Amazon — Appendix

Collection 4 (beta) • Version 1

MapBiomas Brazil • Sentinel (10m) • Land use and Land Cover Map

Algorithm Theoretical Basis Document (ATBD)

General Coordinator

Bruno Ferreira

Team

João Victor Siqueira

Larissa Amorim, Msc

Camila Damasceno

Ives Brandão

Raissa Ferreira

Manoela Athaide

August, 2026

1. OVERVIEW

The Amazon is Brazil's largest biome, characterized by ecological complexity that encompasses the world's largest river basin and unparalleled biodiversity. However, its vast extent and density make terrestrial monitoring a logistical and technical challenge. Land Use and Land Cover (LULC) mapping is a strategic tool that transforms raw satellite data into information on the biome's conservation status. By analyzing historical map series, it is possible to accurately estimate deforestation rates, identify areas of vegetation regeneration, and monitor water dynamics, which is essential for understanding flood cycles and regional water availability.
For this reason, the MapBiomas initiative produces annual LULC maps, initially at 30m scale and now at 10m resolution, enabling direct time-series analysis to distinguish between natural seasonal variations and permanent anthropogenic changes.

Amazon mapping within the MapBiomas 10m initiative is processed entirely in Google Earth Engine (GEE), using high spatial resolution data. The workflow focuses on the classification of seven main classes: Forest Formation, Savanna Formation, Grassland, Pasture, Agriculture, Water, and Non-Vegetated Areas.

The Collection 1.0 (2016–2022) established the project foundation by using stable areas from Collection 7.1 (Landsat) as reference data for algorithm training. The feature space was structured from Sentinel-2 mosaics, integrating spectral bands, vegetation indices, and spectral mixture fractions. Map refinement involved post-classification spatial and temporal filters, along with integration of cross-cutting themes such as Wetlands and Rocky Outcrops.

The Collection 2.0 (2016–2023) continued this approach, extending the time series with the inclusion of 2023 data.

For Collection 3 (2017–2024), the methodological protocol was maintained, with updates to the feature space through the inclusion of Google embeddings bands. This version also incorporated corrections in previously identified specific areas, optimizing the consistency of the LULC data for the 2024 series closure.

For Collection 4 (2017–2025), the same methodology used in Collection 3 was retained, with the sole addition of the year 2025.

All processing scripts and algorithms used in the Collection 4 Sentinel classification are publicly available and can be accessed via the official MapBiomas Amazon repository:

https://github.com/mapbiomas/brazil-amazon

Table 1. Evolution of Amazon mapping collections in the MapBiomas (10m).

Collection	Period	Mapped Classes	Method / Mapping Unit
1.0	7 Years 2016–2022	Forest Formation, Savanna Formation, Grassland, Pasture, Agriculture, Other Non-Vegetated Areas and Water	Random Forest / Annual Sentinel Mosaic

Collection	Period	Mapped Classes	Method / Mapping Unit
2.0	8 Years 2016–2023	Forest Formation, Savanna Formation, Grassland, Pasture, Agriculture, Other Non-Vegetated Areas and Water	Random Forest / Annual Sentinel Mosaic
3.0	8 Years 2017–2024	Forest Formation, Savanna Formation, Grassland, Pasture, Agriculture, Other Non-Vegetated Areas and Water	Random Forest / Annual Sentinel Mosaic + Satellite Embedding
4.0	9 Years 2017–2025	Forest Formation, Savanna Formation, Grassland, Mosaic of Uses, Other Non-Vegetated Areas and Water	Random Forest / Annual Sentinel Mosaic + Satellite Embedding

2. METHODOLOGICAL DESCRIPTION — COLLECTION 4

Each step of the Collection 4 workflow is detailed below, covering mosaic construction and feature space, sampling and classification strategy, pre-integration steps, and post-classification filters applied to the Amazon biome (Figure 1).

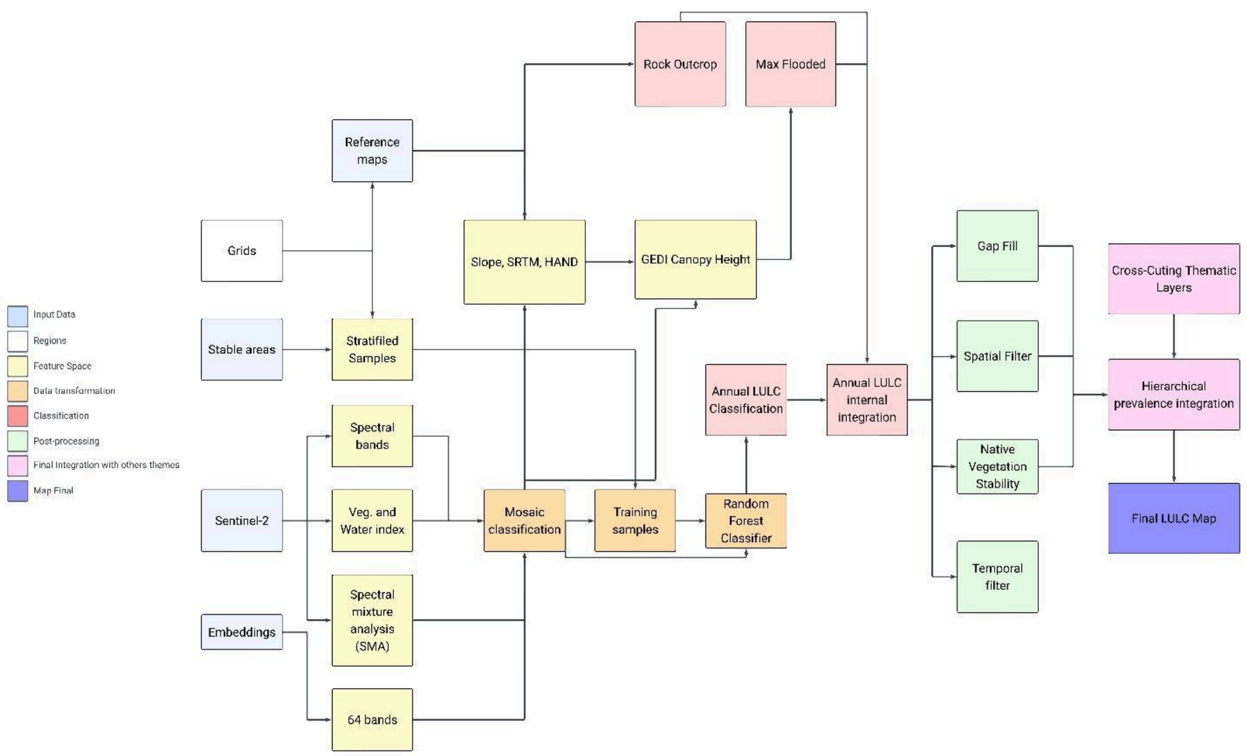

Figure 1. General workflow of the MapBiomas 10m classification process for the Amazon biome.

2.1. Mosaic Construction and Feature Space

The first step consists of generating annual mosaics using the harmonized surface reflectance (SR) collection from **Sentinel-2A e 2B**.

- **Spectral Composition, Metrics and Fractions:** The feature space used for Amazon biome classification comprises a comprehensive set of metrics derived from Sentinel-2 mosaics, organized to capture both spectral response and the temporal dynamics of land cover. As detailed in Table 2, this dataset includes the native spectral bands Blue, Green, Red, Near Infra-Red (NIR), Short Wave Infra-Red (SWIR) 1, and SWIR 2, for which annual medians, dry and wet season-specific medians, and standard deviation are calculated. This seasonal stratification is fundamental for distinguishing targets with distinct hydrological behaviors across the biome.

  In addition to native bands, the workflow integrates the vegetation and water indices Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI), and Enhanced Vegetation Index 2 (EVI2), supplementing these variables with the annual amplitude metric, which enables identification of extreme variations in vegetation cover such as deforestation or seasonal flooding. The model also incorporates Spectral Mixture Analysis – SMA (Souza et al., 2005) fractions such as GV (Green Vegetation), NPV (Non-Photosynthetic Vegetation) and Soil, which decompose the pixel into its basic biophysical components. Complementing this feature space, derived indices such as Normalized Difference Fraction Index (NDFI) focused on degradation detection (Souza et al., 2013) — also the SEFI (Soil Exposure Fraction Index) (Alencar et al., 2020) and WEFI (Wetland Ecosystem Fraction Index) (Rosa, 2020) index, aimed at wetland identification, are used as well. The combination of these statistical and biophysical metrics provides the Random Forest Classifier (RFC) with a robust temporal signature, enabling precise separation between forest, savanna, grassland formations and the various anthropic land use classes. The predictor variables used in classification, including statistical metrics for native bands, spectral indices and spectral mixture fractions, are detailed in Table 2.

- **Collection 3.0 Innovation:** The main advancement in this phase is the integration of Google's Satellite Embedding bands (*GOOGLE/SATELLITE_EMBEDDING/V1/ANNUAL*) (Brown et al., 2025). These 64 bands encode complex contextual and structural information via *deep learning*, which, combined with biophysical fractions, enhance the classifier's ability to distinguish classes in heterogeneous landscapes.

Table 2. Feature Space for Amazon biome classification.

Band or Index	Metrics				
	Median	Median (Dry Season)	Median (Wet Season)	Standard Deviation	Range
Blue	X	X	X	X	-
Green	X	X	X	X	-
Red	X	X	X	X	-
NIR	X	X	X	X	-
SWIR 1	X	X	X	X	-
SWIR 2	X	X	X	X	-

Band or Index	Metrics				
	Median	Median (Dry Season)	Median (Wet Season)	Standard Deviation	Range
NDVI	X	X	X	X	X
NDWI	X	X	X	X	X
EVI2	X	X	X	X	X
GV	X	-	-	-	X
NPV	X	X	X	X	X
Solo (Soil)	X	X	X	X	X
NDFI	X	X	X	X	X
SEFI	X	X	X	X	X
WEFI	X	X	X	X	X

2.2 Sampling and Classification Strategy

The workflow follows a consistent calibration structure for the RFC:

- **Samples from Stable Areas:** Training samples are generated from stable areas of Collection 10.1 (Landsat), ensuring robust reference data for the seven target classes in each year of the historical series.
- **Regionalized Grid Processing:** To handle the biome's heterogeneity, the Amazon was divided into 295 grids. In each grid, classification is performed independently, allowing fine-tuning of hyperparameters and feature selection according to the ecological and climatic particularities of each regional sector (Figure 2).
- **Stratification and Point Density:** Training points were defined for each grid. The distribution of these points among target classes follows a logic proportional to the area occupied by each class within the grid, based on the reference map. This strategy ensures that dominant classes have statistical representation, while the use of stable areas minimizes the inclusion of noise or spurious transitions in model training.

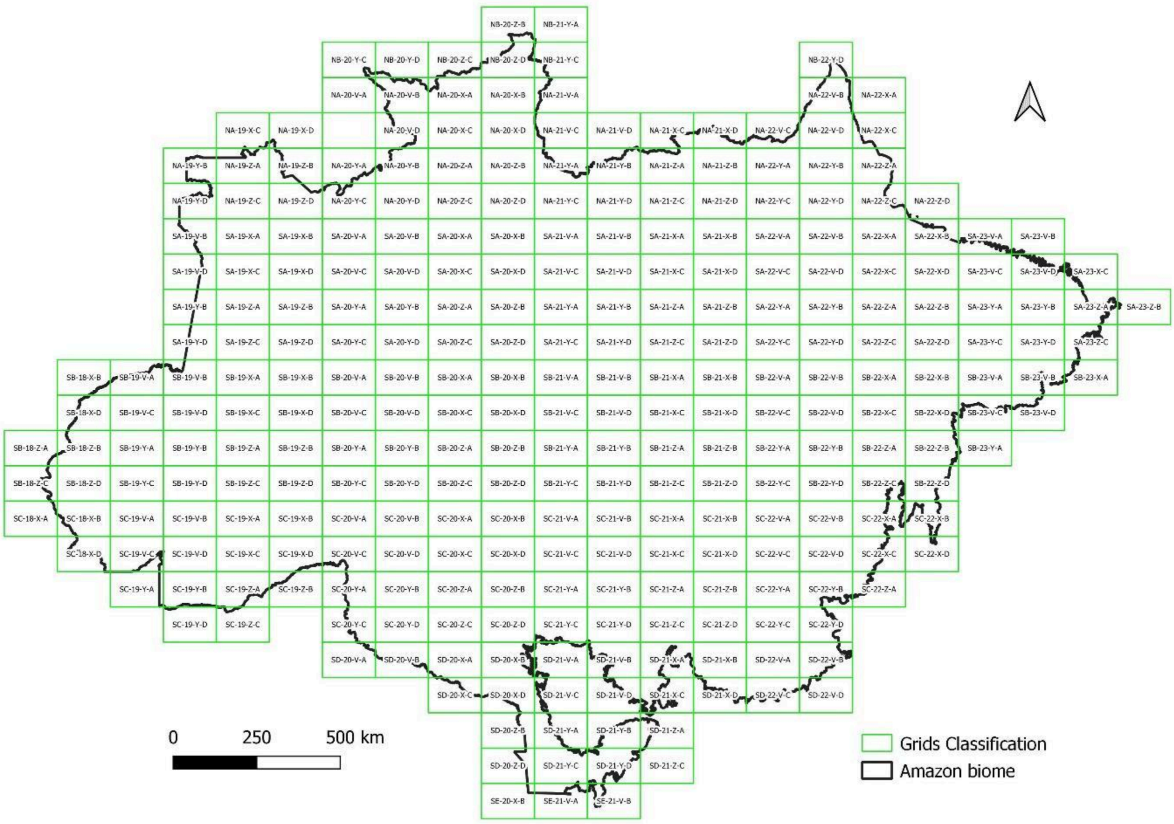

Figure 2. Regionalized processing grid used for Amazon biome classification.

2.3. Classification Algorithm

Digital classification was performed year by year using a variation of the RFC (Breiman, 2001) algorithm, available in Google Earth Engine. The RFC was trained on stable samples using 100 decision trees, defined according to each feature space subset. Each sub-region (grid) has its own sample balancing and classification. After classification processing, the sub-regions were integrated to compose the full territory of the Amazon biome.

2.4. Pre-integration: Wetlands and Rocky Outcrops

Specific themes were mapped separately to reduce commission errors:

- **Wetland and Floodable Forest Mapping**

  For wetland mapping, Sentinel-2 mosaics were used in conjunction with reference maps for sample screening and stratification. The training and calibration of the RFC integrated various structural and topographic datasets, including Global Ecosystem Dynamics Investigation (GEDI), Shuttle Radar Topography Mission (SRTM), Height Above the Nearest Drainage (HAND), Global Canopy Height, and SMA fractions. Reference maps from Gumbricht et al. (2017), Hess et al. (2015), and Tootchi et al. (2018) were used.

  Initially, sampled pixels were automatically classified into a binary map (Wetland and Non-Wetland), serving as the basis for annual mosaic classification. To ensure time-series consistency across the Amazon biome, a maximum value reducer was applied to the annual layers, thereby defining the Maximum Flooded Area (MFA).

Annually, the LULC map is crossed with the MFA layer following the remapping logic below:

- Pixels classified as **Forest Formation** that overlap with the MFA are remapped as **Floodable Forest**.
- Pixels classified as **Savanna Formation** or **Grassland** that overlap with the MFA are remapped as **Wetlands**.

MFA mapping is restricted to Regions of Interest (ROI) defined by training attributes. Due to the resolution difference between some predictors (such as SRTM/HAND) and Sentinel-2, the final mapping in certain areas may appear visually close to 30-meter spatial resolution.

- **Rocky Outcrop Mapping**

  Rocky Outcrop mapping in the Amazon biome used the same annual mosaics, supplemented by random stratified samples (Outcrop and Non-Outcrop) for RFC calibration. The unique Spectro-temporal behavior of these formations is characterized by high elevations, steep slopes, escarpments, and a predominance of exposed rocks and soil.

  To represent these features, spectral mixture model fractions such as Soil, NPV, and GV were used. Terrain morphological characteristics were incorporated using SRTM and HAND data, enabling precise distinction between outcrops and other land cover classes.

2.5. Post-Classification and Filters

After the classification step, a logical sequence of filters is applied to ensure result consistency:

- **Gap Fill:** Fills gaps caused by clouds using the mode of the pixel's class across the time series.
- **Spatial Filter:** Uses the *connectedPixelCount* function (native to GEE) to identify neighboring pixels of the same value. Isolated pixels that do not form groups of minimum size (less than 1 hectare) are reclassified to avoid noise and smooth edges.
- **Native Vegetation Stability:** Pixels that oscillate only among natural classes (Forest, Savanna, and Grassland) are stabilized using the mode, since such transitions are rare without human intervention.
- **3-Year Temporal Filter applied to the first and last year of the time series:** Corrects inconsistent transitions. The persistence rule A-B-A → A-A-A is applied (e.g., Forest-Pasture-Forest becomes Forest-Forest-Forest). Additionally, the rules in Table 3 correct confusion errors where recent patches are classified as Savanna or Grassland following forest vegetation conversion. In the first- and last-year, if a pixel is no observed, the pixel assumes the class of the immediately preceding or following year to avoid spurious fluctuations (Table 3).

Table 3. Temporal filter rules applied in post-classification for the Amazon biome.

Rule	Class Year 1	Class Year 2	Class Year 3	New Class Year 3
1	Forest Formation	Pasture	Grassland	Mosaic of Uses
2	Forest Formation	Pasture	Savanna Formation	Mosaic of Uses
3	Forest Formation	Forest Formation	Grassland	Mosaic of Uses

Rule	Class Year 1	Class Year 2	Class Year 3	New Class Year 3
4	Forest Formation	Forest Formation	Savanna Formation	Mosaic of Uses
5	Floodable Forest	Floodable Forest	Savanna Formation	Mosaic of Uses

In Collection 4.0, specific corrections were implemented to reduce classification errors in the Land Use Mosaic for the state of Roraima. In addition, the time series was updated to include the 2025 map. Figure 4 illustrates the final 2025 LULC map, using Sentinel data with a resolution of 10 meters. The same classes as in the previous collection were mapped. Table 4 presents all the classes mapped for Collection 4.0 in the Amazon biome.

Table 4. Classification scheme of Collection 11 for the Amazon biome (adapted from Souza Jr et al., 2023).

Value	Color	Color Code	Class	Description
3		*#1f8d49*	**Forest Formation**	Vegetation types dominated by tree species with continuous canopy and high density. Also includes mangroves, secondary regeneration, and planted forests.
4		*#7dc975*	**Savanna Formation**	Vegetation types with a tree layer of variable density distributed over a continuous shrubby-herbaceous layer.
6		*#026975*	**Floodable Forest**	Permanently or temporarily flooded forest areas.
11		*#519799*	**Wetland**	Shrublands and natural fields permanently or temporarily covered by water.
12		*#d6bc74*	**Grassland Formation**	Herbaceous vegetation, including patches with a well-developed shrubby-herbaceous layer.
15		*#ffefc3*	**Mosaic of Uses**	Transition class. Farming areas where it was not possible to distinguish pasturelands and croplands. This class may also include areas of secondary vegetation developing on abandoned pasturelands that have not yet reached the size and structure of a forest, as well as small-scale family farming and recent deforested areas.
25		*#db4d4f*	**Other Non-Vegetated Areas**	Impervious surfaces or exposed soil not mapped in other classes.
29		*#ffaa5f*	**Rocky Outcrop**	Naturally exposed rocks on the land surface without soil cover, frequently with partial rupestrian vegetation and steep slopes.
33		*#2532e4*	**River, Lake and Ocean**	Rivers, lakes, dams, reservoirs, and other water bodies.

3. Integration with Cross-Cutting Themes

After applying the post-processing filters, the LULC maps for each of the 9 years (2017–2025) were integrated with the cross-cutting themes using a set of specific hierarchical prevalence rules (Table 5). The result of this step is the final LULC map of the Amazon biome for each year of the time series (Figure 3).

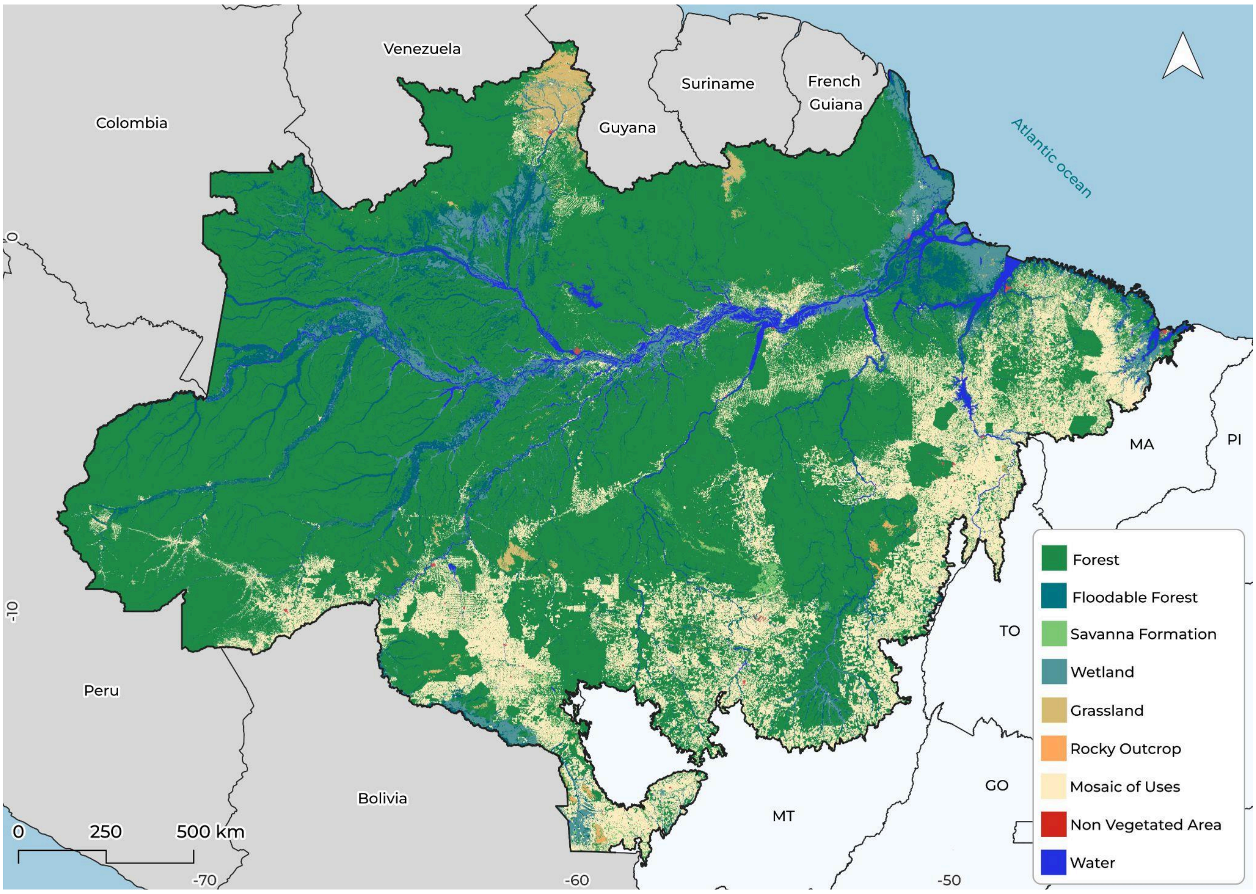

Figure 3. Classified map of the Amazon biome — Sentinel-2, 2025.

Table 5. Prevalence rules for combining digital classification with cross-cutting themes in the Amazon biome — Collection 4 Sentinel.

Order	Class	Class ID	Source
1	Photovoltaic Plant	75	Cross-Cutting Theme
2	Mining	30	Cross-Cutting Theme
3	Beach and Dune	23	Cross-Cutting Theme
4	Mangrove	5	Cross-Cutting Theme
5	Aquaculture	31	Cross-Cutting Theme
6	Salt Flat	32	Cross-Cutting Theme
7	Water (Working Group)	33	Cross-Cutting Theme
8	Urban Infrastructure	24	Cross-Cutting Theme

Order	Class	Class ID	Source
9	Sugarcane	20	Cross-Cutting Theme
10	Soybean	39	Cross-Cutting Theme
11	Rice	40	Cross-Cutting Theme
12	Cotton	62	Cross-Cutting Theme
13	Other Temporary Crops	41	Cross-Cutting Theme
14	Perennial Crops	36	Cross-Cutting Theme
15	Coffee	46	Cross-Cutting Theme
16	Citrus	47	Cross-Cutting Theme
17	Outras Perennial Crops	48	Cross-Cutting Theme
18	Temporary Crops	19	Cross-Cutting Theme
19	Forest Plantation	9	Cross-Cutting Theme
20	Rocky Outcrop	29	Biome
21	Other Non-Vegetated Areas	25	Biome
22	River, Lakes and Ocean	33	Biome
23	Forest Formation	3	Biome
24	Floodable Forest	6	Biome
25	Savanna Formation	4	Biome
26	Arboreal Restinga	49	Biome
27	Wetland	11	Biome
28	Grassland Formation	12	Biome
29	Shrubby Restinga	50	Biome
30	Pasture	15	Cross-Cutting Theme

4. References

Alencar, A., Z. Shimbo, J., Lenti, F., Balzani Marques, C., Zimbres, B., Rosa, M., Arruda, V., Castro, I., Fernandes Márcico Ribeiro, J. P., Varela, V., Alencar, I., Piontekowski, V., Ribeiro, V., M. C. Bustamante, M., Eyji Sano, 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.

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

Brown, C. F., Kazmierski, M. R., Pasquarella, V J., Rucklidge, W. J., Samsikova, M., Zhang, C., Shelhamer, E., Lahera, E., Wiles, O., Ilyushchenko, S., Gorelick, N., Zhang, L. L., Alj, S., Schechter, E., Askay, S., Guinan, O., Moore, R., Boukouvalas, A., & Kohli, P.(2025). AlphaEarth Foundations: An embedding field model for accurate and efficient global mapping from sparse label data. arXiv preprint arXiv.2507.22291

Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D and Moore R. 2017. Google Earth Engine: planetary-scale geospatial analysis for everyone. Remote Sens. Environ., Big Remotely Sensed Data: Tools, Appl. Experiences 202 18–27.

Gumbricht, T., Román-Cuesta, R.M., Verchot, L.V., Herold, M., Wittmann, F., Householder, E., Herold, N., Murdiyarso, D. 2017. An expert system model for mapping tropical wetlands and peatlands reveals South America as the largest contributor. Global Change Biology 23(9):3581–3599. doi: http://www.cifor.org/pid/6419

Hess, L.L., J.M. Melack, A.G. Affonso, C.C.F. Barbosa, M. Gastil-Buhl, and E.M.L.M. Novo. 2015. LBA-ECO LC-07 Wetland Extent, Vegetation, and Inundation: Lowland Amazon Basin. ORNL DAAC, Oak Ridge, Tennessee, USA. https://doi.org/10.3334/ORNLDAAC/1284

Rosa, M. R. (2020). Metodologia de classificação de uso e cobertura da terra para analise de três decadas de ganho e perda anual de cpbertura florestal nativa na mara atlântica (Doctoral Dissertation, Universidade de São Paulo)

Souza JrC. MRobertsD. A.CochraneM. A. (2005). Combining spectral and spatial information to map canopy damage from selective logging and forest fires. *Remote Sens. Environ.*98, 329–343. 10.1016/j.rse.2005.07.013

Souza, C.M., Jr.; Siqueira, J.V.; Sales, M.H.; Fonseca, A.V.; Ribeiro, J.G.; Numata, I.; Cochrane, M.A.; Barber, C.P.; Roberts, D.A.; Barlow, J. Ten-year landsat classification of deforestation and forest degradation in the Brazilian amazon. Remote Sens. 2013, 5, 5493–5513.

Souza C.M. Jr, Oliveira L.A., de Souza Filho J.S., Ferreira B.G., Fonseca A.V. and Siqueira J.V. (2023) Landsat sub-pixel land cover dynamics in the Brazilian Amazon. Frontiers in Forest Global Change 6:1294552. doi: 10.3389/ffgc.2023.1294552

TerraClass 2014. TerraClass 2014 Project [WWW Document] (http://inpe.br/cra/projetos_pesquisas/terraclass2014.php) (Accessed: July 23, 2019).

Tootchi, Ardalan; Jost, Anne; Ducharne, Agnès (2018): Multi-source global wetland maps combining surface water imagery and groundwater constraints. Sorbonne Université, Paris, France, PANGAEA. https://doi.org/10.1594/PANGAEA.892657