Pampa - Appendix

MapBiomas 10m - Collection 3

Version 1

General coordinator
Heinrich Hasenack

Team
Eduardo Vélez Martin
Macleidi Varnier
Juliano Schirmbeck
Eliseu Weber

April, 2026

1. Introduction

This document describes the procedures used to produce the Collection 3 of annual land use and land cover (LULC) maps for the Brazilian Pampa Biome (2017–2024). The mapped area covers 193,935 km² and corresponds to the geographical official limits of the Pampa biome (IBGE 2019).

The biome is predominantly characterized by gently undulating relief with altitudes below 300 meters. Geologically, the terrain is primarily sedimentary, particularly in the central and eastern regions. Additionally, volcanic igneous terrains are present in the northern and western areas, while igneous and metamorphic formations are found in the central-eastern region. A significant pedological diversity is observed, with a predominance of Argisols (Acrisols/Alisols), Neosols (Leptosols/Regosols), and Latosols (Ferralsols) (IBGE, 2021; NASA, 2020; SGB/CPRM, 2025).

The region exhibits a humid subtropical climate, characterized by the absence of distinct dry and rainy seasons. Monthly precipitation is reasonably uniform throughout the year, ranging from 1,200 to 1,800 mm annually. However, in many regions of the biome, periods of water deficit are common during the summer, driven by increased evapotranspiration and sporadic rainfall patterns—a condition that may be aggravated during La Niña events. The seasonality of the biome is clearly defined by thermal variations, where high summer temperatures contrast with low temperatures in the winter months, during which monthly averages fall below 15°C.

The Pampa Biome is mainly renowned for its vegetation, composed primarily of grasses and herbs, and for providing habitat for a diverse and specific biota typical of open ecosystems. It is part of the largest temperate grassland region in South America, covering the great plain of central-eastern Argentina, Uruguay and the south of Brazil. However, over the last 200 years large tracts of native grasslands have been mainly replaced by sown pastures, crops, and tree plantations, placing these grasslands among the most altered ecoregions of the world (Baeza *et al*., 2022).

The Pampa MapBiomas 10m Collection 3 represents the process of continuous improvements throughout successive collections (Table 1). It was built up following a sequence of steps, similar to those used in the previous collections, including:

1) Use of Google Satellite Embedding V1 (SATELLITE_EMBEDDING/V1/ANNUAL) and building annual Sentinel 2 (COPERNICUS/S2_SR_HARMONIZED) mosaics – using pixel values (median, minimum, amplitude and standard deviation) of the scenes within a temporal optimum period,

2) Definition of a feature space – set of remote sensing metrics, including original bands and several indexes used as potential class predictors,

3) Training samples – reference samples for each class, using the temporal stable samples approach,

4) Classification of predictor variables with training samples – using the Random Forest algorithm,

5) Post-classification treatment – a series of filters to remove noise, classification errors and unlikely transitions among years,

6) Final map assembling - integration of the classes mapped by the Pampa biome team with additional cross-cutting classes,

7) Validation - accuracy assessment using a set of independent multi annual validation samples and comparison with other available maps for the biome.

The classification stages for Collection 3 are similar to those of the preceding collections 1 and 2. The classification process was repeated every year aiming to solve the main commission and omission errors identified in the validation of Collection 2. In addition, some of the post-classification filters were enhanced and modified. All codes used for producing Collection 3 are available at GitHub (https://github.com/mapbiomas/brazil-pampa/tree/main/lulc_10m_sentinel/collection_03).

Table 1. Main characteristics of land use and land cover MapBiomas 10m collections for the Pampa biome

Collection	Year Range	Method	Mapped classes*	Key Improvements
1.0	2016–2022	RF	Forest formation, Wetland, Grassland, Mosaic of uses, Other Non-Vegetated Areas, Rocky Outcrop, and River, Lake and Ocean.	Feature space: a complete set of 140 bands and indexes
2.0	2016–2023	RF	Forest formation, Wetland, Grassland, Mosaic of uses, Other Non-Vegetated Areas, Rocky Outcrop, and River, Lake and Ocean.	Feature space: selection of the best 70 bands and spectral indexes. Covariates inclusion: slope, Height Above the Nearest Drainage and two segmentation bands generated with the Simple Non-Iterative Clustering (SNIC)
3.0	2017–2024	RF	Forest formation, Wetland, Grassland, Mosaic of uses, Other Non-Vegetated Areas, Rocky Outcrop, and River, Lake and Ocean.	Incorporation of the Google Satellite Embedding into the same feature space used in Collection 2 Classification of rock outcrops separately and within occurrence zones

*these are the classes mapped by the Pampa team on which the other classes mapped by the teams of the cross-cutting themes were applied

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

2. Methodology Overview

The LULC maps for the Pampa biome were developed following a supervised classification approach. The entire process was carried out in the Google Earth Engine cloud environment using the JavaScript language. Compared to the previous version, the main innovation implemented in Collection 3 is the use of Google Satellite Embeddings (Brown et al. 2025).

Figure 1 shows the workflow applied in the mapping. It was divided into five main steps: (i) generation of training samples; (ii) selection of predictive variables; (iii) LULC classification; (iv) post-classification filters and integration logic; and (v) LULC accuracy assessment.

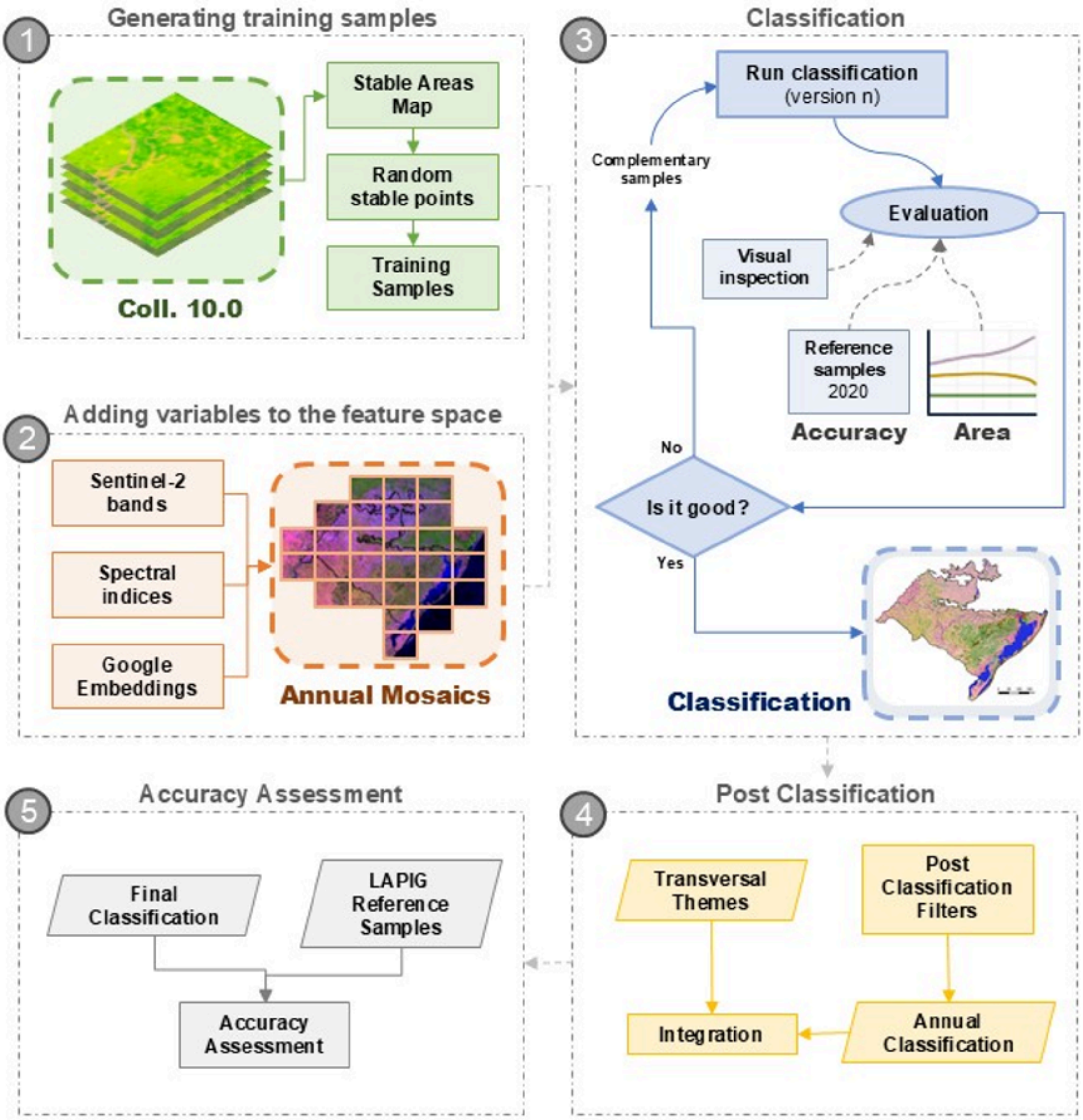


Figure 1. Classification workflow.

The thematic mapping was made for nine classes: Forest formation, Wetland, Grassland, Mosaic of uses, Other Non-Vegetated Areas, Rocky Outcrop, and River, Lake and Ocean, Wooded sandbank vegetation and Herbaceous sandbank vegetation. In the following sections, each methodological step is explained in detail.

2.1 Regionalization

The Pampa biome was divided into seven regions based on the work of Hasenack et al. (2022). In that study, the authors delineated the Pampa biome using a digital elevation model, a soil map, and a phytoecological regions map, resulting in ten biophysical groups. To facilitate the classification process, smaller regions were aggregated into adjacent larger regions, resulting in seven regions (**Figure 2**): 1) Bush Grassland, Atlantic Sub-Montane Grassland, and Coastal Subtropical Forest; 2) Shortgrass Grassland; 3) Shallow Soil Grassland; 4) Aristida Grassland and Sandy Grassland; 5) Park Grassland; 6) Inland Sub-Montane Grassland; and 7) Coastal Grassland. Independent classification models were applied to each region using local samples; the resulting annual classifications were then aggregated to generate the final biome mosaic.

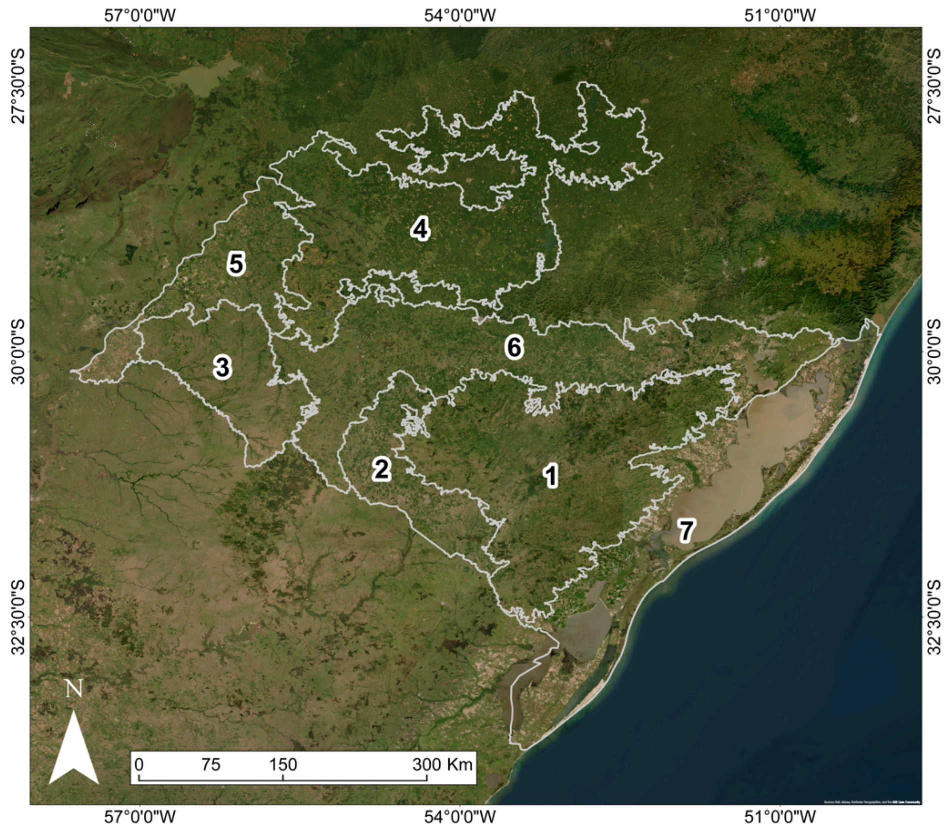
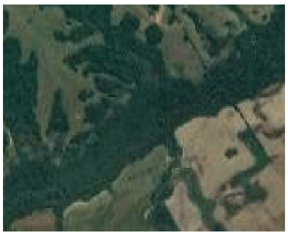
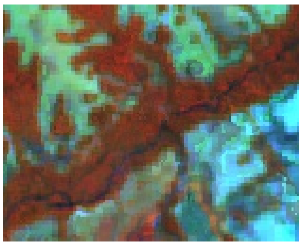
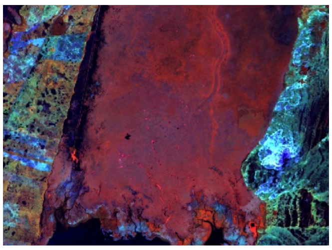


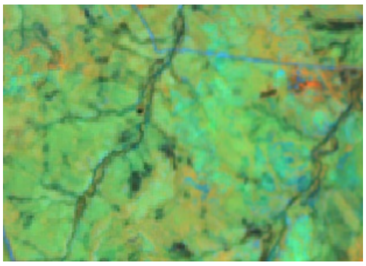
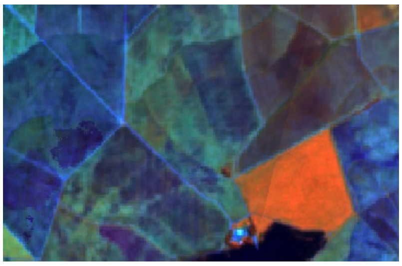
Figure 2. Regions of the Pampa biome used in the classification process.

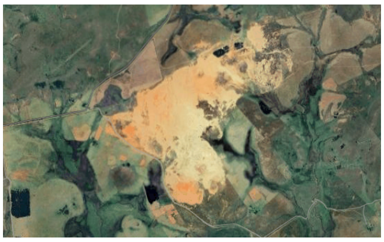
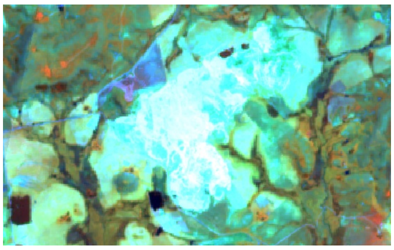
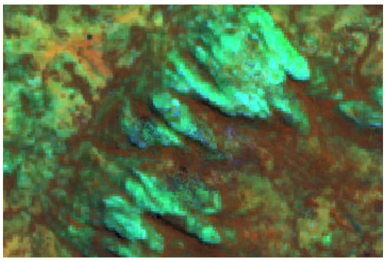
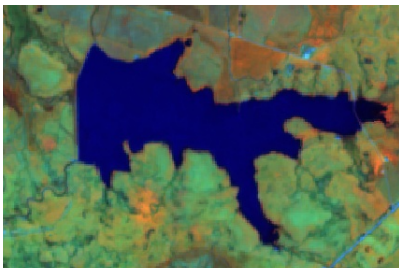
2.2. Mapped Classes

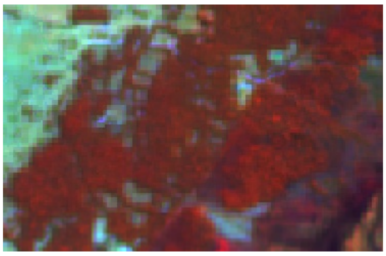
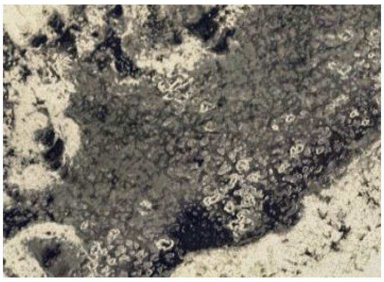
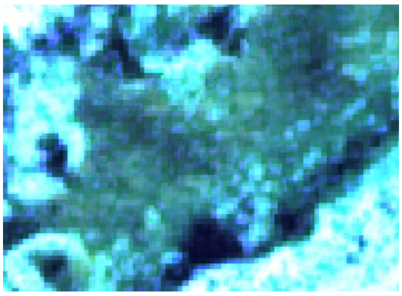
The nine LULC classes mapped by the MapBiomas Pampa team, before the integration with transversal themes, are detailed in **Table 2**. Two of these classes (Wooded Sandbank Vegetation and Herbaceous Sandbank Vegetation) were added through partial reclassification during the post-classification process, and their details are described in **Section 2.5.2.** The integration with cross-cutting themes is also explained in **Section 3**.

Table 2: Classes mapped by the Pampa biome in the MapBiomas Collection 3.

Forest Formation		
Numeric ID: 3	Color	
Google Earth	Sentinel 2 (RGB 8 11 4)	Photo
		
Description: Woody vegetation with trees or shrubs species, with a predominance of a continuous canopy. It includes ombrophilous, deciduous and semi-deciduous forests, and part of the pioneer formations.		
Wetland		
Numeric ID: 11	Color	
Google Earth	Sentinel 2 (RGB 8 11 4)	Photo
		
Description: Swampy areas, regionally called "banhados" or "marismas" (saltmarshes). Typically, hygrophilous vegetation, with emergent, submerged or floating aquatic plants. They occupy plains and depressions in the terrain with waterlogged soil and the shallow margins of ponds or water reservoirs.		
Grassland Formation		

Numeric ID: 12	Color	
Google Earth	Sentinel 2 (RGB 8 11 4)	Photo
		
Description: Vegetation with a predominance of an herbaceous grassy layer, together with herbaceous and subshrub dicotyledons. The botanical composition is influenced by edaphic and topographic gradients and by pastoral management (livestock). They occur in deep to shallow soils, including rocky (rupestrian grasslands) and sandy terrains (sandy or psammophilous grasslands). They occupy well-drained soils (mesic grasslands) or even soils with higher moisture content (humid grasslands - with a marked presence of sedges). In most cases, it corresponds to native vegetation, but patches of invasive exotic wild or forage species (pastures) may be present.		
Mosaic of uses		
Numeric ID: 21	Color	
Google Earth	Sentinel 2 (RGB 8 11 4)	Photo
		
Description: Areas of agricultural use, where it was not possible to distinguish between pasture and agriculture. May include cropland, winter or summer pastures and horticulture. Includes areas left unsown for a period between agricultural crops (fallow).		
Non-Vegetated Areas		
Numeric ID: 22	Color	
Google Earth	Sentinel 2 (RGB 8 11 4)	Photo

		
Description: Mixed class that includes natural and anthropic areas. Natural areas include sandy surfaces such as river beaches and sandy terrains. Anthropic areas include areas of exposed soil and non-permeable surfaces (infrastructure, urban sprawl or mining).		
Rocky Outcrop		
Numeric ID: 29	Color	
Google Earth	Sentinel 2 (RGB 8 11 4)	Photo
		
Description: Rocks naturally exposed, often with partial presence of rupicolous vegetation.		
Water		
Numeric ID: 33	Color	
Google Earth	Sentinel 2 (RGB 8 11 4)	Photo
		
Description: Areas with the presence of permanent surface water. Includes streams, rivers, ocean, lagoons, natural and artificial lakes.		
Wooded sandbank vegetation		
Numeric ID: 49	Color	
Google Earth	Sentinel 2 (RGB 8 11 4)	Photo

		
Description: Arboreal vegetation over sandy soils along the coastal zone of the biome.		
Herbaceous sandbank vegetation		
Numeric ID: 50	Color	
Google Earth	Sentinel 2 (RGB 8 11 4)	Photo
		
Description: Herbaceous vegetations over sandy soils along the coastal zone of biome.		

2.3 Feature Space

The feature space was composed using data from two different datasets: i) Google Satellite Embeddings v1 and ii) Sentinel-2 images.

2.3.1 Embeddings

Google Satellite Embeddings v1 is a product developed by Google DeepMind using the AlphaEarth Foundations model. Each pixel of this product has a spatial resolution of 10 m and a 64 dimensional representation that encodes temporal trajectories of surface conditions at and around each pixel, as measured by multiple Earth observation instruments and datasets, including Sentinel-1 and Sentinel-2, Landsat-8 and Landsat-9, PALSAR-2, GEDI, ERA5-Land, GRACE, GLO-30, and NLCD. Unlike spectral bands or spectral indices, which are direct physical measurements of surface properties, satellite embeddings are feature vectors that summarize relationships across multi-source and multi-modal observations in a less directly interpretable, but more powerful way (BROWN et al., 2025).

The data represent a full calendar year. For Collection 3, the 64 bands per year from the 2017–2024 period were used.

2.3.2 Sentinel-2 mosaics

Annual image mosaics were produced combining bands and several derived indexes using central tendency and variability reducers, calculated over all the images available in a previously defined "optimal period" within the year (September-November) and within temporal ranges defined through higher and lower NDVI values (extracting data only from those scenes with values over the third (Q3) and below the first (Q1) quartile, respectively.

The optimal period was established as the annual timeframe presenting a specific combination of natural vegetation phenology and land preparation interventions in agricultural areas, which maximizes the contrast conditions between the supervised classification targets.

It was also included spectral indexes supplemented with geolocation (latitude and longitude), the ANADEM/UFRGS slope datasets and the Global Height Above the Nearest Drainage (HAND) (Donchyts, et al., 2016). Finally, two segmentation bands generated with the Simple Non-Iterative Clustering (SNIC) algorithm were added to the feature space. The median values of the NDVI and EVI bands and a grid of 20 and 80 pixels as the location spacing of the superpixel seed, respectively, were used to cluster pixels with similar spectral characteristics. The average values per cluster resulting from the NDVI and EVI bands were added to the feature space.

The final feature space comprised 145 variables. However, only 40 variables were selected for classification. This selection considered the variables that showed the highest importance across regions 1 and 3[1], as well as those that consistently exhibited high importance in most years of the collection in a previous classification. Table 2 shows all the bands and indices used in the classification.

[1] These regions were selected for variable selection due to their class complexity and because they well represent most of the extreme dynamics occurring in the Pampa biome.

Table 3. Feature space subset considered in the classification of Collection 3 of the Pampa biome.

Type	Name	Statistics
Sentinel-2 band	Red	median_dry
	Green	stdDev, median, median_dry
	Red Edge 1	stdDev, median_wet
	Red Edge 3	median
	Red Edge 4	median_dry
	NIR	median
	SWIR 1	median_wet
	SWIR 2	median_dry
Spectral Index	Advanced Vegetation Index	median_dry
	Aerosol Free Vegetation Index	median_wet
	Brightness Index	median_wet
	Co2flux	median
	Disease-Water Stress Index 5	median
	Enhanced vegetation index	median, median_wet, median_cluster, median_dry
	Green Chlorophyll Vegetation Index	median_wet
	Green Leaf Index	median_dry
	Global Vegetation Moisture Index	median_dry, median_dry_1, median_wet_1
	Indicator of water index	median_wet
	Land Surface Water Index	median_dry
	Moisture Stress Index	median
	Normalized Difference Drought Index	median_wet
	Normalized Difference Vegetation Index	median_cluster
	Photochemical Reflectance Index	median_wet
	Redness Index	median_dry

Type	Name	Statistics
Spectral Index	Relative Vigor Index	median*, median_1**
	Shape Index	median_dry, median
	Urban Index	median_wet
	Wetness Index	median_wet
Terrain	Height Above Nearest Drainage	Hand value
Coordinates	Latitude	
Embeddings	Google Satellite Embedding v1	Bands A00 - A63

*median = refers only for the months that constitute the mosaics's optimal period,

**median_1 = refers to all months of the year.

2.4 Samples

The training samples were generated using two approaches: i) automatic random sampling from stable areas derived from MapBiomas annual maps (Landsat Collection 10), and ii) and complementary stable samples derived from the visual interpretation of remote sensing imagery. The concept of stable samples refers to a temporal criterion; that is, pixels selected from areas within maps classified in a previous collection where the corresponding class remains unchanged over the years

2.4.1 Stable Samples

The automatic random sampling from stable areas for each class was derived from Landsat based MapBiomas Collection 10. First, the maps were reclassified into the seven classes mapped by the MapBiomas Pampa team: Forest Formation, Wetland, Grassland, Mosaic of Uses, Other Non-Vegetated Areas, Rocky Outcrop, and River, Lake, and Ocean. Subsequently, only pixels that did not change during the 1999–2024 period were retained. Next, for each Pampa region, 2,000 samples per class were randomly generated.

The final number of samples per class in each region used in the classification process was defined respecting the proportional area occupied by class *i* in year *j*. Class weights were calculated as the proportion of each class area based on MapBiomas Collection 10 and multiplied by the 2,000 points available in the training dataset. During training, when systematic errors were observed (omission or commission errors affecting an entire region), the number of samples for a given class was manually adjusted.

2.4.2 Complementary samples

Complementary samples were generated iteratively. When gross classification errors were identified, samples were added at the corresponding locations based on the visual interpretation of high-resolution imagery available in Google Earth Engine and annual Sentinel-2 mosaics (RGB 8–11–4). To ensure that the complementary samples were representative of the entire mapped period, samples were collected only when it was possible to confirm that the class did not change during the 2017–2024 period.

2.5 Classification algorithm

The classification was divided into two independent processes: a general classification and a Rocky Outcrop classification.

2.5.1 General classification

The classification was performed using the Random Forest algorithm (BREIMAN, 2001) with annual predictor variables (presented in **Table 3**) for each year from 2017 to 2024. In the general classification, six LULC classes were considered, including: Forest Formation, Wetland, Grassland, Mosaic of Uses, Non-Vegetated Areas, and Water. The training was performed separately for each region to better represent their specific characteristics. The parameters used in the Random Forest were 100 trees and one variable per split. An iterative approach was adopted: An initial classification was performed using only stable samples and a set of initial sample weights for the classes. Intermediate classifications were generated adjusting manually the weights for those classes with higher commission or omission errors, combined with the inclusion of additional stable samples in areas of misclassification.

The final classification was achieved after several repeated model runs, with results evaluated using a set of reference samples, visual comparison with Sentinel-2 mosaics, and analysis of the mapped area behaviour of each class along the years. To ensure a smooth transition across regional boundaries, a 100 m buffer was applied to each region. The final annual LULC mosaic was generated merging all the seven classified regions for each year. In those areas of classification that overlap within buffers between neighboring regions, the prevalence of the class with the lowest numerical value was adopted as the decision rule for pixel assignment. This criterion prioritizes natural classes over anthropogenic ones, as the MapBiomas legend assigns lower numerical values to natural classes.

2.5.2 Rocky outcrop classification

The Rocky Outcrop class was mapped independently from the general classification, but using the same predictive variables and random forest parameters. However, the classification was restricted to a large set of regional patches previously identified on zones of non-sedimentary geology and potentially containing rocky outcrops (**Figure 3**).

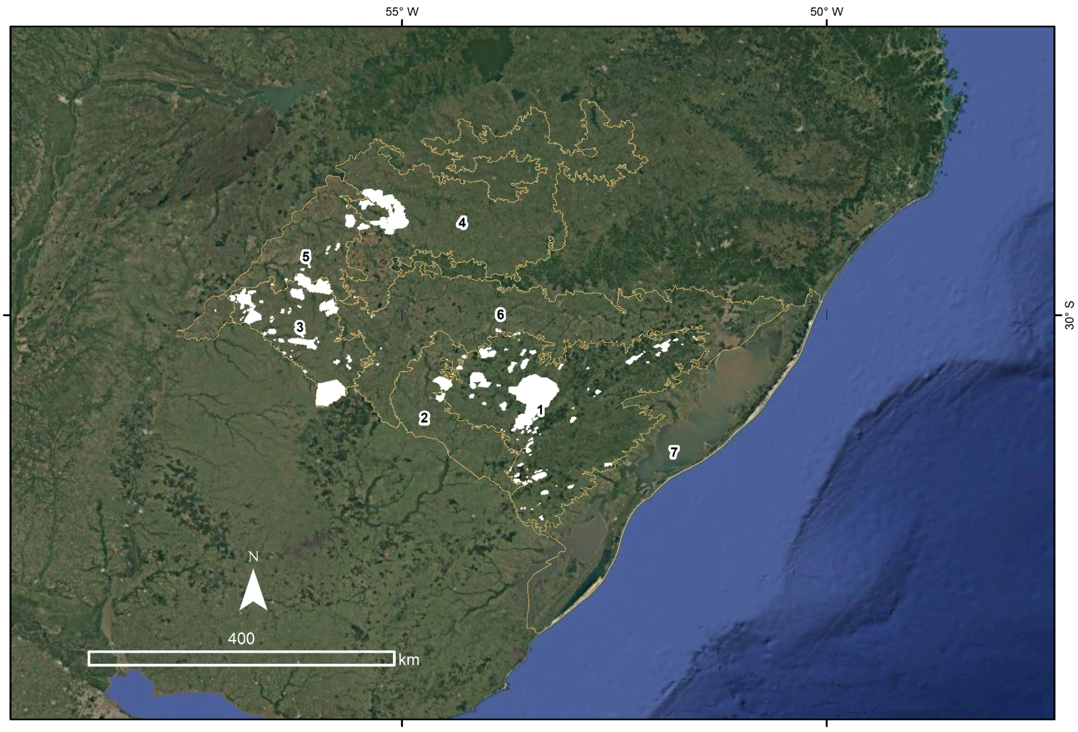


Figure 3. Rocky Outcrop zones (white patches).

A binary classification approach was adopted to represent the presence and absence of Rocky Outcrops. The training dataset comprised 500 stable samples for absence and 250 for presence, as well as 360 complementary samples for absence and 160 for presence. The classification was performed only for the year 2022 and considered as a stable data layer valid for the other years. The final classification involved repeated model runs, with results evaluated using visual comparison with Sentinel-2 mosaics, and analysis of the mapped area.

2.6 Post Classification

The following sections describe the sequence of filters applied in the mapping of MapBiomas Pampa Collection 3.

2.6.1 Gap Fill filter

This filter uses information of previous years to replace pixels classified as Non-observed in a given year through a forward procedure. Then, it is complemented by the information of later years, through a backward procedure, to replace pixels that remained as Non-observed.

2.6.2 Incident filter

The incident filter was applied to correct cases of false transitions observed in pixels placed in the border of patches belonging to different classes. The classification of pixels located at the boundary of two classes, are sensitive to be classified to one or another of these classes along the years because of unwanted effects. They may include small tilts in the acquisition of the original radiometric information by the satellite sensors, associated with issues of georeferencing precision, but also to the effect in data acquisition due to drier or wetter years. The consequence is an increased level of false transitions for these pixels over the years.

The filter selects all pixels presenting more than three incidents over the 8 years (Pontius et al., 2017) and belonging to patches of less than six pixels and replaces them by the modal class. It is assumed that small patches with high incident rates are indicative of confusions that are common in edge pixels.

2.6.3 Temporal filter

The temporal filter uses the information from previous and subsequent years to identify and correct pixel misclassifications for a given year, assuming a set of invalid transitions.

The rules differ for the first, the last and intermediate years of the collection. The process starts looking at the first three years, comparing the class in the year 2017 with the two following. Whenever a pixel in the first year differs from a native vegetation class (3, 11, 12) and is assigned to one in the two next years (2018 and 2019), then it is replaced by the corresponding natural class. For the last three years, the year 2024 is compared with the two precedents and, whenever a pixel is classified as 21 (Mosaic of Uses) in both but is different in the last year, then it is replaced by class 21. Both procedures aim to avoid cases of false positives of regeneration.

The last step applies a 3-year moving window to correct the remaining intermediate years. Whenever the first and the third year of the window have the same class and the middle year is different, it is replaced by their class. This procedure has the purpose of fixing abrupt transitions that are unlikely to happen. The filter was applied, step by step, respecting the following sequence of classes: [22, 21, 11, 3, 12, 33].

2.6.4 Temporal natural areas filter

This filter aims to reduce improbable temporal variations and enforce stability in predominantly natural areas within the 2017–2024 time series. First, classes were categorized into two groups: natural (forest, wetland, grassland, and water) and anthropic (mosaic of uses and non-vegetated areas). A temporal filter was then applied using a 3-year moving window to identify isolated 'natural-anthropic-natural' occurrences, replacing the intermediate year with the subsequent class. Finally, a class frequency filter was applied per pixel: in areas where natural classes occurred in more than 80% of the series, pixels were reclassified using the mode. This process reduces short-term noise, resulting in a more consistent time series.

2.6.5 Frequency filter

Frequency filters were applied to use the information available for each pixel over the years to correct cases of false positives. The general logic of the frequency filters is to search for each pixel a specific combination of classes throughout the 8 years producing a subset of pixels considered eligible for correction. Then the filter detects and overwrites only those years where cases considered as probable false positives using a fixed class value, that usually is the mode of classifications detected along the temporal range. This type of filter was used with parsimony to solve only well delimited cases. Four different variations of the frequency filters were employed in Collection 10, as described below:

i) Frequency filter 1 – Wetlands x other natural classes

The first rule selects pixels where wetland frequency exceeds 70% of the temporal window and the remaining years are classified as other natural classes (grassland, forest, or water). Pixels are considered only when the combined frequency of wetland, grassland, forest, and water exceeds 80% and the temporal incidence is greater than 1, indicating instability among natural wet-environment classes. In these cases, the non-wetland years are considered false positives and are reassigned to the wetland class.

The second rule selects pixels where wetland frequency is lower than 25% of the temporal window and the incidence is greater than 1, indicating temporal instability. In these cases, wetland occurrences are considered false positives and the years classified as wetland are replaced by the temporal mode of the pixel over the 8-year window.

The third rule selects pixels with intermediate wetland frequency (25–70%) occurring within wet environments. Pixels are selected when the combined frequency of wetland, water, and forest exceeds 75% of the temporal window and the frequency of forest is lower than 33%. In these cases, years classified as forest are considered false positives and are replaced by wetland. Water classes are not modified.

ii) Frequency filter 2 – Paddy rice x wetlands

The second frequency filter was designed to reduce confusion between paddy rice and temporary water or wetlands, and was applied only in regions 5, 6, and 7. The filter selects pixels where the combined frequency of agriculture (class 21), wetland (11), and water (33) exceeds 75% of the temporal window and where agriculture frequency is greater than 33%. In these cases, the pixel is interpreted as irrigated agriculture with seasonal flooding, and all years are reassigned to class 21. This correction accounts for the post-harvest flooding period of paddy rice fields, which may otherwise be misclassified as water or wetland.

iii) Frequency filter 3 – Non vegetated areas

The third rule selects all pixels with high frequency (>80%) of class 22 (non-vegetated areas). All the remaining years classified differently from class 25 are then replaced by class 22. This majority criteria is an attempt to minimize the problem of misclassification. The second rule selects all pixels with low frequency (<20%) of class 22. Then those years classified as class 22 are replaced by the modal class.

iv) Frequency filter 4 – Relief shadow x water and wetland

This filter was designed to reduce false positives of water and wetland caused by terrain shadow or spectral confusion in areas where their occurrence is topographically unlikely. It was applied only in regions 1, 2, 3, 4, and 5. Pixels were selected when water (class 33) or wetland (class 11) occurred with low to intermediate frequency within the temporal window and the pixel was located in HAND class 2, representing sloping or elevated areas relatively distant from drainage networks. In such locations, the presence of persistent surface water or wetlands is unlikely.

Occurrences of water and wetland were therefore considered false positives and replaced by the temporal mode of the pixel calculated excluding the respective class.

v) Frequency filter 5 – Agriculture x Nom-vegetated area

The fifth rule aims to correct pixels that show temporal instability between Agriculture (class 21) and Non-vegetated/Anthropic areas (class 22). For each pixel, the frequency of occurrence of these two classes is calculated over an 8-year period, and the correction is applied when the combined frequency of Agriculture and Non-vegetated area exceeds 95%, indicating that the pixel alternates almost exclusively between these two classes, and when Non-vegetated area occurs in more than 45% of the period. In such cases, the pixel is considered unstable or noisy, and its entire time series is replaced by the modal class (the most frequent class over the period), reducing artificial fluctuations between Agriculture and exposed or non-vegetated surfaces.

2.6.6 Time-series start/end filter

This filter was applied to smooth transitions between the penultimate and final years of the time series. A minimum transition area filter was implemented to remove transitions between natural and anthropic classes occurring in patches equal to or smaller than 1 hectare. In these cases, pixels in the final year were reassigned to the same class observed in the penultimate year.

2.6.7 Spatial filter

The same 3-year temporal filter described above (section 8.1.3) was applied to remove accidental and unwanted effects in the filtered classification resulting from the combined application of the filters.

The spatial filter also was applied for Rocky Outcrop classification in order to reduce isolated pixels and very small patches, improving the spatial consistency of the final map. This procedure used a focal mode operator in a 3 × 3 pixel window, replacing pixels belonging to small connected patches (≤ 6 pixels) with the dominant class in the surrounding neighborhood. This approach aimed to minimize classification noise while preserving the spatial pattern of the mapped features.

2.6.8 Granular filter

This filter aims to reduce classification noise and improve spatial consistency. Small patches (smaller than approximately 2 ha) classified as Class 21 were identified using a connected pixel analysis. For each patch, the immediate surrounding context was evaluated within a 30 m buffer, where the proportions of the natural classes (grassland, forest, wetlands, and water) were calculated. Small patches embedded in predominantly natural environments were then reassigned to the dominant natural class in their surroundings. The correction was applied only to eligible areas and integrated into the original annual maps, resulting in a refined classification with fewer isolated fragments and improved spatial coherence across the landscape.

2.6.9 Shadow filter

This filter was applied to reduce false water positives in region 1. After the classification, shadows cast by rocky outcrops were identified as a source of water misclassification. To address this issue, a 100 m buffer was first generated around the Rocky Outcrop polygons. Then, using a slope product derived from NASADEM (NASA/NASADEM_HGT/001) and a flow accumulation layer from Hydrography90 (AMATULLI *et al.*, 2022), water pixels located within the buffer and presenting slopes greater than 15° and drainage areas smaller than 1 km^2 were reassigned based on the mode of a 9 × 9 neighborhood. This approach reduced the misclassification.

2.7 Special classes

After the post-classification process, the annual classifications for each one of the seven regions were merged to compose the biome map. Then three additional classes were added up to the annual maps. First, the rocky outcrop classification was overlaid to each annual map, then the wooded sandbank vegetation and herbaceous sandbank vegetation classes were added through a reclassification process. The class Wooded sandbank vegetation was defined as reclassifying all the patches classified as Forest, and the class Herbaceous sandbank vegetation was defined as reclassifying all the patches classified as Grasslands, both within the limits of the Brazilian marine and coastal system (IBGE, 2019) and belonging to Regosols patches from a soil classification map (IBGE, 2021).

3. Integration with cross-cutting themes

The annual maps resulting from the post-classification filters for each one of the 8 years (2017-2024), and containing nine classes, were integrated with the maps of the cross-cutting themes. This assembling was performed using hierarchical prevalence rules (**Table 4**). The output is a final set of integrated LULC maps for the Pampa biome now containing 14 classes

Additionally, some exceptions to these prevalence rules were applied due to particularities observed in the Pampa biome, including:

1. Class 15 (pasture) converted to class 21 (mosaic of uses). The class pasture is classified without distinction between grasslands and planted pastures and is almost totally overlapped in the integration process by class 12 (grasslands). The inspection of the remaining pixels of class 15 shows that they do not necessarily refer to perennial planted pastures as expected. Most of the remaining pixels within this class correspond to fallow areas or abandoned agricultural lands undergoing initial regrowth of native vegetation. Both cases fall under the mosaic of uses class definition.

2. Classes 11 (wetland), grassland (12) and water (33) prevails over class 19 (temporary crops). This rule was adopted to remove commission errors.

3. Class 11 (wetland) prevails over class 9 (silviculture). The rule was adopted to remove commission errors.

4. Classes 9 (silviculture) and 19 (temporary crops) were not allowed to overlay natural classes within the boundaries of the Lagoa do Peixe National Park. This rule was necessary to remove commission errors.

Table 4: Hierarchical prevalence rules for combining the output of the Pampa classification with the cross-cutting themes in Collection 3. *missing classes in the biome map.

Class	Class ID	Prevalence order
Mining	30	1
Beach, Dune and Sand Spot	23	2
Mangrove*	5	3
Aquaculture*	31	4
Salt Flat*	32	5
Urban area	24	6
Forest plantation	9	7
Rocky Outcrop	29	8
Perennial Crop*	36	9
Temporary Crop	19	10
Herbaceous Sandbank vegetation	50	11
River, Lake and Ocean	33	12
Other Non Vegetated Areas	25	13
Forest Formation	3	14
Savanna Formation*	4	15
Wooded Sandbank vegetation	49	16
Wetland	11	17
Grassland	12	18
Pasture*	15	19
Mosaic of Uses	21	20

4. Validation

The classification performance was assessed using two reference datasets: the LAPIG reference samples and a reference map for the year 2020 available for a portion of the biome (SH-22-Y-D).

4.1 LAPIG dataset

For the LAPIG dataset, 1,936 reference samples were available per year for the period 2017–2024. These reference samples are widely used for the evaluation of mappings produced by the MapBiomas initiative at a 30 m spatial resolution, since the interpreters that produced this dataset used Landsat imagery. Because Sentinel mapping with a 10 m spatial resolution is being validated with this coarser data, the results carry a methodological limitation inherent to this comparison, requiring cautious interpretation. In the absence of a superior validation dataset, this analysis is deemed a reasonable approximation of the true accuracy.

Due to the limited number of samples for some classes, the accuracy evaluation includes the following seven classes: Forest, Forest Plantation, Wetland, Grassland, Agriculture (including Temporary and Perennial Crops), Non-Vegetated Areas (including Beach, Dune, Sand Spot, Urban Area, Mining, and Other Non-Vegetated Areas), and Water. First, the classification and reference samples were reclassified according to the previous classes definitions. Then, for each reference sample, the class value of the corresponding classification pixel was extracted. Using these data, confusion matrices were generated for each year, and global accuracy, user's and producer's accuracies, as well as allocation and quantity disagreements, were evaluated. Finally, to obtain a metric summarizing the entire collection, the cells of all eight yearly matrices were summed and divided by the number of years (8). This resulted in the mean accuracies for Collection 3 in the Pampa biome.

Table 5 shows the yearly global accuracies, as well as allocation and quantity disagreements for the 2017–2024 period. It is observed that global accuracies range from 88% to 89.9%, while allocation disagreements range from 8.4% to 10%, and quantity disagreements from 0.8% to 2.7%. These results demonstrate temporal consistency, with small variations over the years.

Table 5. Global accuracies and disagreements in the classification of Collection 3 of the Pampa biome 2017-2024.

Year	Global accuracy	Allocation disagreements	Quantity disagreements
2017	88.0%	10,0%	2%
2018	88.9%	8.4%	2.7%
2019	89.5%	8.8%	1.8%
2020	89.7%	9.1%	1.2%
2021	89.8%	9.1%	1.1%
2022	89.9%	9.2%	0.9%
2023	89.7%	9.4%	0.9%
2024	89.4%	9.8%	0.8%

Figure 4 shows the mean confusion matrix for the 2017–2024 period. The overall accuracy is 89.3%, with 9.4% allocation disagreement and 1.2% quantity disagreement. For user's accuracy, Forest, Forestry, Agriculture, Non-vegetated areas, and Water achieved values greater than 90%. Grassland achieved 87.6%, while Wetland presented a value of 68.8%. For producer's accuracy, Forest, Forestry, Grassland, Agriculture, and Water achieved values greater than 87%. Wetland achieved 60.5%, while Non-vegetated areas achieved 68.3%. In addition, the greatest disagreement occurred between Grassland and Agriculture. **Figure 5** details the errors of omission and commission for each class, indicating each of the classes related to these errors.

Referência / *Classificação*

	Forest	Forestry	Wetland	Grassland	Agriculture	Nom-vegetated areas	Water	Total	AU	EI
Forest	110	3	2	3	0	0	0	118	0,931	0,069
Forestry	3	55	0	1	2	0	0	60	0,907	0,093
Wetland	0	0	22	10	0	0	0	32	0,688	0,313
Grassland	10	1	8	648	60	13	1	740	0,876	0,124
Agriculture	1	1	4	81	820	0	0	906	0,905	0,095
Nom-vegetated areas	0	0	0	1	1	28	1	31	0,902	0,098
Water	0	0	1	0	1	0	48	49	0,967	0,033
Total	124	59	36	744	884	40	49	1936		
AP	0,886	0,922	0,605	0,871	0,929	0,683	0,969	AG	0,893	
EO	0,114	0,078	0,395	0,129	0,071	0,317	0,031	DG	0,107	

Figure 4. Mean Confusion Matrix (2017-2024) for collection 3 using LAPIG validation dataset.

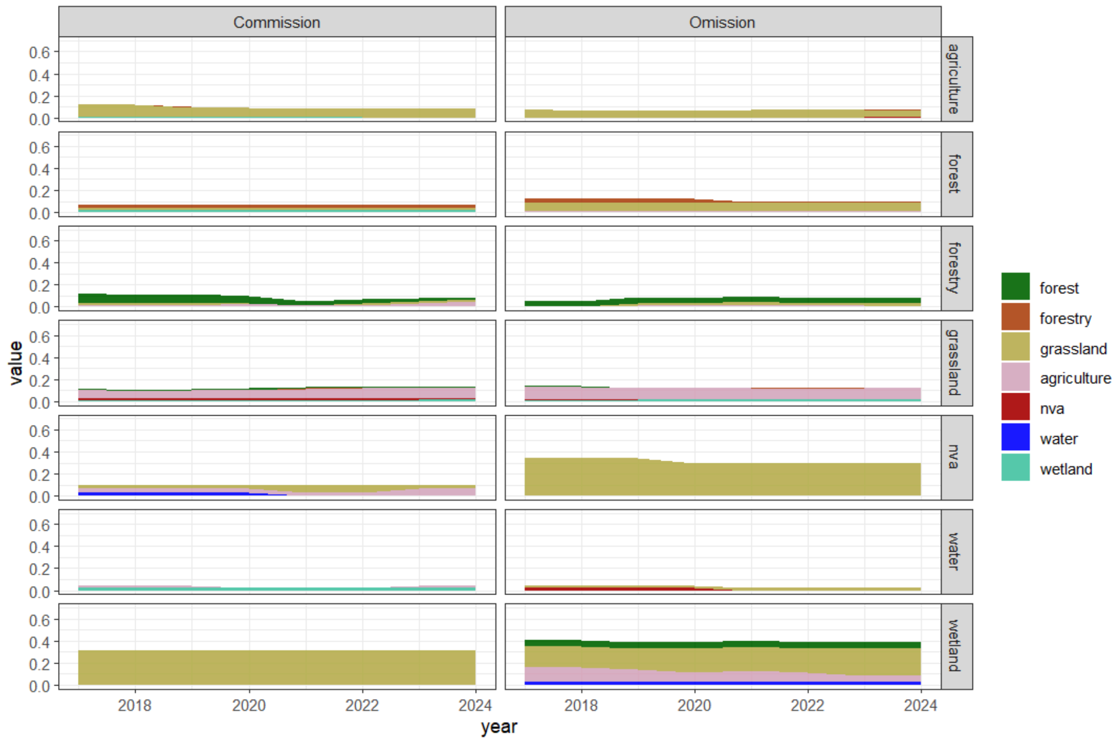


Figure 5. Commission and omission errors of the classes mapped in collection 3 according to LAPIG dataset.

4.2 SH-22-Y-D Reference Chart classification

For the SH-22-Y-D chart, MapBiomas provides a LULC classification map for the year 2020, produced through visual interpretation of Planet imagery. Its spatial resolution is 10 meters and contains 12 classes, including: Forest (3); Forestry (9); Wetland (11); Grassland (12); Temporary Crop (19); Mosaic of Uses (21); Beach, Dune and Sand Spot (23); Urban Area (24); Other Non-Vegetated Areas (25); Rocky Outcrop (29); River, Lake and Ocean (33); Wooded Sandbank Vegetation (49); and Herbaceous Sandbank Vegetation (50). Collection 3 was compared with this dataset, considering the chart map as the reference. The comparison was performed on a pixel-by-pixel basis, computing agreements and disagreements.

Figure 6 presents the confusion matrix for this assessment. The overall accuracy is 82.2%, with 8.88% allocation disagreement and 8.8% quantity disagreement. **Figures 7** and **8** show the omission and commission errors for each one of the mapped classes.

Carta Pelotas

Coleção 3 integração

	3	9	11	12	19	21	23	24	25	29	30	33	49	50	Total	AU	EI
3	85782	9127	281	4636	7236	183	19	4	189	2	6	246	208	0	107918	0,795	0,205
9	645	37357	149	736	252	110	149	9	111	0	0	6	203	1590	41319	0,904	0,096
11	704	248	42447	5170	1473	91	33	5	57	0	0	3110	2504	1809	57653	0,736	0,264
12	5818	1204	12348	105557	38330	1481	413	206	1289	6	33	936	2322	1044	170988	0,617	0,383
19	412	323	1006	7683	171609	4287	64	11	306	1	2	113	96	3166	189079	0,908	0,092
21	3144	294	1382	13018	55734	2647	38	8	1746	10	17	468	85	505	79097	0,033	0,967
23	0	423	189	67	159	23	12206	27	45	0	0	127	13	1830	15109	0,808	0,192
24	18	3	19	401	6	12	2	6612	571	0	2	8	65	12	7731	0,855	0,145
25	36	68	70	640	230	345	1201	271	1464	2	177	132	47	821	5506	0,266	0,734
29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,000	1,000
30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0,000	1,000
33	240	16	749	157	131	15	252	2	21	0	2	543111	35	51	544782	0,997	0,003
49	961	3045	898	2060	547	159	40	27	113	0	0	159	19739	1270	29018	0,680	0,320
50	2	1010	2128	896	1746	102	1267	8	188	0	0	184	287	25554	33373	0,766	0,234
Total	97764	53118	61665	141022	277453	9456	15686	7190	6099	22	232	548601	25606	37654	1281573		
AP	0,877	0,703	0,688	0,749	0,619	0,280	0,778	0,920	0,240	0,010	0,000	0,990	0,771	0,679	AG	0,822	
EO	0,123	0,297	0,312	0,251	0,381	0,720	0,222	0,080	0,760	0,990	1,000	0,010	0,229	0,321	DG	0,178	

Figure 6. Confusion Matrix (2020) for collection 3 using SH-22-Y-D reference map.

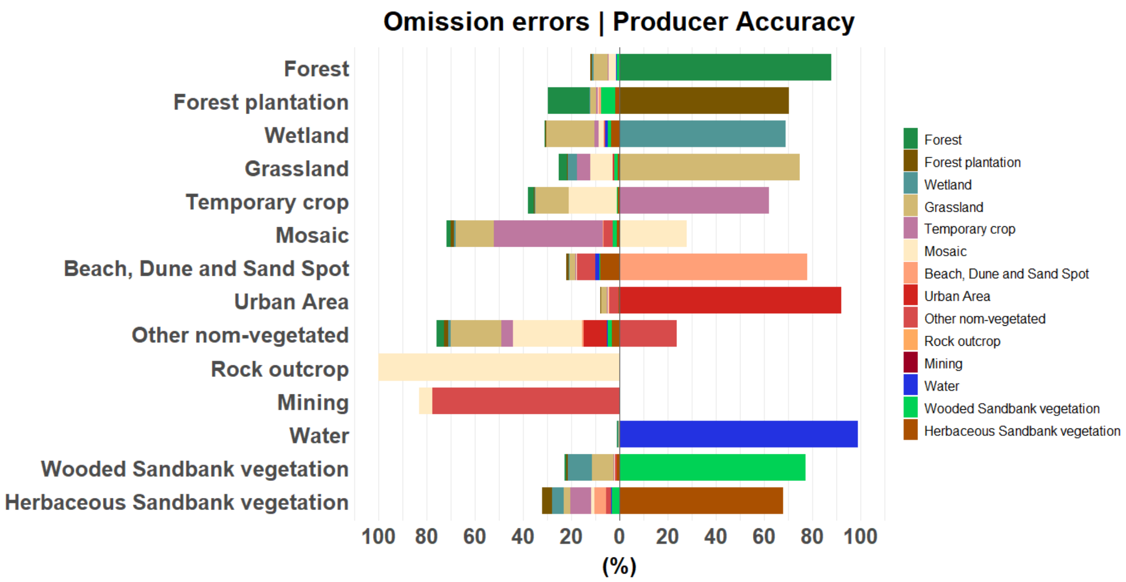


Figure 7. Omission errors of the classes mapped in collection 3 in reference map (SH-22-Y-D).

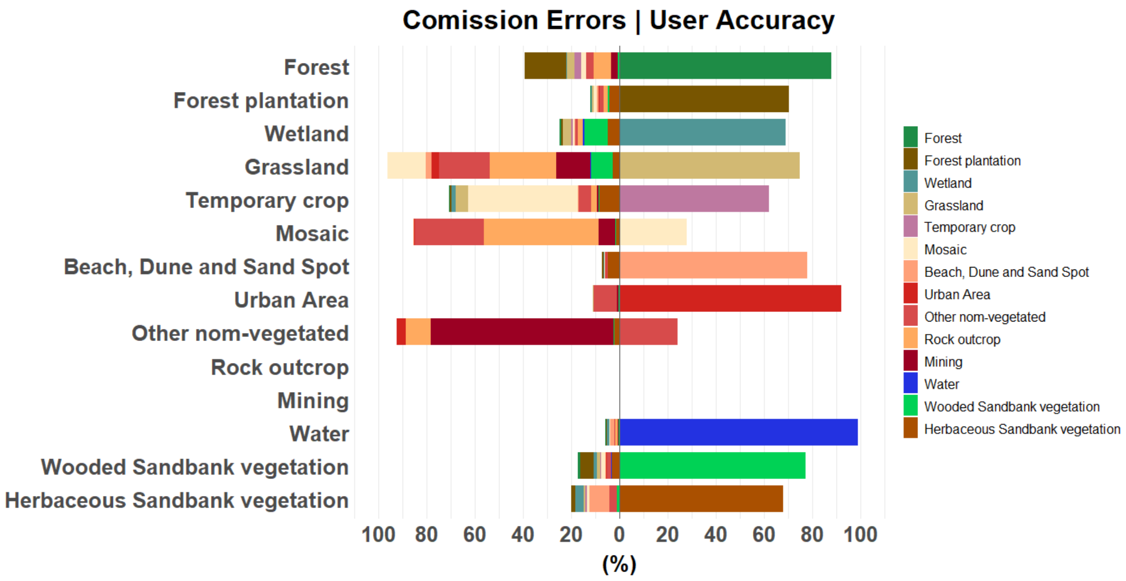


Figure 8. Commission errors of the classes mapped in collection 3 in reference map (SH-22-Y-D).

5. References

AMATULLI, Giuseppe; MARQUEZ, Jaime Garcia; SETHI, Tushar; KIESEL, Jens; GRIGOROPOULOU, Afroditi; ÜBLACKER, Maria M.; SHEN, Longzhu Q.; DOMISCH, Sami. Hydrography90m: a new high-resolution global hydrographic dataset. Earth System Science Data, [S.L.], v. 14, n. 10, p. 4525-4550, 17 out. 2022. Copernicus GmbH. http://dx.doi.org/10.5194/essd-14-4525-2022.

BAEZA, S.; VÉLEZ-MARTIN, E.; ABELLEYRA, D. de; BANCHERO, S.; GALLEGO, F.; SCHIRMBECK, J.; VERON, S.; VALLEJOS, M.; WEBER, E.; OYARZABAL, M.. Two decades of land cover mapping in the Río de la Plata grassland region: the mapbiomas pampa initiative. Remote Sensing Applications: Society and Environment, [S.L.], v. 28, p. 100834, nov. 2022. Elsevier BV. http://dx.doi.org/10.1016/j.rsase.2022.100834.

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

BROWN, Christopher F. et al. AlphaEarth Foundations: An embedding field model for accurate and efficient global mapping from sparse label data. Google Deepmind, California, v. , n. , p. 1-63, out. 2025.

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

GOOGLE EARTH ENGINE, Satellite Embedding V1. Disponível em: https://developers.google.com/earth-engine/datasets/catalog/GOOGLE_SATELLITE_EMBEDDING_V1_ANNUAL. Acesso em: 11 fev. 2026.

GOOGLE EARTH ENGINE, Harmonized Sentinel-2 MSI: MultiSpectral Instrument, Level-2A (SR). Disponível em: https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_SR_HARMONIZED?hl=pt-br. Acesso em: 11 fev. 2026.

HASENACK, Heinrich; WEBER, Eliseu José; BOLDRINI, Ilsi Iob; TREVISAN, Rafael; FLORES, Carlos Alberto; DEWES, Homero. Biophysical delineation of grassland ecological systems in the State of Rio Grande do Sul, Southern Brazil. Iheringia, Série Botânica., [S. l.], v. 78, 2023. DOI: 10.21826/2446-82312023v78e2023001. Disponível em: https://isb.emnuvens.com.br/iheringia/article/view/1296. Acesso em: 11 fev. 2026.

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

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

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

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

PONTIUS, Robert Gilmore; KRITHIVASAN, Roopa; SAULS, Laura; YAN, Yan; ZHANG, Yujia. Methods to summarize change among land categories across time intervals. Journal Of Land Use Science, [S.L.], v. 12, n. 4, p. 218-230, 30 jun. 2017. Informa UK Limited. http://dx.doi.org/10.1080/1747423x.2017.1338768.

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

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