MAPBIOMAS
PARAGUAY
COLECCIÓN 3

MapBiomas General "Handbook"

Algorithm Theoretical Basis Document (ATBD)

Collection 3

Version 1.0

September, 2026

Country coordinators

Edder Ortíz

Fernando Palacios

Paraguay team

José Baéz

Camilo Benítez

Lourdes Cristaldo

Claudia Palacios

Paula Paniagua

Table of Contents

Executive Summary

MapBiomas Paraguay Collection 3 is the first annual land use and land cover (LULC) mapping collection produced entirely by the national team under a unified methodology, in contrast to Collections 1 and 2, which were generated through the a posteriori integration of the MapBiomas Chaco and MapBiomas Atlantic Forest products. The collection covers the full Paraguayan territory at 30 m spatial resolution for an annual time series beginning in 1985, consolidating a classification, filtering, and integration architecture proper to the Paraguayan context.

This Algorithm Theoretical Basis Document (ATBD) describes the conceptual framework, data sources, processing architecture, classification methods, post-classification procedures, integration rules, validation strategies, and analytical products underlying Collection 3.

Territorial coverage is organized into eight classification regions defined specifically for Paraguay, integrating the country's main biomes under a single national regionalization, rather than inheriting Mapbiomas Chaco and the Mapbiomas Atlantic Forest zones used in previous collections. The legend follows the MapBiomas hierarchical structure and was not changed compared to Collection 2.

MapBiomas Paraguay is embedded in an institutional trajectory that began in 2018 with Guyra Paraguay joining the MapBiomas network through MapBiomas Chaco, continued in 2019 with WWF Paraguay joining through MapBiomas Atlantic Forest, and consolidated in 2023 with the formal creation of MapBiomas Paraguay under the bi-institutional co-leadership of both organizations. Whereas Collection 2 was released in 2024, extending the time series through 2023, Collection 3 was not released on the standard annual cycle in 2025 precisely because it is the first collection produced entirely in-house by the national team, requiring the full redesign of the classification, filtering, and integration architecture. Following this one-year pause, Collection 3 is released in 2026 covering the two missing years, 2024 and 2025, and marks the transition from a logic of integrating inherited products to a logic of sovereign production with its own methodology, regionalization, and calibration.

Quality assessment for Collection 3 is carried out through an independent external validation, for which the MapBiomas Paraguay network established a formal partnership with the Facultad de Ciencias Agrarias of the Universidad Nacional de Asunción (FCA-UNA), following established good practices for land-cover map accuracy

assessment based on statistical validation with independent samples covering the national territory and the full time series. Validation results —global accuracy by legend level and allocation and area disagreement components— will be incorporated into subsequent versions of this document upon completion of the FCA-UNA assessment.

All MapBiomas Paraguay maps and datasets are freely available on the project platform (https://paraguay.mapbiomas.org/), and all computational algorithms used in the classification are published on GitHub (https://github.com/mapbiomas). The combination of cloud computing and open-source technologies makes this type of national mapping operationally feasible and replicable, sustaining transparency, reproducibility, and MapBiomas Paraguay's contribution to territorial planning, environmental management, scientific research, and public-policy formulation

1. Introduction

1.1. Scope and content of the document

This document describes the theoretical basis, objectives, and methods applied to produce annual maps of land use and land cover (LULC) in Paraguay from 1985 to 2025 of the MapBiomas Collection 3.

Unlike Collections 1 and 2, which were built from the a posteriori integration of the MapBiomas Chaco and MapBiomas Atlantic Forest products, Collection 3 is the first collection produced entirely by the MapBiomas Paraguay team under a unified national methodology. Accordingly, this document describes a single, integrated classification process rather than the aggregation of independent biome-specific pipelines.

This document covers the classification methods of Collection 3, the image processing architecture, the national regionalization into eight classification regions, and the approach used to integrate them into the national mosaic. In addition, the document presents a historical context and background information, a general description of the satellite imagery datasets, feature inputs, the post-classification filter chain, and the accuracy assessment method applied.

The classification algorithms are available on the MapBiomas GitHub (https://github.com/mapbiomas).

1.2. Region of interest

MapBiomas Paraguay was established to produce annual LULC maps for the entire Paraguayan territory, encompassing the country's two main geographic regions —the Occidental Region (Región Occidental or Chaco) and the Oriental Region (Región Oriental)— defined by the Paraguay River and characterized by contrasting vegetation types associated with distinct geomorphological and climatic conditions. The national territory integrates five main biomes: the Dry Chaco and Humid Chaco in the Occidental Region, and the Cerrado, Pantanal, and Atlantic Forest of Alto Paraná in the Oriental Region.

MapBiomas Paraguay Collection 3 organizes the national territory into eight classification regions (Figure 3), delineated on the basis of biophysical, spectral, phenological, and landscape criteria that group areas with comparable land-cover composition and seasonal behavior (Figure 3). This regionalization was designed specifically to capture the environmental heterogeneity of the country —from the semi-arid Dry Chaco in the northwest to the humid subtropical forests of the eastern region— reducing the internal variability within each classification unit and improving the ability of the Random Forest classifier to discriminate spectrally similar classes under contrasting ecological conditions.

The eight regions are the operational units on which most of the classification pipeline is performed: sampling of stable pixels, empirical calibration of per-class sampling limits, training and application of the Random Forest classifier, while post-classification filtering is applied at country level. Regional outputs are subsequently integrated into a single national map through a pixel-by-pixel prevalence rule that resolves the overlaps between adjacent regions along their shared borders.

2. Overview and Background Information

2.1. MapBiomas Network in Paraguay

MapBiomas Paraguay is a joint initiative of WWF Paraguay and Guyra Paraguay, formally established in 2023 under the bi-institutional co-leadership of both organizations. This co-leadership consolidated under a single national platform the collaborative efforts previously carried out by Guyra Paraguay through MapBiomas Chaco (since 2018) and by WWF Paraguay through MapBiomas Atlantic Forest (since 2019).

In Collections 1 and 2, each organization coordinated the mapping of the biome under its historical lead, and national coverage was achieved through a posteriori integration of both biome-level products. Starting with Collection 3, both organizations jointly coordinate a single, unified national mapping process, sharing responsibility over the entire territory, the classification methodology, the regionalization scheme, and the post-classification pipeline. MapBiomas Paraguay collections released to date:

- Collection 1: 1985–2022 (launched in 2023),
- Collection 2: 1985–2023 (launched in 2024),
- Collection 3: 1985–2025 (launched in 2026), **first collection produced entirely under the unified national framework**.

2.1.1. Remote Sensing Data

MapBiomas Paraguay Collection 3 is built on the historical archive of the Landsat program, jointly operated by NASA and USGS, at 30 m spatial resolution. The time series (1985–2025) integrates observations from four sensors: the Thematic Mapper (TM) on board Landsat 5, the Enhanced Thematic Mapper Plus (ETM+) on board Landsat 7, the Operational Land Imager and Thermal Infrared Sensor (OLI/TIRS) on board Landsat 8, and the OLI-2/TIRS-2 sensor on board Landsat 9. All imagery is accessed and processed through Google Earth Engine.

Collection 3 uses the USGS Landsat Collection 2 Tier 1 Level-2 Surface Reflectance (SR) products, a shift from the top-of-atmosphere (TOA) reflectance products used in earlier MapBiomas Chaco and Mapbiomas Atlantic Forest collections. Surface reflectance provides improved radiometric and geometric calibration, more consistent atmospheric correction across sensors, and better cross-sensor comparability throughout the time series. The specific per-year sensor assignment and the mosaic construction procedure are described in Section 3.2.

2.1.2. Google Earth Engine and MapBiomas Computer Applications

MapBiomas image processing is based on Google technology, which includes image processing in cloud computing infrastructure, programming with Javascript and Python via Google Earth Engine, and data storage using Google Cloud Storage.

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

- **Javascript scripts** - these computer codes were written directly in the Google Earth Engine Code Editor and were used to prototype new image processing algorithms and test large-scale image processing. Most image classification and post-classification of Collections 1 and 2 were written in Javascript.

- **Python scripts** – this code category was used to optimize the image processing of large datasets in Google Earth Engine. In addition, the map integration, some post-classification tasks, and statistical analysis were all performed in Earth Engine Python API.

- **Toolkits** – User's Toolkits are collections of scripts with a friendly user interface in Google Earth Engine to download MapBiomas data by state, biome, municipality, or any other geometry. These toolkits were developed for the general public that is not familiar with programming languages used in Google Earth Engine. They are often reviewed and improved. The toolkits for MapBiomas Paraguay are available at https://paraguay.mapbiomas.org/herramientas/.

- **Github repository** - all Javascript, Python, R, toolkit, and dashboard codes are available at the public GitHub repository: https://github.com/mapbiomas.

- **Mapbiomas.org (Dashboard)** - the web platform of the MapBiomas initiative presents the Landsat image mosaics and their quality, land cover and land use annual maps of Collection 3, transitions analysis, statistics, and all the methodological information about the ATBD, tools, scripts, fact sheets, tutorial videos, and accuracy analysis.

2.2. Historical Perspective: Existent mapping initiatives

Systematic monitoring of land use and land cover (LULC) dynamics in Paraguay has evolved substantially over the past three decades, transitioning from isolated regional assessments to nationally coordinated monitoring systems and, more recently, to globally harmonized datasets.

Early initiatives focused primarily on forest cover change detection in the Atlantic Forest and Chaco regions, based on Landsat imagery and rule-based classification approaches (Huang et al., 2006; Huang et al., 2008). These studies laid the methodological foundation for later national and global-scale monitoring efforts. The subsequent expansion of cloud-computing platforms, the increased availability of freely accessible satellite time series, and advances in machine learning have enabled the development of higher-resolution, near-real-time, and globally consistent LULC products, of which MapBiomas Paraguay Collection 3 is part.

2.2.1. Global Land Cover and Land Use Data

Several global datasets are routinely referenced in Paraguayan LULC analyses. Among the most influential is the Global Forest Change dataset (Hansen et al., 2013), derived from Landsat time-series analysis to quantify annual forest loss and gain at 30 m resolution, and updated annually. This dataset has been widely used to characterize deforestation dynamics in both the Atlantic Forest and the Gran Chaco regions. In recent years, global LULC datasets have diversified considerably:

- **Dynamic World v2** (Brown et al., 2022) provides near-real-time (NRT) global LULC classifications derived from Sentinel-2 imagery at 10 m resolution, including per-pixel class probabilities.

- **ESRI 10 m Annual Land Use Land Cover** (Karra et al., 2021) delivers annual global maps (2017–present) based on Sentinel-2 imagery and deep learning models.

- **ESA GlobCover** (Arino et al., 2012) and **ESA WorldCover 10 m 2020** (Zanaga et al., 2021) offer globally consistent land cover products at medium to high resolution.

While global datasets provide spatial consistency and frequent updates, they often present limitations in local thematic accuracy due to generalized legends, limited regional training data, and ecological heterogeneity that global models cannot fully capture. In highly dynamic frontiers such as the Paraguayan Chaco, regionally calibrated models frequently outperform global products in class discrimination and thematic detail.

MapBiomas Paraguay Collection 3 addresses this gap by combining global methodological advances with region-specific training data, national expert validation, and a classification framework designed for the biophysical heterogeneity of the Paraguayan territory.

2.2.2. National Land Cover and Land Use Data

At the national level, Paraguay has developed several LULC mapping initiatives over time, primarily focused on forest monitoring, deforestation tracking, and REDD+ reporting.

One of the earliest comprehensive efforts was the Vegetation and Land Use Map of the Western Region (CIF/FIA/UNA/GTZ, 1991), derived from Landsat imagery for 1986–1987. Subsequent work by the SACH/SERMA–DOA/BGR Project (1998) produced a vegetation map of the Paraguayan Chaco at 1:250,000 scale based on Landsat TM imagery. The Global Land Cover Facility (GLCF) produced forest cover change maps for 1985–2001 using Landsat TM and ETM+ imagery, in collaboration with Guyra Paraguay and the University of Maryland (GLCF, 2009).

Within the REDD+ framework, the Joint National Programme on Reducing Emissions from Deforestation and Forest Degradation (UN-REDD+, 2016) generated historical deforestation maps for 2000, 2005, 2011 and 2015, based on supervised classification of Landsat imagery, distinguishing forest and non-forest categories. Similarly, the National Forestry Institute (INFONA) has produced annual Land Use Change maps supporting official forest monitoring and the establishment of national emissions reference levels.

Additional relevant national or regional products include:

- **FCA/CIF/FFPRI (2013)**: National Land Cover Map (2011) based on Landsat 5 TM (minimum mapping unit: 100 ha).
- **Caldas et al. (2013)**: Annual land-cover change maps in the Paraguayan Chaco (2000–2011) derived from MODIS Terra imagery (250 m resolution).

- **Guyra Paraguay (2019)**: Monthly deforestation monitoring in the Gran Chaco Americano (2010–2018) using Landsat time series.

- **DLR/INFONA/WWF (2018)**: Land use and cover map of the Paraguayan Chaco (2016) using Landsat 8 OLI.

Despite these advances, most national products were either limited temporally, restricted geographically (typically to the Chaco), binary (forest/non-forest), or produced at coarse spatial resolution. Furthermore, methodological heterogeneity has historically hindered long-term harmonized time-series analysis.

MapBiomas Paraguay Collection 3 overcomes these limitations by:

- Providing annual wall-to-wall national coverage (1985–2025);
- Applying a standardized hierarchical legend consistent with the MapBiomas network;
- Using consistent Landsat Collection 2 surface reflectance imagery across the entire time series;
- Ensuring temporal consistency through a comprehensive post-classification filter chain;
- Maintaining full transparency of algorithms and open data access.

In this context, MapBiomas Paraguay Collection 3 represents the first nationally integrated, open-access, high-resolution historical LULC time series for Paraguay produced entirely under a unified national methodology, aligning the country with international best practices in land monitoring systems.

3. Methodological Description

3.1. Overview of the classification workflow

The MapBiomas Paraguay Collection 3 classification workflow produces annual land use and land cover (LULC) maps of the entire Paraguayan territory at 30 m spatial resolution for the 1985–2025 period. The workflow is organized as a single, unified national pipeline in which classification, calibration, and post-classification filtering are performed at the regional level over eight national classification regions (Figure 3), and the regional outputs are subsequently integrated into a single national mosaic. This differs from Collections 1 and 2, in which national coverage resulted from the a posteriori integration of two independent biome-level pipelines (MapBiomas Chaco and MapBiomas Atlantic Forest).

The workflow is implemented primarily in Google Earth Engine using JavaScript for mosaic construction, classification, and post-classification filtering, and complemented by some steps in Python using JupyterLab (earthengine-api, geemap, sklearn) for the empirical calibration of per-class sampling limits and for batch processing across regions and years.

The methodology combines annual Landsat surface reflectance mosaics, a feature space composed of approximately 90 spectral, temporal, and topographic variables, a Random Forest classifier trained per region and per year on stable-pixel samples, and a sequential post-classification filter chain designed to address specific classification limitations and improve the spatial and temporal coherence of the annual LULC trajectories. The complete workflow is illustrated in Figure 1 and can be summarized in the following major steps:

a. **Landsat mosaics:** Annual multi-temporal surface reflectance mosaics are generated from Landsat Collection 2 Tier 1 Level-2 data to ensure optimal spectral quality, minimize atmospheric noise, and maintain temporal consistency across sensors and years.

b. **Classification regions:** The national territory is divided into eight classification regions defined on the basis of biophysical, spectral, phenological, and landscape criteria, reducing internal variability within each classification unit and improving the ability of the Random Forest classifier to discriminate spectrally similar classes under contrasting ecological conditions.

c. **Feature space:** A set of predictor variables comprising spectral bands, spectral indices, seasonal reduction metrics, and static topographic variables, designed to capture the spectral, temporal, and environmental complexity of the Paraguayan landscape.

d. **Training samples and classification:** Stable-pixel samples derived from the previous MapBiomas Paraguay collection are complemented by manually collected and externally provided samples, and used to train a Random Forest classifier operating independently for each region and year. Per-class sampling limits are empirically calibrated to balance class representation without biasing the classifier toward dominant classes.

e. **Integration:** The regional outputs are integrated into a single national map through a pixel-by-pixel prevalence rule, and specific cross-cutting products from the MapBiomas network are incorporated according to predefined prevalence rules to ensure consistency with the regional MapBiomas framework.

f. **Post-classification filtering:** A sequential chain of fourteen filtering stages is applied to correct classification artifacts, remove spurious transitions, and reinforce the temporal consistency of LULC trajectories throughout the 1985–2025 series.

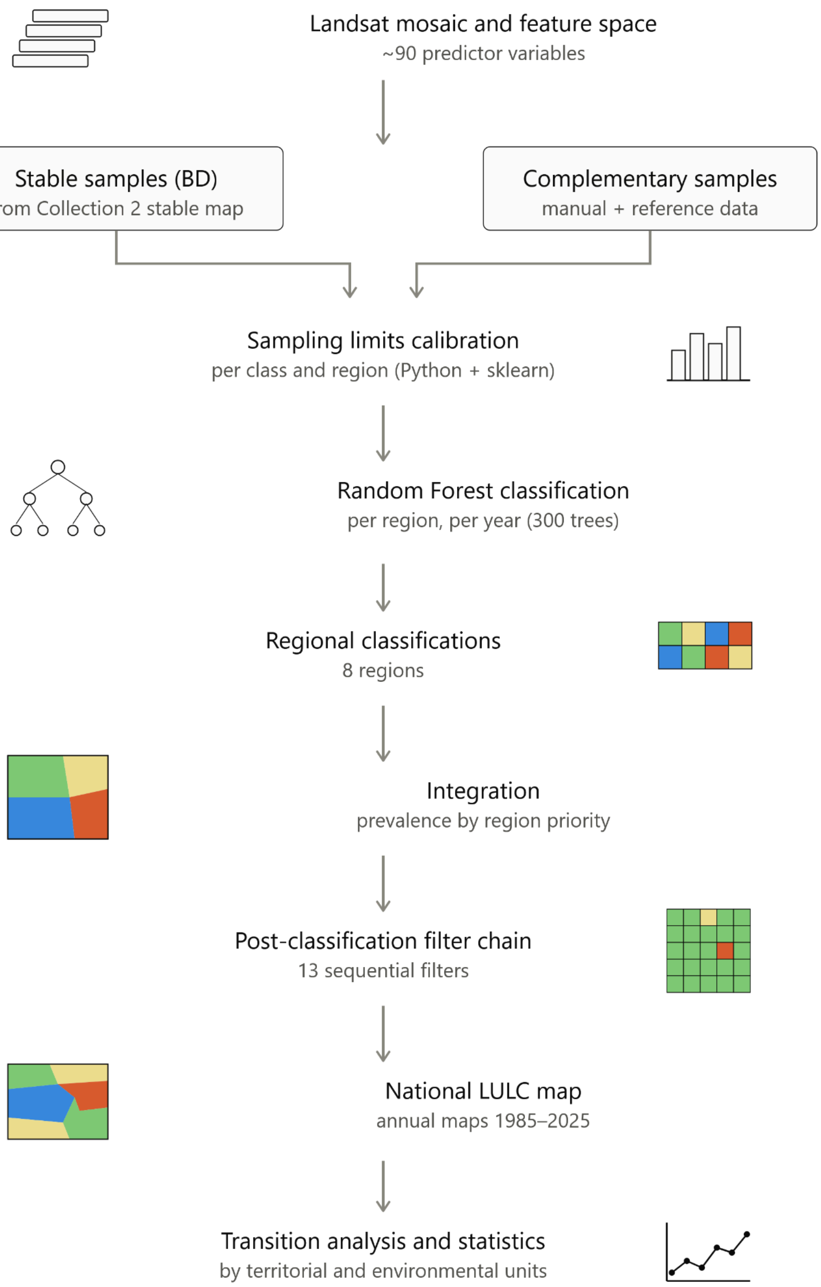

Figure 1. Overview of the MapBiomas Paraguay Collection 3 classification workflow. Annual Landsat surface reflectance mosaics and their associated predictor variables are combined with training samples derived from stable pixels of the previous collection (BD) and complementary samples obtained through visual interpretation and reference datasets. Per-class sampling limits are empirically calibrated for each classification region, and a Random Forest classifier is trained independently for each region and each year (1985–2025), producing eight regional classifications that are subsequently merged into a single national mosaic through a pixel-by-pixel prevalence rule based on regional priority. The integrated national map is then processed through a sequential chain of thirteen post-classification filters designed to correct classification artifacts and reinforce the temporal consistency of the annual

land use and land cover trajectories. The resulting annual national LULC maps constitute the primary output of Collection 3 and serve as the basis for transition analyses and statistics computed by territorial and environmental units.

The final annual LULC maps undergo visual inspection throughout the workflow and are undergoing an independent external accuracy assessment, conducted in partnership with the Facultad de Ciencias Agrarias of the Universidad Nacional de Asunción (FCA-UNA). The following sections describe each of these methodological components in detail.

3.2. Landsat Image Mosaics

The first step in the classification workflow consists of generating annual Landsat mosaics that serve as the primary input for the Random Forest classifier. All mosaics for Collection 3 are constructed from USGS Landsat Collection 2 Tier 1 Level-2 Surface Reflectance products at 30 m spatial resolution, integrating observations from four sensors across the historical series: Landsat 5 (TM), Landsat 7 (ETM+), Landsat 8 (OLI-TIRS), and Landsat 9 (OLI-2/TIRS-2). The adoption of Surface Reflectance across the full time series is a methodological consolidation of Collection 3, replacing the top-of-atmosphere (TOA) reflectance products used in earlier collections and providing improved radiometric consistency and cross-sensor comparability.

The temporal assignment of sensors follows a prioritization strategy designed to maintain uninterrupted national coverage throughout the 1985–2025 period (Table 1). Landsat 5 (TM) is the primary source for 1985–1999 and 2003–2011. Landsat 7 (ETM+) fills the intermediate periods 2000–2002 and the transitional year 2012. Landsat 8 (OLI-TIRS) is the primary source from 2013 to 2022. From 2023 onwards, Landsat 8 is combined with Landsat 9 (OLI-2/TIRS-2) to increase temporal density during the most recent years, and the final year (2025) is derived exclusively from Landsat 9. Mosaic construction is organized following the CIM 1:250,000 grid, which subdivides the Paraguayan territory into 43 tiles processed independently and buffered by 100 m to reduce edge artifacts. Scenes with more than 80% cloud cover are excluded from the input collection; the remaining scenes are masked pixel by pixel using the Collection 2 quality flags, the CloudScore algorithm, and a temporal-outlier detection procedure (TDOM), with a 4-pixel dilation applied to buffer cloud edges. The annual mosaic is built from all valid observations within the calendar year (January 1 – December 31), with per-pixel seasonal separation based on NDVI percentiles as detailed in Section 3.4.

Table 1. Landsat sensor assignment along the 1985–2025 time series of MapBiomas Paraguay Collection 3. All imagery is drawn from USGS Landsat Collection 2, Tier 1, Level-2 Surface Reflectance products.

Period	Primary sensor	Satellite	GEE Collection ID
1985–1999	TM	Landsat 5	LANDSAT/LT05/C02/T1_L2
2000–2002	ETM+	Landsat 7	LANDSAT/LE07/C02/T1_L2
2003–2011	TM	Landsat 5	LANDSAT/LT05/C02/T1_L2
2012	ETM+	Landsat 7	LANDSAT/LE07/C02/T1_L2
2013–2022	OLI/TIRS	Landsat 8	LANDSAT/LC08/C02/T1_L2
2023–2024	OLI/TIRS + OLI-2/TIRS-2	Landsat 8 + Landsat 9	LANDSAT/LC08/C02/T1_L2 + LANDSAT/LC09/C02/T1_L2
2025	OLI-2/TIRS-2	Landsat 9	LANDSAT/LC09/C02/T1_L2

3.2.1. Gap-filling with legacy MapBiomas mosaics

The reprocessing of Paraguayan mosaics under Collection 3 using stricter cloud, shadow, and edge-artifact filtering and masks than earlier collections, produced incomplete coverage in a subset of years, most notably 1985–1987, in which the reduced number of surviving valid observations left persistent gaps across parts of the territory. Rather than propagating these gaps into the classification stage or filling them with adjacent-year values (which would introduce spurious temporal transitions), Collection 3 implements a band-harmonized temporal gap-filling procedure drawing on the legacy MapBiomas mosaics used in Collections 1 and 2, originally produced by the MapBiomas Chaco and MapBiomas Atlantic Forest working groups. Because both the new Paraguayan reprocessing and the legacy mosaics share the same schema by construction (identical reduction metrics, identical NDVI-percentile dry/wet definitions, and identical SMA endmembers per sensor) a lossless pixel-level merge is possible without any spectral rescaling.

The gap-filling is applied per year and per pixel: where the new Paraguayan mosaic has a valid observation its value is retained, and only where it is masked and the legacy mosaic has a valid observation is the legacy value written into the pixel. Two operational rules govern the procedure: (i) for the years 2000–2003 and 2012 (periods in which the primary MapBiomas Atlantic Forest mosaic collection is not populated) only the Landsat 7 variant of the Atlantic Forest mosaics and the Chaco mosaics are eligible for use in the fill; and (ii) gap-filling is applied only through 2023, so that 2024 and 2025 are derived exclusively from the new Paraguayan reprocessing. The output is a coherent annual mosaic stack for the entire 1985–2025 period, produced under the Collection 3 processing standard where possible and complemented by band-compatible legacy observations where necessary (Figure 2). Still adjustments are needed for future collections to improve the quality of the mosaics for earlier years.

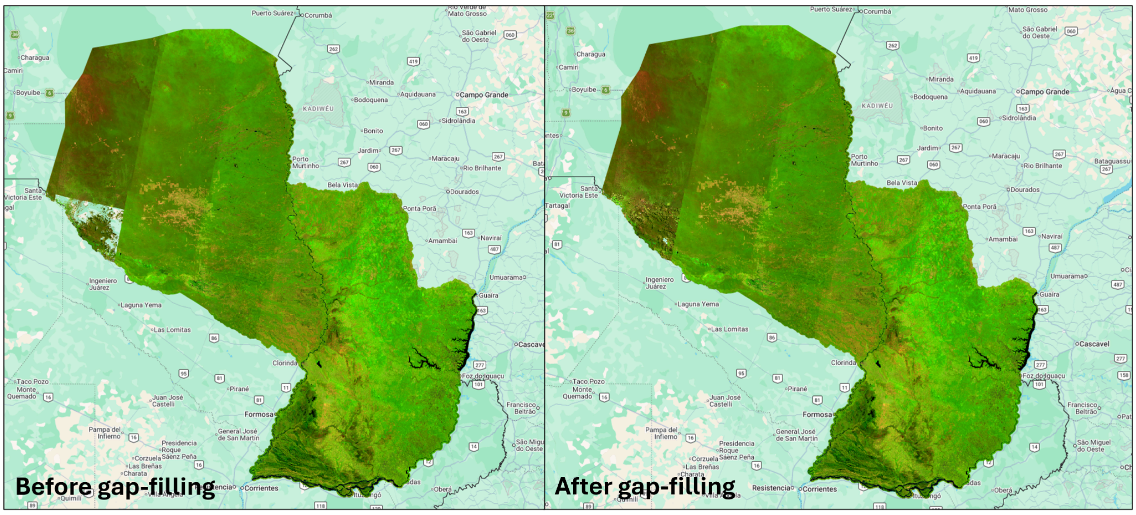

Figure 2. Effect of the band-harmonized gap-filling procedure on the annual Landsat mosaic for 1985, SWIR1-NIR-Red false-color composite. Panel A: the Collection 3 Paraguayan mosaic before gap-filling, showing persistent unmasked gaps produced by the stricter cloud and edge-artifact masks combined with limited valid observations in early Landsat 5 years. Panel B: the same mosaic after pixel-level merging with the legacy MapBiomas Chaco and MapBiomas Atlantic Forest mosaics, harmonized under the same 90-band schema.

3.3. Classification regions

Supervised land use and land cover classifications perform better on smaller and more homogeneous areas, where training samples are more spatially representative of class characteristics than over larger and more heterogeneous domains. Following this principle MapBiomas Paraguay Collection 3 partitions the national territory into eight classification regions, which serve as the operational units on which the entire classification pipeline is executed independently: annual Random Forest classification, empirical calibration of per-class sampling limits, application of most post-classification filters, and generation of country stable-pixel training samples.

The eight regions are not defined as ecological biomes, but as spatially coherent operational units designed to reduce internal heterogeneity within each classification unit while remaining practical to delineate, interpret, and maintain across successive collections. Their boundaries follow ecoregional limits as a first structural criterion, grouping areas with comparable climate, vegetation, and land-use composition, and are then adjusted using natural and infrastructural features that behave as effective landscape discontinuities: watershed boundaries, main river courses, and major roads. These physical breaks are useful because they frequently coincide with genuine changes in land-use patterns (irrigated versus rainfed, subsistence versus mechanized, forest fragment versus consolidated cropland), and because they provide stable, easily interpretable region boundaries that do not shift year to year.

A third criterion, incorporated on the basis of the operational experience accumulated across previous collections, was the identification of areas that had historically produced classification artifacts, most notably abrupt class discontinuities during the region integration process. The delineation of the Collection 3 regions deliberately shifts region boundaries away from these historically problematic zones and, where necessary, redistributes the internal composition of transitional areas so that spectrally difficult land-cover mixtures are handled within a single region rather than split across two. This adjustment reduces the incidence of sharp classification breaks in the integrated national mosaic and was one of the practical motivations for redesigning the regionalization for Collection 3, rather than inheriting the boundaries used in the preceding collections.

The resulting regionalization captures the main ecological and land-use gradients of the country without imposing rigid bioma-based limits where the underlying landscape is transitional. The regionalization is used as-is for the entire 1985–2025 series; regional boundaries are not modified between years. Figure 3 shows the spatial distribution of the eight classification regions over the national territory.

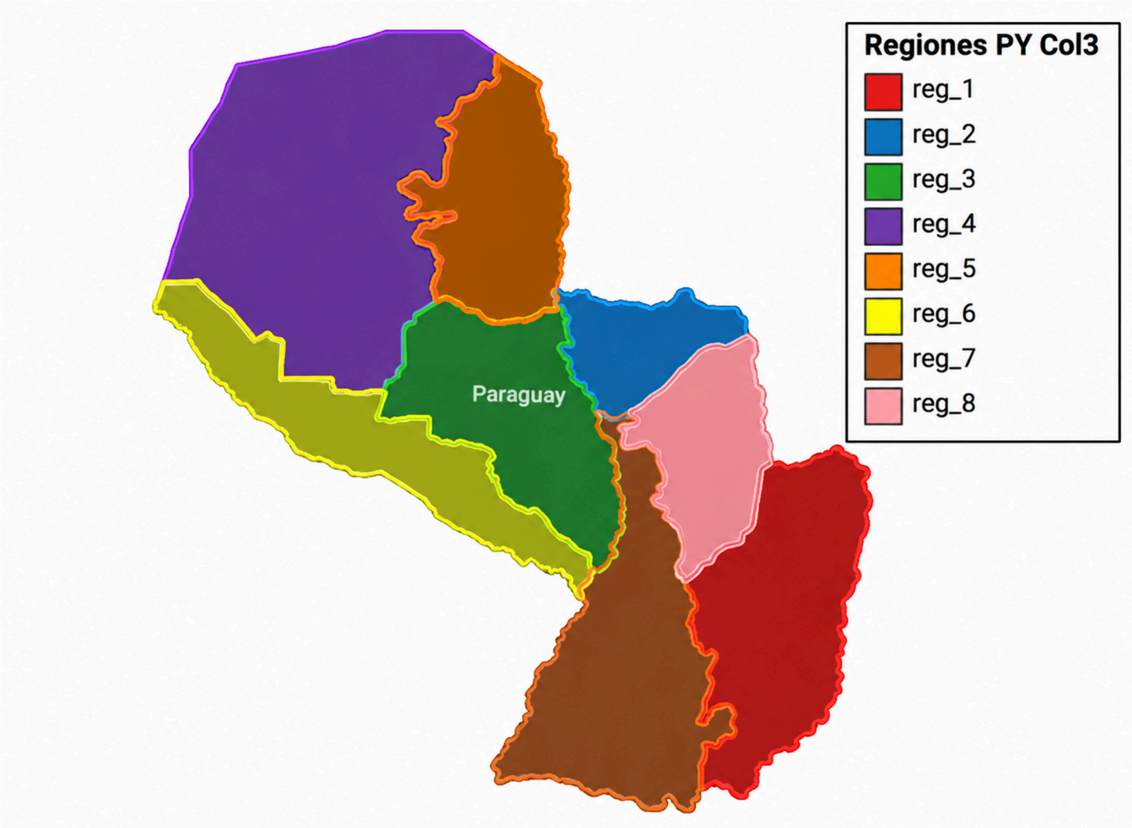

Figure 3. Spatial distribution of the eight classification regions of MapBiomas Paraguay Collection 3 over the national territory.

3.4. Feature space

The feature space used for classification comprises 90 predictor variables per pixel per year, designed to capture the spectral, phenological, structural, and topographic dimensions that discriminate land cover and land use classes across the biophysical heterogeneity of the country. The variables are drawn from six thematic components: Landsat reflectance bands (31 bands), spectral indices computed on reflectance (26), vegetation-structure indices (5), Spectral Mixture Analysis (SMA) fractional-cover bands (15), SMA-derived indices (12), and one topographic variable (slope). The complete inventory, including descriptions, reduction metrics retained, and bibliographic references, is presented in Table 2.

A distinctive component of the MapBiomas Paraguay feature space is the systematic use of Spectral Mixture Analysis (SMA) to decompose each Landsat pixel into fractional abundances of six spectral endmembers: green vegetation (GV), green vegetation shade (GVS), non-photosynthetic vegetation (NPV), soil, cloud, and shade using sensor-specific endmember sets following the MapBiomas convention (Souza et al., 2005). From these fractional bands, three SMA-derived indices are computed: the Normalized Difference Fraction Index (NDFI), the Savanna Ecosystem Fraction Index (SEFI) from Alencar et al. (2020), and the Wetland Ecosystem Fraction Index (WEFI) from Rosa (2020). SMA is particularly valuable in the Paraguayan context, where large parts of the Chaco are dominated by mixed pixels combining tree canopy, herbaceous cover, and exposed soil at sub-pixel scales, and where the Pantanal and Humid Chaco include seasonally flooded classes that benefit from the wetland-specific fraction index.

For each variable, a set of temporal reduction metrics is computed from all cloud-masked Landsat observations acquired within the calendar year. The overall annual median is retained for every variable as the baseline reduction. The dry-season median (median of observations with NDVI ≤ the 25th percentile of the annual distribution at the pixel) and the wet-season median (median of observations with NDVI ≥ the 75th percentile) are retained for variables sensitive to phenological seasonality. The amplitude (annual maximum minus annual minimum) is retained for variables in which the intra-annual range carries discriminatory information for seasonal vegetation and land-use classes. The standard deviation is retained as a measure of temporal variability where dispersion contributes discriminatory power. The annual minimum is retained only for the six reflectance bands, where it captures the darkest observation of the year and is diagnostic for surface water and shadowed classes. This selective retention of reduction metrics, rather than the full Cartesian product of variables × reductions, was defined empirically across previous MapBiomas Paraguay collections to maximize classification performance while keeping the feature space computationally tractable in Google Earth Engine. The specific reductions retained for each variable are listed in Table 2.

Two additional variables complete the feature space. A GLCM entropy texture computed on the annual median of the green band is included as a proxy for local spatial heterogeneity, contributing to the discrimination of classes with similar spectral signatures but different landscape organization (for example, fragmented cropland versus continuous pasture). The terrain slope, derived from a digital elevation model, is the only static (non-temporal) variable in the schema and is included to discriminate spectrally similar classes that occur in different terrain positions, most notably surface water from terrain shadows in areas of complex relief. The complete feature space is applied consistently across the eight classification regions and across the entire 1985–2025 time series.

Table 2. Feature space used for classification in MapBiomas Paraguay Collection 3. For each variable, the 'Reductions retained' column lists the temporal reduction metrics preserved in the final band schema. Reductions are computed over all cloud-masked Landsat observations acquired within the calendar year: median = annual median; median_dry = median of observations with NDVI ≤ p25 of the annual distribution; median_wet = median of observations with NDVI ≥ p75; amp = annual amplitude (max − min); stdDev = annual standard deviation; min = annual minimum. 'Identity' indicates the variable is used directly, without temporal reduction. Spectral indices and CAI are stored with a +1 offset applied to the native computation to avoid negative values; SMA-derived indices are stored as byte data in the range after the transformation value×100 + 100; SMA fractional-cover bands are stored as byte data in the range [0, 100]. The complete schema comprises 90 bands per pixel per year.

Type	Name	Description / Formula	Reductions retained	Reference
Landsat band	Blue	Blue surface reflectance band	median, median_dry, median_wet, min, stdDev	USGS
	Green	Green surface reflectance band	median, median_dry, median_wet, min, stdDev	USGS
	Red	Red surface reflectance band	median, median_dry, median_wet, min, stdDev	USGS
	NIR	Near-infrared surface reflectance band	median, median_dry, median_wet, min, stdDev	USGS
	SWIR1	Shortwave infrared 1 surface reflectance band	median, median_dry, median_wet, min, stdDev	USGS

Type	Name	Description / Formula	Reductions retained	Reference
	SWIR2	Shortwave infrared 2 surface reflectance band	median, median_dry, median_wet, min, stdDev	USGS
	Green texture	GLCM entropy computed on green_median (local spatial heterogeneity)	Identity	This collection
Spectral index	NDVI	Normalized Difference Vegetation Index — (NIR − Red) / (NIR + Red)	median, median_dry, median_wet, amp, stdDev	Rouse et al., 1974
	NDWI	Normalized Difference Water Index (Gao) — (NIR − SWIR1) / (NIR + SWIR1)	median, median_dry, median_wet, amp, stdDev	Gao, 1996
	EVI2	Enhanced Vegetation Index 2 — 2.5 × (NIR − Red) / (NIR + 2.4 × Red + 1)	median, median_dry, median_wet, amp, stdDev	Jiang et al., 2008
	SAVI	Soil Adjusted Vegetation Index — 1.5 × (NIR − Red) / (0.5 + NIR + Red), L = 0.5	median, median_dry, median_wet, stdDev	Huete, 1988
	PRI	Photochemical Reflectance Index (Landsat adaptation) — (Blue − Green) / (Blue + Green)	median, median_dry, median_wet	Gamon et al., 1992
	GCVI	Green Chlorophyll Vegetation Index — (NIR / Green) − 1	median, median_dry, median_wet, stdDev	Gitelson et al., 2003

Type	Name	Description / Formula	Reductions retained	Reference
Vegetation structure	CAI	Cellulose Absorption Index (SWIR ratio adaptation) — SWIR2 / SWIR1	median, median_dry, stdDev	Adapted from Nagler et al., 2003
	Hall Cover	Canopy-cover regression — exp(−0.017·Red − 0.007·NIR − 0.079·SWIR2 + 5.22)	median, stdDev	Hall et al., 2006
SMA fractional cover	GV	Green vegetation fraction from linear spectral unmixing	median, amp, stdDev	Souza et al., 2005
	GVS	Shade-normalized green vegetation — GV / (GV + NPV + Soil) × 100	median, median_dry, median_wet, stdDev	Souza et al., 2005
	NPV	Non-photosynthetic vegetation fraction from linear spectral unmixing	median	Souza et al., 2005
	Soil	Soil fraction from linear spectral unmixing	median, amp, stdDev	Souza et al., 2005
	Cloud	Cloud fraction from linear spectral unmixing	median, stdDev	Souza et al., 2005
	Shade	Shade fraction (residual) — |100 − (GV + NPV + Soil)|	median, stdDev	Souza et al., 2005
SMA-derived index	NDFI	Normalized Difference Fraction Index — (GVS − (NPV + Soil)) / (GVS + (NPV + Soil))	median, median_dry, median_wet, amp, stdDev	Souza et al., 2005

Type	Name	Description / Formula	Reductions retained	Reference
	SEFI	Soil Enhanced Fraction Index — ((GV + NPV)▯ − Soil) / ((GV + NPV)▯ + Soil), where (GV + NPV)▯ = (GV + NPV) / (GV + NPV + Soil) × 100	median, median_dry, stdDev	MapBiomas SMA framework
	WEFI	Water Enhanced Fraction Index — ((GV + NPV) − (Soil + Shade)) / ((GV + NPV) + (Soil + Shade))	median, median_wet, amp, stdDev	MapBiomas SMA framework
Topography	Slope	Terrain slope derived from digital elevation model	Identity	Tadono et al., 2014

3.5. Training samples and Classification

The classification is performed independently for each of the eight classification regions and for each year of the 1985–2025 series, using a Random Forest classifier trained on a curated per-region training set. The training set is built from two complementary sources: a stable-pixel training map derived from the previous MapBiomas Paraguay collection and refined for recent land-cover changes, and a set of complementary samples, manually delineated by the team's interpreters or drawn from external reference datasets, that reinforce classes underrepresented in the stable base. Prior to model training, the per-class contribution of the stable base is empirically optimized through a metrics-driven calibration procedure that balances general classification accuracy, discrimination at class boundaries, and inter-annual prediction stability. The following paragraphs describe each of these components in the order in which they are produced.

3.5.1. Stable-pixel training map

The stable-pixel map is derived from the integrated annual classifications of MapBiomas Paraguay Collection 2. For each pixel of the national territory, the number of distinct land-cover classes assigned across the full Collection 2 time series is computed, and only pixels that maintain the same class in every year are retained as candidate reference pixels. Two auxiliary Collection 2 classes used to encode information not directly applicable to the Collection 3 legend are excluded, and two class identifiers are consolidated to align with the Collection 3 legend (class 19 remapped to class 18 for agriculture; class 33 remapped to class 26 for water). The resulting stable pixel base contains ten target classes: forest formation (3), open woody vegetation (4), flooded woody vegetation (6), forest plantation (9), flooded grassland (11), grassland (12), pasture (15), agriculture (18), non-vegetated area (22), and water (26).

Because a strict "same class in every year" criterion is unable to capture land-cover changes that occurred at the tail end of the Collection 2 series, the stable-pixel base is subsequently refined through a recent-change screening procedure based on Dynamic World probability data (Brown et al., 2022). For each 10 m Dynamic World pixel, the per-class probabilities of every Dynamic World observation acquired within each of three temporal windows —2022–2023 as the baseline, 2024 as the target, and 2025 as confirmation— are averaged into a single per-class probability distribution per window, and the dominant class of each window is identified as the class with the highest average probability. Pixels for which the dominant class changed between the baseline and target windows, and for which the change persisted into 2025, are flagged as having undergone a genuine land-cover change during 2024. The 10 m change flags are aggregated to the 30 m stable-pixel grid using a 7-of-9 majority rule, and pixels for which the majority of sub-pixels reflect a persistent change are removed from the stable base. This step ensures that the training data reflects the actual land-cover state of each pixel throughout the classification period rather than the state it held prior to a recent transition and accounts for the 2-years gap between collection 2 and 3.

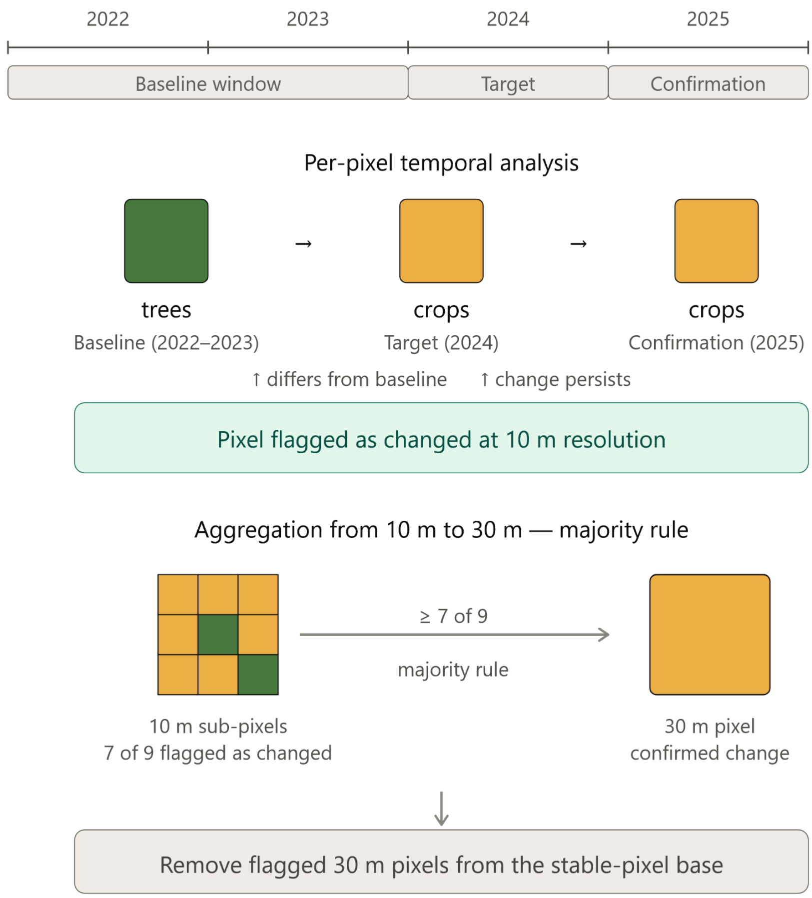

Figure 4. Illustrated example of the Dynamic World-based recent-change screening applied to the stable-pixel training map of MapBiomas Paraguay Collection 3. Top: three temporal windows defined on the 2022–2025 span (baseline 2022–2023, target 2024, confirmation 2025). Middle: per-pixel temporal analysis of a representative 10 m Dynamic World pixel, where the dominant class in each window is determined by first averaging the per-class probabilities of all Dynamic World observations within the window, and then selecting the class with the highest average probability; a pixel is flagged as having undergone a change when both the target and the confirmation dominant classes differ from the baseline dominant class (illustrated by a hypothetical trees-to-crops transition that persists in 2025). Bottom: aggregation from the native 10 m Dynamic World grid to the 30 m stable-pixel grid via a 7-of-9 majority rule (illustrated for a 30 m cell where 8 of the 9 sub-pixels are flagged as changed). Flagged 30 m pixels are removed from the stable-pixel base prior to sample allocation.

3.5.2. Complementary samples

For classes that are underrepresented in the stable-pixel base or that show high spectral or temporal dynamics, most notably forest plantation (class 9), and agriculture (class 18) in some regions, the training set is enriched with two types of complementary samples:

- **Manually delineated polygons** are collected by the team's interpreters over reference imagery for each classification region.

- **External reference datasets** are additionally incorporated where available, most notably the INFONA forest plantation cartography for 2024, which supplies a spatially exhaustive reference of plantation extent. Both manual and external polygons are converted to per-pixel samples at 30 m resolution during model training, with the number of samples generated per polygon and per year configurable to balance the classifier's exposure to each complementary class.

3.5.3. Per-region sampling from the stable base and per-year feature extraction

From the refined stable-pixel base, a stratified random sample of 800 points per class per region is drawn independently for each of the eight classification regions, using a fixed random seed for reproducibility. Prior to sampling, the outer boundary of the region collection is dissolved and eroded inward by 300 m to prevent sampling of edge pixels to accumulate classification noise; internal region boundaries remain unaffected by this operation. For each sampled point, the values of the 90 predictor variables that compose the feature space are extracted from each annual gap-filled mosaic covering 1985–2025, and each observation is tagged with its acquisition year. Samples with missing feature values in a given year are dropped for that year. The resulting per-region tables of stable-pixel samples, each containing all sample points enriched with 41 years of feature-space values, constitute the input to the sampling-limits calibration procedure described below.

3.5.4. Sampling-limits calibration

The uniform allocation of 800 samples per class per region does not necessarily correspond to the class distribution that maximizes classification performance. Class-balanced training tends to over-represent minority classes at the expense of accuracy on dominant classes, while area-proportional training tends to bias the classifier toward dominant classes and worsen discrimination at class boundaries. To find a per-region allocation that balances these effects empirically, for Collection 3 an automated calibration procedure was implemented. The procedure operates entirely outside Google Earth Engine on the per-region sample tables produced upstream, using Python and scikit-learn.

The calibration optimizes a composite objective function that combines three complementary metrics computed across the full 1985–2025 series. The first metric is the macro-averaged F1 score on a stratified held-out validation set (30% of the samples per class per year, disjoint from any training pool used during the search), which measures overall classification quality under balanced class prevalence. The second metric is the macro-averaged F1 score on a boundary-focused validation set, constructed by identifying, for each of eleven predefined confusable class pairs (for example, closed forest versus flooded woody vegetation, pasture versus agriculture, agriculture versus non-vegetated area), the thirty samples of each class of the pair that are closest in the standardized feature space to the centroid of the opposite class of the pair. This set of ambiguous boundary samples measures the classifier's ability to resolve mixtures rather than pure signatures.

The third metric is a temporal-inconsistency term, computed as the mean fraction of pixels whose predicted class changes between consecutive years in a fixed validation set of 100 samples per class drawn from the middle year of the series and classified independently by the model trained for each year; this term penalizes allocations that produce temporally unstable classifications. The composite objective assigns equal weight to the two accuracy terms and a lower weight to the temporal term (with a penalty coefficient of 0.5), which was found empirically to prevent the search from converging to allocations that stabilize the time series at the expense of accuracy.

The optimization proceeds in two stages. In the first stage, uniform allocations at eight candidate levels (10, 15, 25, 50, 100, 200, 400, and 800 samples per class) are evaluated across all regions, and the best-performing uniform allocation is adopted as the starting point of the second stage. In the second stage, a coordinate-wise iterative search is applied: for each class in order, all candidate values in the per-class grid are evaluated while holding the allocations of the other classes fixed, and the value maximizing the composite objective is retained. The procedure iterates over classes for a maximum of two rounds, or until a full pass produces no further improvement. The per-class grid is capped at the class-level upper bound determined by the minimum yearly availability of that class across the full 1985–2025 series, ensuring that the search only proposes allocations that are actually deliverable in every year. To assess the robustness of the resulting allocation, the optimized vector is re-evaluated under five distinct random seeds, and the standard deviation of the resulting scores is compared with the improvement over the best uniform baseline; when the two are comparable, the simpler uniform allocation is preferred for reproducibility.

3.5.5. Random Forest classification

For each classification region and each year of the 1985–2025 series, a Random Forest classifier is trained on the combined training set constituted by (i) the stable-pixel samples for that region, trimmed to the calibrated per-class allocation, and (ii) the complementary samples for that region and year, in the configuration established for the region by the team. The classifier uses 200 decision trees, with the number of variables considered at each split set to the default value of the square root of the total number of predictors (approximately nine variables per split, from the 90-band feature space). The trained classifier is then applied pixel by pixel to the annual mosaic for that region and year, producing a single-band raster of predicted class values that serves as input to the post-classification chain.

3.6. Integration

The eight regional annual classifications are merged into a single national land use and land cover raster covering the full 1985–2025 time series. This integration step is the last operation before the post-classification filter chain and produces the national annual stack on which all subsequent filtering operations are performed.

The merge is executed pixel by pixel. For pixels that fall entirely within a single classification region, the value of that region's classification is directly retained. For pixels that fall in the overlap between two or more regions —which can occur along the shared boundaries between adjacent regions, where region polygons are buffered outward for edge protection during classification, a per region prevalence rule determines which region classification prevails.

The prevalence order was defined empirically by the MapBiomas Paraguay team on the basis of observed regional classification performance across previous collections and iterative visual inspection of Collection 3 outputs at region boundaries. From highest to lowest prevalence, the order is:

Region 4 > Region 5 > Region 7 > Region 3 > Region 6 > Region 2 > Region 1 > Region 8.

Regions with historically higher classification stability at their boundary zones are placed at the top of the prevalence order, so that their outputs govern the integrated map wherever they overlap with adjacent regions. This empirical prioritization complements the region-delineation criterion introduced earlier which already shifts region boundaries away from historically problematic transition zones, and together the two criteria minimize the incidence of sharp classification discontinuities in the integrated national mosaic. The output of the integration step is a single national annual map, one classification band per year, that serves as the input to the post-classification filter chain described in the following section.

3.7. Post-classification

ed empirically to ensure that (i) mechanical preparation of the temporal series precedes any pattern-based correction, (ii) simple spatial cleanups precede temporal cleanups so that residual spatial noise does not trigger false temporal corrections, and (iii) filters that incorporate country-specific ecological rules or that rely on external auxiliary datasets are applied at defined stages of the sequence, so that their effects are preserved in the final output. Table 3 summarizes the filter sequence in execution order, and the following sub-sections describe each filter grouped by functional type.

Table 3. Sequence of post-classification filters applied to MapBiomas Paraguay Collection 3, listed in execution order. Filters are grouped into four functional types — data preparation, spatial filters, temporal filters, and trajectory-pattern filters — plus a final set of other filters addressing specific spectral and thematic corrections. Each filter is applied sequentially to the integrated national annual land use and land cover maps to correct classification artifacts and reinforce temporal consistency.

N°	Filter	Purpose
1	**Gap fill**	Fill missing years per pixel via forward and backward temporal fill
2	**First spatial filter**	Remove small patches (≤ 0.5 ha) via 9×9 focal-mode replacement; four native/water classes protected
3	**Urban and non-vegetated areas filter**	Reclassify GISD30 urban pixels with Random Forest confirmation; small-patch cleanup; Open-Buildings-based stabilization of consolidated urban surfaces
4	**Topographic filter**	Correct water and flooded-grassland errors on steep slopes using SRTM

5	**Spatial-temporal transitions**	Correct A–B–A temporal patterns co-occurring with small spatial extent
6	**Temporal general filter**	Main temporal cleanup: 5-, 4-, and 3-year kernel windows with PY-specific exceptions and protection of recent anthropic conversions
7	**Frequency hierarchical filter**	Stabilize pixels temporally dominated by native classes (≥ 90% native frequency) to a hierarchically selected dominant native class
8	**Trajectories filter**	Correct grassland intrusions in anthropic series and specific inter-anthropic intrusions; waterlogged-terrain exception applied via HAND auxiliary
9	**False regrowth filter**	Correct short native blocks embedded in consolidated anthropic series (pasture, agriculture)
10	**Silviculture filter**	Correct forest-plantation classification errors using the INFONA national plantation cadastre
11	**Waterlogged terrain filter**	Correct anthropic and silviculture intrusions into HAND-waterlogged native environments (rivers, meanders, floodplains)
12	**Temporal single-year filter**	Correct one-year isolated occurrences on temporally stable pixels via replacement with the pixel's temporal mode

13	**Second spatial filter (final)**	Final cleanup of patches ≤ 5 pixels (~0.45 ha) via 3×3 focal-mode replacement; same four protected classes as filter 2

Three of the filters in the chain incorporate country-specific corrections that formalize ecological and land-use observations accumulated by the MapBiomas Paraguay team across previous collections. The Silviculture filter targets short blocks of native forest cover assigned inside otherwise persistent forest plantation series; plantations in the Paraguayan context are stable long-term land uses that do not revert to native forest within the classifier's timeframe, so short native intrusions in plantation trajectories are reverted to plantation. The False Regrowth filter targets the inverse pattern: short blocks of native vegetation classified within consolidated pasture or agriculture series, which in the Paraguayan context are not consistent with the slow multi-year signature of genuine regeneration and are consequently reverted to the surrounding anthropic class. Both the Trajectories filter and the Waterlogged Terrain filter incorporate an explicit exception for pixels flagged as waterlogged by an auxiliary raster derived from the Height Above Nearest Drainage (HAND) index —a hydrological descriptor that expresses, for each pixel, the vertical height above the nearest downstream drainage cell, with values effectively at zero indicating pixels most likely to be periodically flooded or waterlogged: in the Humid Chaco, along riparian corridors, and on floodplains, oscillations between flooded and non-flooded native classes are legitimate hydrological cycles rather than classifier noise, and the generic corrections that would smooth them are disabled where the auxiliary layer indicates waterlogged terrain.

3.7.1. Data preparation

Gap fill

The first step in the chain fills temporal gaps in the integrated map. For each pixel, the values across the annual bands are traversed forward in time and any masked year inherits the value of the preceding non-masked year; the traversal is then repeated backward in time to fill gaps that remain at the beginning of the series. The output is a fully populated 41-band annual stack with no temporal gaps, on which all subsequent filters operate. Filling temporal gaps early in the chain is important because most downstream filters compare each pixel's class values across consecutive years, and temporal gaps would otherwise create spurious "transitions" from and to null values.

3.7.2. Spatial filters

First spatial filter

This filter removes small isolated patches of classification that are typically the result of pixel-level classifier noise rather than genuine landscape features. For each year, connected-component patches smaller than six pixels (approximately 0.5 ha at 30 m resolution) are replaced by the modal class value within a 9×9 pixel focal neighborhood. Four classes are explicitly excluded from this filter to preserve legitimate small-scale ecological features: forest formation, flooded woody vegetation, flooded grassland, and water bodies, which routinely occur as small legitimate patches such as gallery forests, wetland fragments, and small ponds. The filter is applied year by year, without temporal context.

Second spatial filter

This filter is applied at the end of the chain, it is a stricter spatial cleanup applied as the closing operation of the post-classification pipeline. Patches of five or fewer connected pixels (approximately 0.45 ha) are replaced by the modal class of a 3×3 pixel focal neighborhood. This threshold is more restrictive than that of the First Spatial filter; the rationale is that any small patch that has survived every temporal and spatial cleanup upstream and still measures five pixels or less is almost certainly residual noise rather than legitimate landscape structure. The same four classes protected in the first Spatial remain untouched here.

3.7.3. Temporal filters

Spatial-temporal transitions filter

While the First Spatial filter operates on each year independently, this filter identifies transitions that are anomalous in both space and time. A three-year temporal window is applied over every pixel of every year: if the middle year is classified into a thematic category different from the preceding and following years, and those neighboring years agree with each other, the middle year is a candidate for correction. The correction proceeds only if the spatial patch of the anomalous transition is also small (fewer than 0.5ha), on the reasoning that transitions that are simultaneously ephemeral in time and small in space are almost always classifier noise. The filter uses three thematic categories: native vegetation (forest formation, open woody, flooded woody, flooded grassland, grassland), anthropic use (plantation, pasture, agriculture), and other (non-vegetated area, water). Transitions such as native → anthropic → native, or anthropic → native → anthropic, that meet both temporal and spatial conditions trigger the correction and the middle year is reverted to the class of the preceding year. Stable large patches of genuine change, and temporally coherent isolated pixels, are left untouched.

Temporal general filter

This is the main temporal cleanup of the pipeline. It applies moving-window rules of decreasing size 5-year, 4-year, and 3-year kernels to each pixel's temporal series, in a class-hierarchical order that gives priority to correcting errors that would compromise native classes if left uncorrected. For each combination of class and window length, if a pixel's class in a candidate year is inconsistent with the surrounding years of the window, the candidate year is smoothed to match the surrounding context. The 4-year window is applied only to a specific pattern (native–agriculture–agriculture–native) that the MapBiomas Paraguay team, across previous collections, identified as a persistent source of false transitions; while the generic 4-year cleanup was found to introduce more false positives than it corrected in the Paraguay context. Second, the two most recent years of the series are protected from cleanup of anthropic-to-native transitions, because it could be genuine anthropic conversions in these years are not yet supported by a long observation history so it should not be smoothed away. After the generic pipeline, three country-specific exception patterns catch residual cases: single-year (Forest, Agriculture, Forest), single-year (Forest, Pasture, Forest), and two-year (Forest, Agriculture, Agriculture, Forest) intrusions of anthropic classes within stable forest are reverted to forest.

Frequency hierarchical filter

This filter stabilizes pixels whose temporal series is dominated by native vegetation but contains oscillations between different native classes. For each pixel, the temporal frequency of each native class across the full 1985–2025 series is computed, and the pixel is classified as "stable native" when the combined frequency of all native classes exceeds 90% of the years. For each stable-native pixel, a single dominant native class is assigned by applying a hierarchical set of per-class frequency thresholds ordered so that forest formation takes precedence over any other native class it competes with. In each year of the pixel's series, the class is corrected to the dominant native class only if the pixel is stable native, the current-year class is a native class different from the dominant one, and the current-year class does not form a persistent block of at least three consecutive years with itself.

The three-year persistence safeguard prevents the filter from collapsing genuine multi-year fluctuations of native cover (for example, three consecutive years of open woody vegetation in a pixel that is otherwise dominantly closed forest could reflect real inter-annual variability rather than classifier noise).

Temporal single-year filter

This filter is a targeted single-year cleanup that complements the earlier temporal filters. For each pixel and each year, if the class in that year differs from both the preceding and following years, forming a one-year isolated occurrence in the pixel's temporal series, the year is reverted to the pixel's temporal mode, provided that the mode class occurs in more than 40% of the years of the pixel's series (i.e., the pixel has an unambiguously dominant class over time). The first and last years of the series are not evaluated because they lack a complete temporal context. This filter catches residual single-year errors that survived the more sophisticated temporal filters upstream, particularly on pixels with stable long-term trajectories whose isolated flicker years were not addressed by the pattern-specific rules of earlier filters.

3.7.4. Trajectory-pattern filters

Trajectories filter

This filter targets two groups of trajectory patterns (specific sequences of classes across a pixel's time series that the MapBiomas Paraguay team identified as consistent classifier errors).

The first group are intrusions of grassland within consolidated pasture or agriculture series, on the assumption that these are typically short-lived spectral confusions between herbaceous cover types rather than genuine transitions. Three patterns are corrected: (i) a single year of grassland embedded within surrounding pasture or agriculture; (ii) two consecutive years of grassland similarly embedded; and (iii) a single year of grassland at the end of the series (2025), preceded by at least three consolidated years of the surrounding class. In all three cases the intruding years are reverted to the surrounding class.

The second group concerns intrusions between anthropic classes themselves. Single-year and two-year patterns of pasture within agriculture, agriculture within pasture, and plantation within agriculture are corrected to the surrounding class. No correction is applied on pixels flagged as 'waterlogged' by the Height Above the Nearest Drainage (HAND) auxiliary layer. This exception explicitly tries to preserve ecologically meaningful oscillations that a generic temporal cleanup would otherwise smooth away.

False regrowth filter

This filter targets short blocks of native vegetation embedded in consolidated anthropic-use series, which are treated as classifier errors rather than genuine regeneration. The correction depends on how long the native block is: a single-year intrusion is reverted when preceded by at least two anthropic years and followed by one; a two-year intrusion requires three preceding and two following; three- and four-year blocks require five preceding and two following. Shorter blocks demand less surrounding context because they are more likely to be noise. In all cases the corrected years take the anthropic class that follows the native block, so legitimate transitions between anthropic uses (for example, agriculture converting to pasture) are respected.

A separate rule anchors the extremes: pixels that are anthropic in both 1985 and 2025 have all internal native blocks shorter than five years reverted to the surrounding anthropic class; longer blocks are respected as potentially genuine regeneration.

Waterlogged terrain filter

This filter corrects spurious intrusions of anthropic into genuinely waterlogged native environments. Pixels flagged as waterlogged by the HAND auxiliary layer (HAND value effectively zero, indicating riverbed, meander, or persistently anegated terrain) are inspected for short blocks that intrude into an otherwise native series; these short blocks are replaced by a pre-computed pixel-wise native mode. The plantation rule is applied only outside the INFONA plantation mask, to preserve legitimate riparian plantations. This filter targets a specific and recurrent classifier error: bare-soil or bright-vegetation spectral signatures within riverine and floodplain environments are occasionally misclassified as anthropic land use, producing narrow strips of pasture or agriculture along river courses that are inconsistent with the surrounding hydrological context. The HAND-based restriction ensures that the correction only affects zones that are unambiguously waterlogged, and does not touch anthropic classes that are legitimately located near, but above, the drainage network.

3.7.5. Other filters

Urban and non-vegetated areas filter

The non-vegetated area class combines two surface types: anthropic (urban, infrastructure), which is usually irreversible once consolidated, and natural (rocky outcrops, sand riverbanks), which oscillates legitimately with climate. The classifier does not distinguish between the two, so truly urban pixels often appear as pasture, agriculture, or grassland in some years and non-vegetated in others. This filter identifies genuinely urban pixels and locks them to non-vegetated area across the whole series.

Urban pixels are identified from two sources. From the Global Impervious Surface Dataset (GISD30), which records the year each pixel became urban: these pixels are locked to non-vegetated area from that year onward, if the classifier had originally assigned them to pasture, agriculture, or non-vegetated area, and had flagged them as non-vegetated in at least five years. From Google Open Buildings v3 footprints (with a 30 m buffer): pixels the classifier had flagged as non-vegetated in at least fifteen years are locked to non-vegetated area in the years where the classifier confuses them with pasture, agriculture, or grassland. Between the two, a spatial cleanup replaces small non-vegetated patches (fewer than six connected pixels) with the focal-mode class of a 9×9 neighborhood.

Topographic filter

This filter corrects two spectral errors caused by terrain relief: terrain shadows on steep slopes misclassified as water, and bright wet-looking pixels misclassified as flooded grassland. Using the SRTM digital elevation model, three rules reclassify flagged pixels to forest formation: flooded grassland on slopes above 12%; water on slopes above 15% if the pixel has at least two years of native vegetation history (protecting genuine reservoirs in hilly terrain); and water on slopes of 40% or more, unconditionally. The filter is essentially inactive in the Chaco and takes effect mainly in eastern regions with meaningful relief (Alto Paraná, Cordillera de los Altos, Ybytyruzú).

Forest plantation filter

This filter targets two symmetric errors involving forest plantations. Because plantations do not revert to native forest within the classifier's timeframe, short native blocks inside otherwise persistent plantation series are treated as classifier errors; conversely, short plantation blocks inside otherwise stable non-plantation series are also errors, because a genuine plantation would produce a persistent multi-year signature. We use the public API of the National Forestry Institute (INFONA) to access to Forest plantation data for 2024 and use it as a mask. Inside INFONA layer, single-year and two-year native intrusions are reverted to plantation, native regeneration blocks of up to eight years within a longer plantation series are also reverted, and an additional rule stabilizes plantation trajectories in the recent years of the series. Outside the mask, plantation blocks of one to three years embedded in a stable series of any other class are reverted to the surrounding class.

3.8. Map Legend

The MapBiomas Paraguay Collection 3 legend follows the hierarchical structure adopted by the MapBiomas network and is inherited without changes from MapBiomas Paraguay Collection 2 (Table 4).

Table 4. Hierarchical land use and land cover legend of MapBiomas Paraguay Collection 3, adapted from the MapBiomas Paraguay Collection 2 legend." (ajustá la redacción según lo que diga la primera línea de la sección: "The MapBiomas Paraguay Collection 3 legend follows the hierarchical structure adopted by the network and is inherited without changes from MapBiomas Paraguay Collection 2

Class Level 1	Class Level 2	ID	Legend color	Description
Woody Natural Vegetation	Closed Natural Woodlands	3		Land covers dominated by trees and shrubs, grouped together as woody vegetation without distinction between growth forms. Includes only natural or semi-natural vegetation covers with a woody canopy of more than 65%.
	Open Natural Woodlands	4		As above, but with a woody canopy between 15% and 65%.

	Flooded Natural Woodlands	6	Transition areas between purely terrestrial and purely aquatic systems, where the water table is generally at or near the surface (waterlogged areas). Natural vegetation formed by trees, shrubs, and palm savannas, transitional between the two systems and significantly influenced by water or dependent on flooding.
Herbaceous Natural Vegetation	Flooded Grassland	11	Transition areas between purely terrestrial and purely aquatic systems, where the water table is generally at or near the surface (waterlogged areas). Natural herbaceous vegetation significantly influenced by water or dependent on flooding, including esteros, bañados, cañadas, marshes, and aquatic beds.
	Grassland	12	Areas of natural herbaceous vegetation. The category admits the presence of woody vegetation, provided its cover falls approximately between 1–5% and 20%.
Agricultural Areas	Pasture	15	Areas of cultivated herbaceous species intended for forage (livestock production).

Agriculture	18		Areas of annual herbaceous crops, mainly cereals and legumes cultivated for grain or fibre production.
Forest Plantation	9		Areas of cultivated tree species (for example pine and eucalyptus plantations, citrus and fruit orchards).
Non-vegetated Area	22		Areas of artificial cover such as buildings, roads, quarries, and mining infrastructure. Also includes areas without vegetation cover and without artificial cover (natural bare surfaces such as rocky outcrops, sand riverbanks, and salt flats).
Water Body	26		Areas covered by water, whether natural (rivers, lakes) or artificial (reservoirs, canals, artificial lakes).
Not Observed	27		Assigned to missing values or other errors that may arise from the classification process.

4. Map Collections and Analysis

MapBiomas Paraguay Collection 3 produces a set of analytical products derived from the annual land cover and land use classifications described in Section 3. These include annual national LULC maps for each year of the 1985–2025 series, transition maps between selected pairs of years, statistics of area and number of changes summarized by administrative and environmental territorial units, and derived thematic analyses.

All products are freely available through the MapBiomas Paraguay platform at https://paraguay.mapbiomas.org/, together with the code that generates them.

4.1. Transition analysis

Transition maps and their associated area statistics describe how land cover and land use classes shift between two selected years of the series. For each pair of years, a transition map is generated by concatenating the two annual classifications on a pixel-by-pixel basis, producing a single raster in which each pixel value encodes both its class in the starting year and its class in the ending year. From this raster, transition-area matrices are computed at the national level and at every territorial unit for which the platform reports statistics (departments, districts, biomes, watersheds, and other environmental units), quantifying the total area that shifted from each class to each other class between the two selected years.

Transition maps are generated on demand through the MapBiomas Paraguay platform, which allows users to select any pair of years within the 1985–2025 range. Fixed reference transitions, typically 1985 vs. 2025, and decadal transitions, are pre-computed and available directly as downloadable products, together with their corresponding transition-area matrices.

5. Accuracy Assessment

The accuracy of MapBiomas Paraguay Collection 3 is being assessed through an independent external validation, marking a methodological milestone for the initiative. For this collection, the MapBiomas Paraguay network established a formal partnership with the Facultad de Ciencias Agrarias of the Universidad Nacional de Asunción (FCA-UNA), which is conducting the accuracy assessment of the product using a statistical validation with independent samples covering the national territory and the full 1985–2025 time series. The validation follows the good-practice framework of Olofsson et al. (2014)

and Stehman (2019), which is the reference approach adopted across the MapBiomas network. Also the sample size is allocated per stratum following the equations of Cochran (2007) as summarized by Olofsson et al. (2014), targeting a specified overall standard error of the estimated global accuracy. Each sample point is independently labelled by the FCA-UNA team through visual interpretation of Landsat imagery time series and high-resolution reference imagery, without prior knowledge of the Collection 3 assignment for that point.

Accuracy will be reported at each level of the legend (Level 1 and Level 2) and will consists of three complementary metrics: (i) overall accuracy, the proportion of validation samples for which the Collection 3 class and the reference class agree; (ii) allocation disagreement, the portion of total disagreement attributable to the spatial allocation of classes across the territory; and (iii) quantity disagreement, the portion attributable to differences between the total area of each class in the Collection 3 map and in the reference. Metrics will be also reported per year of the series and aggregated over the full time series, so that both temporal and thematic patterns of classification quality can be inspected.

At the time of publication of this document, the FCA-UNA validation is in progress and its results are not yet available. Full accuracy results —including per-year global accuracy at Level 1 and Level 2, allocation and quantity disagreement components, and per-class user's and producer's accuracies— will be incorporated into subsequent versions of this ATBD upon completion of the assessment, and will be made publicly available on the MapBiomas Paraguay platform (https://paraguay.mapbiomas.org/) together with the map products.

References

1. Arino, O., Ramos Perez, J. J., Kalogirou, V., Bontemps, S., Defourny, P., & Van Bogaert, E. (2012). Global land cover map for 2009 (GlobCover 2009). European Space Agency (ESA) & Université Catholique de Louvain.
2. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324
3. Brown, C. F., Brumby, S. P., Guzder-Williams, B., Birch, T., Hyde, S. B., Mazzariello, J., Czerwinski, W., Pasquarella, V., Haertel, R., Ilyushchenko, S., Schwehr, K., & Radeloff, V. C. (2022). Dynamic World, near real-time global 10 m land use land cover mapping. Scientific Data, 9, 251. https://doi.org/10.1038/s41597-022-01307-4
4. Caldas, M. M., Goodin, D., Sherwood, S., Campos Krauer, J. M., & Wisely, S. M. (2013). Land-cover change in the Paraguayan Chaco: 2000–2011. Journal of Land Use Science, 10(1), 1–18. https://doi.org/10.1080/1747423X.2013.80731
5. CIF/FCA/UNA & FFPRI. (2013). Mapa de cobertura de la República del Paraguay (2011). Facultad de Ciencias Agrarias, Universidad Nacional de Asunción & Forestry and Forest Products Research Institute of Japan.
6. GLCF (Global Land Cover Facility). (2009). Forest cover change map of Paraguay (1985–2001). University of Maryland.
7. Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S., Tyukavina, A., Thau, D., Stehman, S., Goetz, S., Loveland, T., Kommareddy, A., Egorov, A., Chini, L., Justice, C., & Townshend, J. R. G. (2013). High-resolution global maps of 21st-century forest cover change. Science, 342(6160), 850–853. https://doi.org/10.1126/science.1244693
8. Huang, C., Kim, S., Altstatt, A., Townshend, J. R. G., Davis, P., Song, K., Tucker, C. J., Rodas, O., Yanosky, A., & Clay, R. (2008). Assessment of Paraguay's forest cover change using Landsat observations. Global and Planetary Change, 67(1–2), 1–12.
9. Huang, C., Kim, S., Song, K., Townshend, J. R. G., Davis, P., Altstatt, A., Rodas, O., Yanosky, A., & Clay, R. (2006). Rapid loss of Paraguay's Atlantic forest and the status of protected areas—A Landsat assessment. Remote Sensing of Environment, 106(4), 460–466.
10. INFONA. (2018–2022). Mapas de cambio de uso del suelo en Paraguay. Instituto Forestal Nacional, Paraguay.
11. Souza, C. M., Shimbo, J. Z., Rosa, M. R., Parente, L. L., Alencar, A., Rudorff, B. F. T., Hasenack, H., Matsumoto, M., Ferreira, L. G., Souza-Filho, P. W., 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., Pareyn,

F., 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. (2020). Reconstructing three decades of land use and land cover changes in Brazilian biomes with Landsat archive and Earth Engine. Remote Sensing, 12(17), 2735. https://doi.org/10.3390/rs12172735

12. UN-REDD+. (2016). Documento técnico metodología mapa de cambios de cobertura forestal Paraguay (2000–2015). Programa Nacional Conjunto ONU-REDD+ Paraguay.
13. USGS (United States Geological Survey). (2017). Landsat Collection 1 Level-1 quality assessment band. U.S. Geological Survey.
14. USGS. (2021). Landsat Collection 2 Surface Reflectance product guide. U.S. Geological Survey.
15. Wulder, M. A., Loveland, T. R., Roy, D. P., Crawford, C. J., Masek, J. G., Woodcock, C. E., Allen, R. G., Anderson, M. C., Belward, A. S., Cohen, W. B., Dwyer, J., Erb, A., Gao, F., Griffiths, P., Helder, D., Hermosilla, T., Hipple, J. D., Hostert, P., Hughes, M. J., Zhu, Z. (2019). Current status of Landsat program, science, and applications. Remote Sensing of Environment, 225, 127–147.
16. Zanaga, D., Van De Kerchove, R., De Keersmaecker, W., Brockmann, C., Kirches, G., Wevers, J., Cartus, O., Santoro, M., Fritz, S., & Herold, M. (2021). ESA WorldCover 10 m 2020 product description and validation report. European Space Agency.