpa – UnB [Trabalho de conclusão de curso, Universidade de Brasília]. Repositório Institucional da UnB.

Macena, F., Assad, E. D., Steinke, E. T., & Müller, A. (2008). Clima do bioma Cerrado. In A. C. S. Albuquerque & A. G. Silva (Eds.), Agricultura tropical: Quatro décadas de inovações tecnológicas, institucionais e políticas. Embrapa Informação Tecnológica.

Nagler, P. L., Inoue, Y., Glenn, E. P., Russ, A. L., & Daughtry, C. S. T. (2003). Cellulose absorption index (CAI) to quantify mixed soil–plant litter scenes. Remote Sensing of Environment, 87(2–3), 310–325. https://doi.org/10.1016/j.rse.2003.06.001

Parente, L., & Ferreira, L. (2018). Assessing the spatial and occupation dynamics of the Brazilian pasturelands based on the automated classification of MODIS images from 2000 to 2016. Remote Sensing, 10(4), 606. https://doi.org/10.3390/rs10040606

Qi, J., Chehbouni, A., Huete, A. R., Kerr, Y. H., & Sorooshian, S. (1994). A modified soil adjusted vegetation index. Remote Sensing of Environment, 48(2), 119–126. https://doi.org/10.1016/0034-4257(94)90134-1

Ribeiro, J. F., & Walter, B. M. T. (2008). As principais fitofisionomias do bioma Cerrado. In S. M. Sano, S. P. de Almeida, & J. F. Ribeiro (Eds.), Cerrado: Ecologia e flora (Vol. 1, pp. 151–212). Embrapa Informação Tecnológica; Embrapa Cerrados.

Rikimaru, A., Roy, P. S., & Miyatake, S. (2002). Tropical forest cover density mapping. Tropical Ecology, 43(1), 39–47.

Rosa, M. R. (2020). Metodologia de classificação de uso e cobertura da terra para análise de três décadas de ganho e perda anual da cobertura florestal nativa na Mata Atlântica [Doctoral dissertation, Universidade de São Paulo].

Rouse, J. W., Jr., Haas, R. H., Schell, J. A., & Deering, D. W. (1974). Monitoring vegetation systems in the Great Plains with ERTS. In Third Earth Resources Technology Satellite-1 Symposium: Vol. 1. Technical presentations, Section A (pp. 309–317). National Aeronautics and Space Administration.

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. (2019). Cerrado ecoregions: A spatial framework to assess and prioritize Brazilian Savanna environmental diversity for conservation. Journal of Environmental Management, 232, 818–828. https://doi.org/10.1016/j.jenvman.2018.11.108

Souza, C. M., Jr., Roberts, D. A., & Cochrane, M. A. (2005). Combining spectral and spatial information to map canopy damage from selective logging and forest fires. Remote Sensing of Environment, 98(2–3), 329–343. https://doi.org/10.1016/j.rse.2005.07.013

Xiao, J., Shen, Y., Tateishi, R., & Bayaer, W. (2006). Development of topsoil grain size index for monitoring desertification in arid land using remote sensing. International Journal of Remote Sensing, 27(12), 2411–2422. https://doi.org/10.1080/01431160600554363

Xu, H. (2006). Modification of normalized difference water index (NDWI) to enhance open water features in remotely sensed imagery. International Journal of Remote Sensing, 27(14), 3025–3033. https://doi.org/10.1080/01431160600589179

Zarco-Tejada, P. J., Rueda, C. A., & Ustin, S. L. (2003). Water content estimation in vegetation with MODIS reflectance data and model inversion methods. Remote Sensing of Environment, 85(1), 109–124. https://doi.org/10.1016/S0034-4257(02)00197-9