

Atlantic Forest - Appendix

MapBiomass 10m - Collection 3 (beta)

Version 1

General coordinator

Natalia Crusco

Luis Guedes Pinto

Team

Fernando Frizeira Paternost

Yasmin Kloosterman Gelli

Marcos Reis Rosa

Jacqueline Freitas

MapBiomass Brasil - Land Use and Land Cover

May, 2026

1. Overview

The annual land use and land cover maps of the Atlantic Forest biome, corresponding to MapBiomias Collection 3 (beta), were produced from annual mosaics of Sentinel-2 satellite imagery. Table 1 shows the evolution of the collections produced within the project. From Collection 3 onwards, geospatial embeddings developed by Google were also used. The AlphaEarth Foundations annual embeddings encode temporal landscape trajectories in a 64-dimensional representation per 10-m pixel, integrating multiple data sources such as optical, radar, and LiDAR sensors. The 64 embedding bands were added to the Sentinel-2 mosaics and used as input variables in the classification process.

Table 1. Methodological evolution of the Atlantic Forest mapping collections in the MapBiomias 10m.

Collection (Beta)	Period	Mapped classes	Method	Key improvements
1.0	2016-2022	Forest Formation, Savanna Formation, Grassland, Mosaic of Uses, Non-vegetated areas, Rock Outcrop, Forest Plantation, and Water	Random Forest	First version of collection in the MapBiomias 10m initiative.
2.0	2016-2023	Forest Formation, Savanna Formation, Grassland, Mosaic of Uses, Non-vegetated areas, Rock Outcrop, Forest Plantation, and Water	Random Forest	Random Forest multiprobability approach; Improvement of post-classification filters;
3.0	2017-2024	Forest Formation, Savanna Formation, Grassland, Herbaceous Sandbank Vegetation, Mosaic of Uses, Non-Vegetated Areas, Rocky Outcrop, Forest Plantation, and Water	Random Forest	New approach for Herbaceous Sandbank Vegetation; Introduction of Satellite Embedding dataset;

In order to reduce spectral confusion among classes and improve sample balance, the classification was carried out in a regionalized manner. The Atlantic Forest biome was subdivided into 30 homogeneous regions, defined based on native vegetation types described by IBGE (2021). Given the high diversity of

phytophysionomies within the biome, the definition of forest formations primarily considered characteristics related to canopy cover and canopy height.

The classification of the annual mosaics, covering the period from 2017 to 2024, was performed using the Random Forest algorithm. The year 2016 was not included due to the unavailability of Google embeddings for that period. The mapped classes include Forest Formation, Savanna Formation, Grassland, Herbaceous Sandbank Vegetation, Mosaic of Uses, Non-Vegetated Areas, Rocky Outcrop, Forest Plantation, and Water. Additionally, in specific regions of southeastern and northeastern Brazil, a complementary classification of the Agriculture class was conducted to reduce omission errors and confusion with natural areas. Subsequently, the Agriculture and Forest Plantation classes were converted into the Mosaic of Uses class in the final biome-level product.

The feature space used for classification consisted of a subset of the 20 most relevant variables for each region, selected from an initial set of 203 variables. These variables included original Sentinel-2 spectral bands, vegetation indices, variables derived from spectral mixture modeling, topographic attributes, and Google embeddings. Variable selection was based on a feature importance analysis generated by the Random Forest algorithm, considering 500 iterations.

The Herbaceous Sandbank Vegetation class, which in previous collections had been assigned solely through post-classification procedures, became part of the full classification workflow starting in Collection 3, from the definition of stable samples to the application of final filters. Nevertheless, the IBGE soil map continued to be used as a spatial constraint after classification to prevent the occurrence of this class in areas where it is not naturally expected.

After the individual classification of each region and year, the regional results were integrated into a single biome-wide mosaic, with overlapping boundaries between regions to avoid spatial discontinuities. The priority order for integration was visually defined based on classifications from previous collections and the results of the current collection. Subsequently, spatial and temporal post-classification filters were applied to reduce noise and stabilize the time series. These filters included gap-filling procedures, spatial filters based on a minimum connectivity of 25 pixels (approximately 0.25 ha), noise corrections for the first and last years of the series, mode filters, three-year temporal windows, corrections for spurious regeneration, minimum transition area thresholds between natural and anthropogenic classes, and specific stabilization filters for forest areas. Exceptions to the spatial filter were applied to Forest and Wetland areas located in riparian zones, based on the Height Above Nearest Drainage (HAND) product.

Finally, the Atlantic Forest classification was integrated with data produced by the MapBiomias cross-cutting themes, following predefined prevalence rules. The final product consists of a continuous historical series of annual land use and land cover maps for the Atlantic Forest biome, covering the period from 2017 to 2024, with 21 classes at level 3 of the legend. All codes used are available on our public GitHub (<https://github.com/mapbiomas/brazil-atlantic-forest>).

2. Image Pre-processing:

The annual mosaics were generated from Sentinel-2 image collection ("COPERNICUS/S2_SR_HARMONIZED"), using its original bands, blue, green, red, red-edge 1, red-edge 2, red-edge 3, red-edge 4, near-infrared (NIR), shortwave infrared 1 (SWIR1), and shortwave infrared 2 (SWIR2) and several spectral indexes were also generated and appended as bands in the mosaic.

In addition to the spectral bands, embeddings developed by Google ("GOOGLE/SATELLITE_EMBEDDING/V1/ANNUAL") were used for Collection 3. The Google Satellite Embedding is a global dataset, analysis-ready collection of learned geospatial embeddings. It encodes temporal trajectories of surface representation within its 64-dimensional representation for each 10-meters pixel and assimilates multiple datastreams including optical, radar, LiDAR, and other sources. The 64 bands were added in the Sentinel-2 mosaic and used as input to the classification process, which were made in the same way as the previous collections.

3. Definition of regions for classification

The classification was done in homogenous regions to reduce confusion of samples and classes, as well as to allow a better balance of samples and results. The Atlantic Forest biome was divided into 30 regions based in native vegetation types in the Atlantic Forest biome (IBGE, 2021) (Figure 1).

Due to great diversity of phytophysiognomies in Atlantic Forest, different characteristics are considered in mapping the forest formation within the biome, primarily related to canopy cover and height.

Forest Formation includes natural forest (exclude Forest Plantation) areas of more than 0.5 hectares (ha) with trees with a minimum height of 5 meters (m) and tree canopy cover that varies for each type of original forest formation:

- Dense Ombrophiles Forest - tree crown cover of more than 80%
- Mixed Ombrophiles Forest- tree crown cover of more than 80%
- Open Ombrophiles Forest - tree crown cover of more than 60%
- Seasonal Deciduous Forest- tree crown cover of more than 60%

- Seasonal Semideciduous Forest- tree crown cover of more than 60%

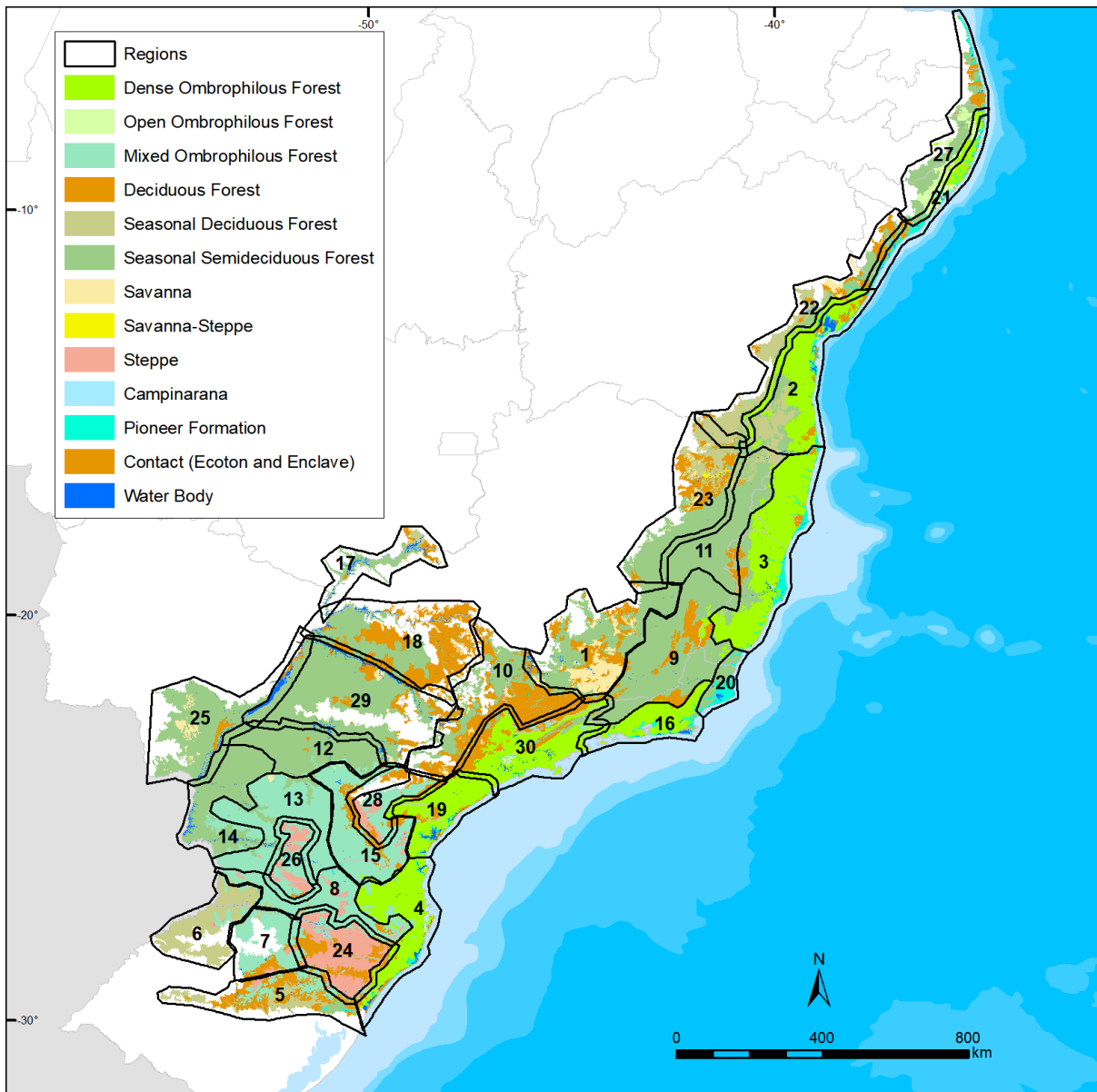


Figure 1. Regions used in the classification of Atlantic Forest biome. Each black polygon represents a classification region, while numbers within each polygon indicate the region ID.

4. General Map Classification Algorithm

4.1 Classification scheme

The MapBiomas 10 m Collection 3 (beta) LCLU dataset for the Atlantic Forest biome classified Sentinel-2 annual mosaics from 2017 to 2024 using the Random Forest algorithm (Figure 2). For this Collection, the year 2016 wasn't included because it is not available on the Google Satellite Embedding dataset.

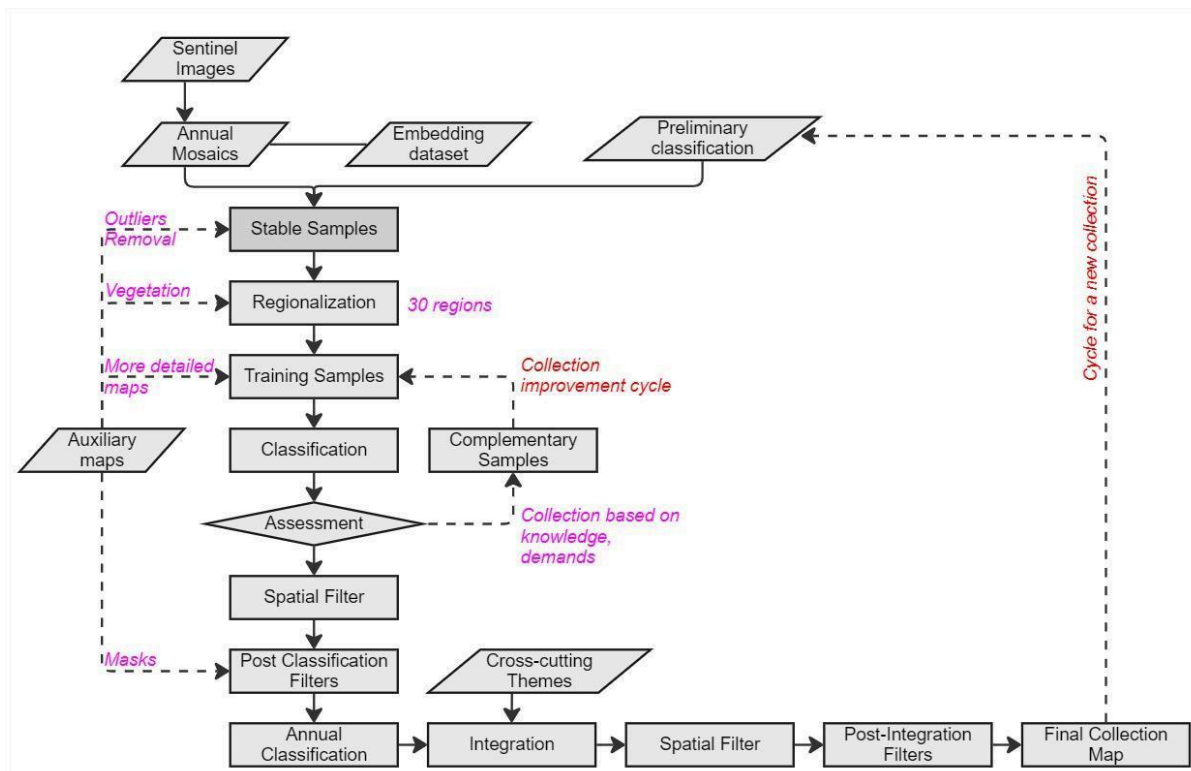

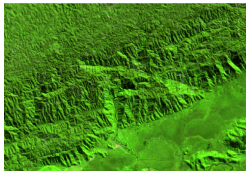

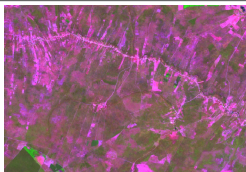

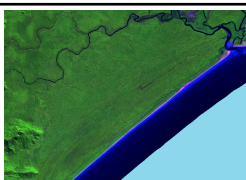

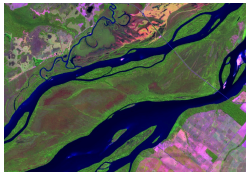

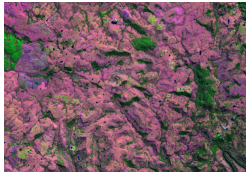

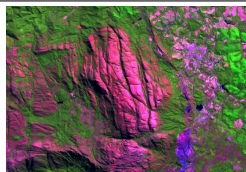









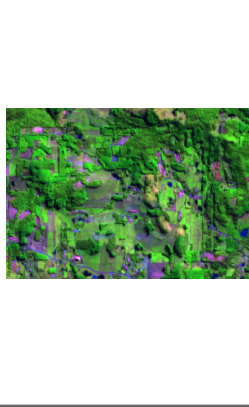

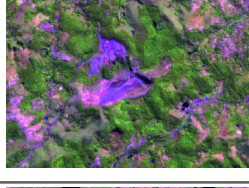

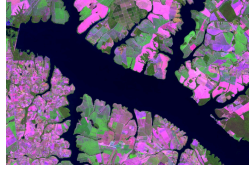
Figure 2. Classification process of MapBiomias 10 m Collection 3 (beta) in the Atlantic Forest biome. Inclined gray rectangles represents databases, while linear rectangles point key-steps in the workflow. Solid arrows indicate main flow, and dashed arrows point to assessment and evaluation cycles. Pink and red text shows additional information for specific steps.

The LCLU classes mapped with Random Forest classification in the Atlantic Forest were: Forest Formation, Savanna Formation, Grassland, Herbaceous Sandbank Vegetation, Mosaic of Uses, Non-vegetated areas, Rock Outcrop, Forest Plantation, and Water. Additionally, an extra classification for the Agriculture class was carried out in specific regions of the biome: Southeast and Northeast (table 2).

With the defined regions, samples were collected in agricultural areas and then classified following the same process as the other classes using Random Forest. The Agriculture and Forest Plantation classification aimed to reduce potential errors in the classification of natural areas and minimize omissions. Later, these two classes were converted to class 21 (Mosaic of uses) in the final Atlantic Forest dataset.

Table 2. Land use land cover categories used for the Sentinel mosaics classification for the Atlantic Forest biome in MapBiomas Collection 3.

Classes Level 1	Classes Level 2	ID	Color	RGB composite (SWIR1-NIR-Red)	Description
Forest	Forest Formation	3			Vegetation types characterized by the predominance of tree species, with high tree density, closed canopy, and vertical stratification. It includes forest typologies: Dense, Open and Mixed Ombrophilous Forest, Semi-deciduous and Deciduous Seasonal Forest, and Pioneer Formation.
	Savanna Formation	4			Vegetation type characterized by the presence of sparse tree and shrub species, with a semi-continuous canopy. It includes: Steppe, Forested and Wooded Savannah.
	Wooded Sandbank Vegetation	49			Forest formations on sandy soils in the coastal region
Herbaceous and Shrubby Vegetation	Wetland	11			Floodplain or grassland vegetation influenced by fluvial and/or lacustrine dynamics, characterized by the predominance of hygrophilous vegetation, including emergent, submerged, or floating aquatic plants.
	Grassland	12			Vegetation dominated by herbaceous species and grasses, with few scattered trees and shrubs, generally featuring an open or absent canopy. It occurs on soils ranging from deep to shallow, including rocky terrains (rupestrian grasslands). Included: Park and Grassland Steppe Savannas, Steppe and Shrub and Herbaceous Pioneers.
	Rocky Outcrop	29			Naturally exposed rocks without soil cover, often with the partial presence of rupicolous vegetation and high slope.
	Herbaceous Sandbank Vegetation	50			Herbaceous vegetation that is established on sandy soils or on dunes in the coastal zone.

Classes Level 1	Classes Level 2	ID	Color	RGB composite (SWIR1-NIR-Red)	Description
Farming	Other Temporary Crops*	19			Areas occupied with short or medium-term agricultural crops, generally with a vegetative cycle of less than one year, which after harvesting need to be planted again to produce.
	Forest Plantation*	9			Tree species planted for commercial purposes (e.g. pinus, eucalyptus, araucaria).
	Mosaic of Uses	21			Areas intended for agricultural and livestock use where it was not possible to distinguish between pastures and croplands, including fallow lands. These areas may also include peri-urban zones such as small farms, rural properties, and residential estates. Transitional areas are also included, where secondary vegetation is developing in abandoned pastures or in ecological restoration processes, prior to reaching forest size and structure.
Non Vegetated Area	Other non Vegetated Areas	25			Natural areas with exposed soil resulting from climatic events (landslides, flooding) and areas with non-permeable surfaces (infrastructure, urban expansion, or mining) not mapped within their respective classes.
Water	River, Lake and Ocean	33			Rivers, lakes, dams, reservoir and other water bodies

*Exceptionally, in regions 01, 03, 09, 10, 11, 16, 19, 20, 21, 27 and 30 we also included the class “3.2.1. Temporary Crop” (ID: 19) and in regions 01, 03, 08, 10, 13, 15, 23, 24, 28 and 30 we also included the class “3.3 Forest Plantations” (ID: 9). These two classes are shared with the agriculture team to pass through the specific filters and are then converted to 21 in the final Atlantic Forest dataset.

Maintaining the procedure adopted in the previous collection, stable samples derived from Sentinel data were used in the classification process. In addition, Global Forest Canopy Height (GFCH) data based on GEDI were employed to filter stable areas for two classes: **Forest Formation** (canopy height ≥ 8 m) and **Grassland** (canopy height ≤ 7 m). Reference maps from the states of São Paulo (IF, 2020), Minas Gerais (IEF), Paraná (IAT, 2020; <https://geopr.iat.pr.gov.br/portal/apps/dashboards/1eca83bf72e44193ae62f282574d>

a52f), and Espírito Santo (IEMA, 2015; <https://geobases.es.gov.br/links-para-mapas1215>) were also used to filter stable samples in natural areas. All reference maps are available on the MapBiomias Brazil website (<https://brasil.mapbiomas.org/en/mapas-de-referencia/>).

The feature space used to classify the Atlantic Forest biome comprised a subset of the 20 most relevant bands for each region, out of a group of 203 variables (table 3). These variables include original Sentinel reflectance bands, vegetation indices, spectral mixture modeling-derived variables, and terrain morphometry (e.g., slope). Additionally, bands from Google Satellite Embedding dataset were used. The definition of the subset was made based on a feature importance analysis produced with Random Forest classification using all bands and 500 interactions. All codes used are available on our public GitHub (<https://github.com/mapbiomas/brazil-atlantic-forest>).

Table 3. Feature space subset with the most frequent variables between regions considered in the classification of the Atlantic Forest biome Sentinel image mosaics in the MapBiomias 10 m Collection 3 (2017-2024).

Type	Description	Formula	Reference
Embeddings Bands	Bands A00 to A63	Satellite Embedding V1	
Sentinel Bands	Bands Blue, Green, Red, NIR, SWIR1, SWIR2	Original reflectance bands	
	Ratio Vegetation Index (RVI)	$(N * R) / (G ** 2.0)$	Jordan (1969)
	Ratio (RATIO)	N/R	Pearson and Miller (1972)
	Normalized Difference Water Index (NDWI)	$(G - N) / (G + N)$	McFeeters (1996)
	Automated Water Extraction Index (AWEI)	$B + 2.5 * G - 1.5 * (N + S1) - 0.25 * S2$	Feyisa et al. (2014)
Spectral Index	Inverse Intensity Index (IIA)	R/N	
	Enhanced Vegetation Index (EVI)	$2.5 * (N + R) / (N + 6R - 7.5B + 1)$	Huete et al. (1994)
	Green Chlorophyll Vegetation Index (GCVI)	$(N/G) - 1$	Gitelson et al. (2005)
	Global Environmental Monitoring Index (GEMI)	$(2 * (N^2 - R^2) + 1.5N + 0.5R) / (N + R + 0.5)$	Pinty and Verstraete (1992)

Chlorophyll Vegetation Index (CVI)	$(N \times R) / (G^2)$	Vincini et al. (2008)
Green Leaf Index (GLI)	$(2G - R - B)/(2G + R + B)$	Lourenço et al. (2021)
Advanced Vegetation Index (AVI)	$(N \times (1 - R) \times (N - R))/ 1000$	Loi et al. (2017)
Bare Soil Index (BSI)	$((S1 + R) - (N + B))/((S1 + R) + (N + B))$	Rikimaru et al. (2002)
Broadband Reflectance-Based Albedo (BRBA)		Liang (2001)
Dynamic Surface Water Index 5 (DSWI5)	$G / (N + S1 + R)$	Fisher et al. (2016)
Land Surface Water Index (LSWI)	$(N - S1)/(N + S1)$	Xiao et al. (2002)
Modified Bare Soil Index (MBI)	$((S1 - S2 - N)/(S1 + S2 + N)) + 0.5$	Nguyen et al. (2021)
Urban Index (UI)	$(S2 - N) / (S2 + N)$	Kawamura et al. (1997)
Optimized Soil Adjusted Vegetation Index (OSAVI)	$(N - R) / (N + R + 0.16)$	Rondeaux et al. (1996)
Redness Index (RI)	$(R - G)/(R + G)$	Mathieu et al. (1998)
Global Vegetation Moisture Index (GVMI)	$((N + 0.1) - (S1 + 0.02)) / ((N + 0.1) + (S1 + 0.02))$	Ceccato et al. (2002)
CONTRAST NIR bands (NIR_CONTRAST)	1/14 GLCM metrics proposed by Haralick, Textural Features for Image Classification	Haralick et al. 1973
CONTRAST RED bands (RED_CONTRAST)	1/14 GLCM metrics proposed by Haralick, Textural Features for Image Classification	Haralick et al. 1973
Normalized Difference Drought Index (NDDI)	$(NDVI - NDWI) / (NDVI + NDWI)$	Gu et al. (2007)
Normalized Difference Vegetation Index (NDVI)	$(N - R) / (N + R)$	Rouse et al. (1974)

Coords	Latitude and Longitude	Latitude, Longitude
--------	------------------------	---------------------

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

4.2 Classification Process of *Herbaceous Sandbank Vegetation Class*

The Herbaceous Sandbank Vegetation class, since the first Sentinel Collection, when it began to be mapped, was classified only during the post-classification process. At that time, only areas originally classified as “Other Non-Forest Natural Formation” (ID13) and located within predefined regions based on the IBGE Soil Map (ESPODOSOL and NEOSOL) were converted to class 50 (Herbaceous Sandbank Vegetation).

Starting with Collection 3, this class began to follow the full classification workflow applied to other natural classes in the biome — from the definition of stable samples to the application of post-classification filters. Nevertheless, the IBGE Soil Map continued to be used as a spatial constraint after the classification to prevent this class from appearing in regions where it is not naturally present.

4.3. Wetland Classification

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

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

From Collection 3, the use of Embeddings made unnecessary a distinct classification for wetlands. This class was included in the main classification chain.

4.4. Final classification

Final classification was performed for all regions and years with stable and complementary samples. All years used the same subset of samples, which was trained in the same mosaic as the year that was classified.

At this stage, the classifications of the 30 regions within the biome are merged. The boundaries between regions are overlapped to prevent gaps in classification along their edges. The priority order for merging the regions is defined visually, based on the classifications from previous collections combined with the classification generated for the current collection.

4.4.1. Post Classification

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

- **Gap-Fill Filter:** No-data values (gaps) are theoretically not allowed and were replaced by the temporally nearest valid classification. In this procedure, if no “future” valid position was available, the no-data value is replaced by its previous valid class. Therefore, gaps should only exist if a given pixel has been permanently classified as no-data throughout the entire temporal domain.
- **Spatial Filter:** The spatial filter avoids unwanted modifications to the edges of pixel groups (blobs). A spatial filter was built based on the "connectedPixelCount" function. Native to the GEE platform, this function locates connected components (neighbors) that share the same pixel value. Thus, only pixels that do not share connections to a predefined number of identical neighbors are considered isolated. In this filter, at least 25 connected pixels are needed to reach the minimum connection value. Consequently, the minimum mapping unit is directly affected by the spatial filter applied, which was defined as 25 pixels (~0.25 ha). For Collection 3, an exception was applied aiming to keep small areas of forest and wetland that were classified in Permanent Protected Areas along rivers. Using Height Above Nearest Drainage (HAND) dataset, the spatial filter was not applied to these classes, (3 and 11) where they were overlapped by the HAND layer.

- Temporal Filters: Some temporal filters were applied to stabilize the classification.
 - First and last year noises correction: This filter removes small deforestation and regeneration from the first and last years, assuming that little gain or losses of vegetation are classification noises.
 - Mode filter: It stabilizes natural classes by avoiding a pixel to be classified as different natural classes during the series. It uses a statistical mode approach to reduce "flickering" between different types of natural vegetation and define to each class the pixel should belong.
 - Three years window: This filter analyses a window of three years and changes the middle year when it is: different from a class and when the year before and the year after are equal to this class. The central year is modified based on the classification assigned to that pixel in the previous year.
 - Regeneration correction: It removes regeneration events from stable pixels that show a change just in the first or last year.
 - Minimum transition area: This is a spatial-temporal filter that sets a minimum transition area of 0,5 ha to allow transition between natural classes and Mosaic of Uses (21).
 - Stabilization filter: This filter counts how many years a pixel was assigned to that class. If the class appears for a majority of the time (e.g., 4 or more years), it is considered "stable" and classifies that pixel as that class for all years. Forest (3) class is not considered in this filter.
 - Forest Stabilization filter: This filter counts how many times a pixel classified as forest changes over time and how connected it is with other forest pixels. It aims to stabilize temporally and remove noises.






















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

5. Integration with cross-cutting themes

The final classification result, including the application of post-classification filters generated by the Atlantic Forest team, is integrated with the data produced by the cross-cutting themes for all years in the historical series (2017–2024). This step is carried out based on predefined prevalence rules (Table 3). These rules

determine which classes take priority in cases of overlap, ensuring consistent decision-making across biomes and over time. The final integrated map for the Atlantic Forest biome consists of 21 classes at level 3 of the legend (Figure 3).

Table 3. General prevalence rules - MapBiomias Collection 3

Class	ID	Prevalence order	Color
Photovoltaic power plant	75	1	
Mining	30	2	
Beach, Dune and Sand Spot	23	3	
Mangrove	5	4	
Aquaculture	31	5	
Hypersaline Tidal Flat	32	6	
Urban Infrastructure	24	7	
Forest Plantation	9	8	
Rocky Outcrop	29	9	
Temporary Crops	19	10	
Perennial Crops	36	11	
Herbaceous Sandbank Vegetation	50	12	
River, Lake and Ocean	33	13	
Other non Vegetated Areas	25	14	
Forest Formation	3	15	
Savanna Formation	4	16	
Wooded Sandbank Vegetation	49	17	
Wetland	11	18	
Grassland Formation	12	19	
Pasture	15	20	
Mosaic of Uses	21	21	

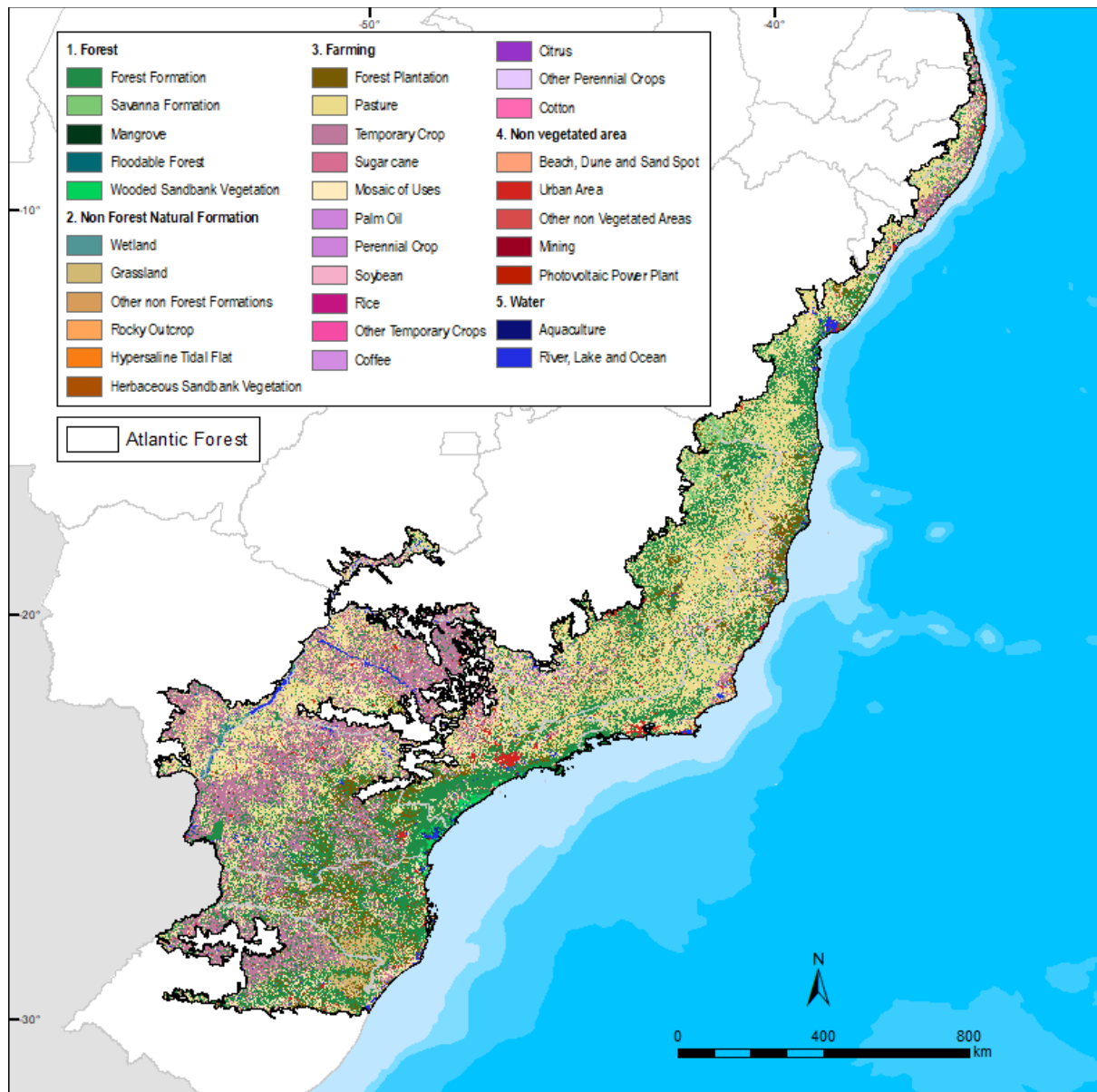


Figure 3. Final land use and land cover map of the Atlantic Forest biome - 2024

6. References

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

IBGE. Vegetação RADAM. Disponível em:

<ftp://geoftp.ibge.gov.br/informacoes_ambientais/acervo_radambrasil/vetores/>. Accessed in: may, 30 2018.

P. Potapov, X. Li, A. Hernandez-Serna, A. Tyukavina, M.C. Hansen, A. Kommareddy, A. Pickens, S. Turubanova, H. Tang, C.E. Silva, J. Armston, R. Dubayah, J. B. Blair, M. Hofton (2020) Mapping and monitoring global forest canopy height through integration of GEDI

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).