MapBiomas Democratic Republic of the Congo

Collection 1

Algorithm Theoretical Basis Document (ATBD)

August, 2026

General coordinator

Landing Mané

Team

Humid Zone

Biome Coordinator

Jean-Pierre Pitchou Meniko

Developers

Corneille Ewango

Landry Nkaba

Jean-Fiston Mikwa

Nelson Nkumunue

Jean-Paul Tokala

Jean-Aimé Didier Mbenga

Esther Ndongo

Joseph Nzambo

André Kayombo

Jean-Michel Mutombo

Valentine Biwata

Arsène Kayengi

Serge Diroken

Gaetan Mwanza

Miombo

Biome Coordinator

Jonathan Muledi

Developers

Séraphin Irenge

Marie-Michelle Kifinga

François Munyemba

Fiston Lumbala

Mosaic

Biome Coordinator

Elena Ndinga

Developers

Judith Mongai

Eric Lutete

Julien Iwewe

Gaston Nlandu

Junior Ngoyi

Anaclet Madibua

Mountain / Terra Firma

Biome Coordinator

Céline Sikulisimwa

Developers

Joel Tungi Tungi

Angali Zabibu

Gracia Lukusa

Serge Nziavake

Charles Ilunga

Gloire Kayitoghera

Technical support

Bárbara Costa, Brazil Team

Dhemerson Conciani, Brazil Team

Victor Kadiata, DRC Team

Introduction

The objective of this document is to describe the theoretical basis, justification, and methods applied to produce annual maps of land use and land cover (LULC) in the Democratic Republic of Congo (DRC) from 2000 to 2025 (Collection 1). The document presents a general description of the satellite image processing, the feature inputs, and the step-by-step process applied to obtain the annual classifications.

1.2 Region of interest

The initiative was launched in Africa in 2025, with the Democratic Republic of Congo (DRC) as the pilot country for producing annual LULC maps for the country's territory. The generation of the maps was carried out by expert teams and organizations with regional and historical knowledge about LULC changes in the major ecoregions of DRC. To account for the specific characteristics of the country's diverse ecosystems, the national territory was divided into five operational ecoregions (biomes): the Humid Zone, the Miombo Zone, the Dryland Zone, the Mountain Zone, and the Mosaic Zone (Figure 1).

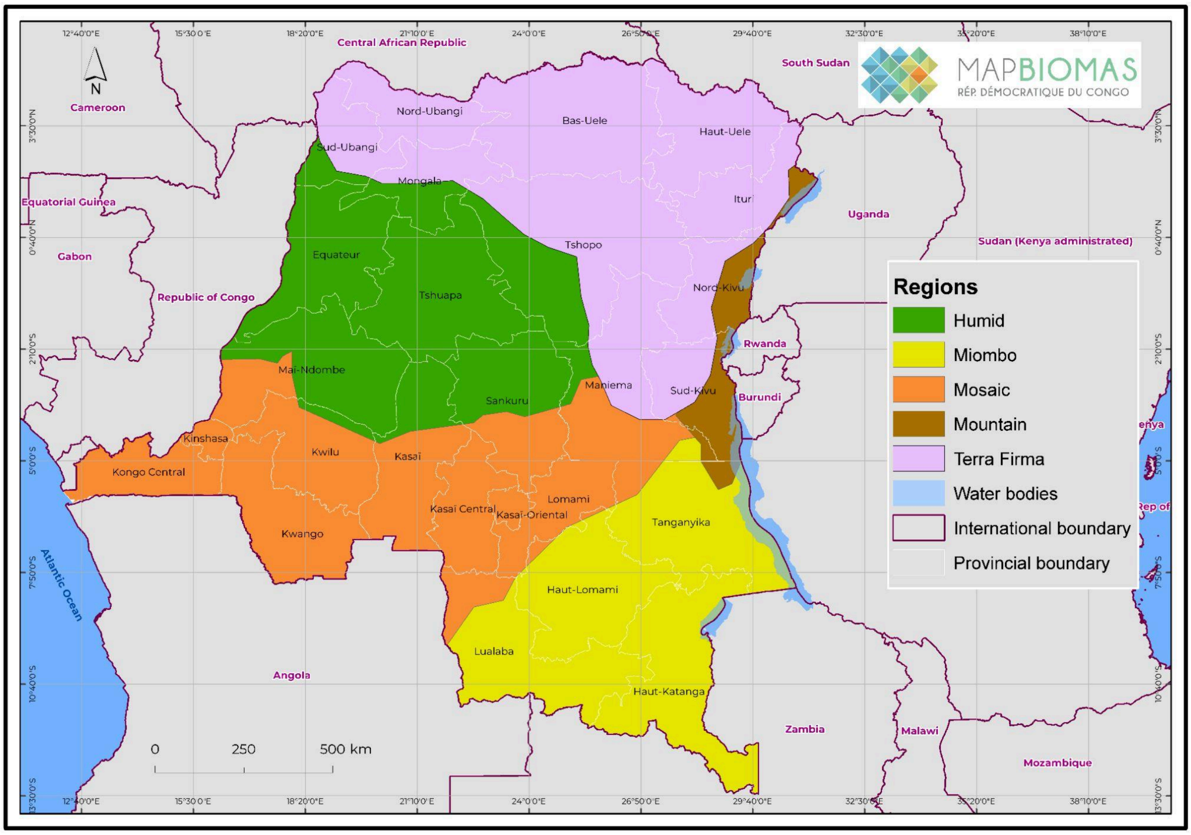

Figure 1. Location of the Democratic Republic of Congo (DRC) and the operational ecoregions used in MapBiomas DRC Collection 1.

We provide a brief description of each operational DRC ecoregion below:

Humid Zone

These environments are characterized by high seasonal variability, marked ecological heterogeneity, and frequent spectral confusions between classes (open water, aquatic vegetation, flooded forests, agro-humid mosaics, etc.), which makes class discrimination more complex and can affect the temporal stability of the results. The "Humid Zone" group operates across a west-east gradient of the Congo Basin, extending from west of Mai-Ndombe, through Mbandaka, Lisala, and Bumba, to Kisangani. This axis encompasses floodplains, marshes, and humid forests subject to contrasting hydrological regimes, including areas of permanent and seasonal flooding. The diversity of landscape units encountered along this corridor provides a relevant framework for characterizing spatio-temporal transitions and consolidating a robust reference base for Humid-specific classes.

Miombo

The Miombo of Katanga region in the Democratic Republic of Congo is a forest-savanna ecosystem of major importance. Dominated by shrubby and tree legumes of the genera Brachystegia, Julbernardia, and Isoberlinia, it provides essential ecological functions: climate regulation, habitats for biodiversity, carbon storage, and support for local human activities. The Miombo region faces increasing pressures: deforestation and fuelwood extraction, expansion of slash-and-burn agriculture, industrial and artisanal mining, urban expansion, and recurring fires. These transformations necessitate rigorous and continuous monitoring of land cover and land use dynamics.

Mosaic

The Mosaic zone extends from Kongo Central to western Katanga, passing through the dismembered provinces of Bandundu and Kasai. It also encompasses part of Maniema and South Kivu. Unlike other ecoregions (biomes), the "Mosaic" zone is not dominated by a particular land cover class. It contains a mix of several classes, sometimes highly fragmented: forests, savannas, agricultural areas, urban areas, etc.

Its climate is characterized by an alternation between a rainy season and a dry season, more pronounced than in the Cuvette Centrale. It serves as a refuge for several species, including forest elephants, buffaloes, and antelopes. This zone is of major economic importance because human activities—such as agriculture, livestock farming, and urbanization—are most dense here.

However, the "Mosaic" zone is vulnerable to various pressures, particularly deforestation, which transforms forest patches into grassy savanna to meet the demand for fuelwood (makala) and land for slash-and-burn agriculture. It is also subject to the impacts of climate change, manifested by reduced rainfall. This phenomenon accelerates the "savannization" of the region, making the soil less fertile in the long term."

Terra Firma

The Terrestrial biome is located in the north of the DRC, specifically in the provinces of Sud-Ubangi, Nord-Ubangi, Mongala, Bas-Uele, Haut-Uele, Ituri, Sud-Kivu, Nord-Kivu, Maniema, and Tshopo. The Terrestrial biome is primarily composed of the vast, continuous rainforest (dense rainforest) that covers most of the country's central basin. However, its extreme northeastern part is dominated by a forest-savanna mosaic. This biome is characterized by dry soil. The trees are giant, reaching over 40 meters in height, and remain green year-round. The forest is very dense and rich in biodiversity, with precious woods such as ebony, wenge, and iroko. Evergreen forests, receiving 1600 to 2000 mm of rain per year, are locally dominated by gregarious species such as *Brachystegia laurentii*, *Gilbertiodendron dewevrei*, or *Julbernardia seretii*, forming specific associations depending on the region (Germain and Evrard, 1956).

Mountain Biome

The Mountain Biome is located in the far eastern part of the DRC, especially in the provinces of Tanganyika, South Kivu, North Kivu, and Ituri. It is dominated by mountain forests and lakes. The forests of the mountain biome are situated between 1600 and 2100 m in altitude, where the ecological conditions—high humidity, large temperature variations, and filtered light—favor medium-sized evergreen formations. The vegetation structure varies with altitude: the rainforest (1600–1900 m), rich in species such as *Podocarpus latifolius* and *Prunus africana*, is followed by the undifferentiated Afromontane forest at lower elevations, which is often replaced after fires by *Juniperus procera* or *Hagenia abyssinica*. On dry slopes, sclerophyllous stands dominated by Juniperus procera develop, while on some eastern slopes there are almost pure forests of *Hagenia abyssinica*. Bamboo groves of *Sinarundinaria alpina* dominate in the high-altitude areas (2300–2600 m). Under the influence of human activities, particularly deforestation and grazing, these formations generate secondary forests characterized by pioneer species such as *Croton macrostachys* or *Polyscia fulva* (Devred, 1960; Lebrun and Gilbert, 1954).

The total mapped area for the entire DRC was 234.5 million hectares (Mha) per year from 2000 to 2025.

1.3 Institutions

MapBiomas Democratic Republic of the Congo is a collaborative network of NGOs, Universities, and Research Institutes that is constantly growing. The network is a group of specialists in remote sensing, ecology, Geographic Information Systems (GIS), and programming from different regions of the country, focused on producing annual collections of land cover and land use maps of the DRC since 2000 (Table 1).

Table 1. Co-creators by regions:

DRC Region	Institutions
Humid Zone	- Université de Kisangani (UNIKIS) - Université de Mbandaka (UNIMBA) - Institut Facultaire Agronomique de Yangambi - University of Kinshasa/ Congo Basin Water Resources Research Center (UNIKIN/CRREBaC) - Ecosystems Restoration Associates – Congo (ERA-CONGO) - Congolese Institute for Nature Conservation/Eala Botanical and Zoological Garden (ICCN/JBZE) - Woodwell Climate Research Center (WCRC Foundation)
Miombo	- Kamina University (UNIKAM) - University of Kolwezi (UNIKOL) - University of Lubumbashi / Observatory of Open Forests in the Democratic Republic of Congo(UNILU/OFCC)
Mosaic	- Satellite Observatory of Central African Forests (OSFAC) - Actions for the Promotion and Protection of Endangered Peoples and Species (APEM) - National Institute for Agricultural Study and Research (INERA/ Ngandajika) - National League of Indigenous Pygmy Associations in Congo (LINAPYCO) - Official University of Mbuji-Mayi (UOM) - President Joseph Kasa-Vubu University (UKV)

	- CEEDD
Terra Firma & Mountain	- Ministry of Public Health, Hygiene and Social Welfare/Taskforce on Health and the Environment (MSPHPS/TS&E) - Satellite Observatory of Central African Forests (OSFAC) - Regional Postgraduate School for Integrated Planning and Management of Tropical Forests and Territories (ERAIFT) - Catholic University of Graben / Butembo (UCG) - University of Kinshasa / Laboratory of Hydrochemistry and Environmental Chemistry (UNIKIN/LAHYE) - Official University of Bukavu (UOB) - Official University of Mbuji-Mayi (UOM)

2. Remote Sensing Data

2.1 Landsat Collection

The imagery dataset used in the MapBiomas Democratic Republic of the Congo (DRC), Collection 1, was obtained from the Landsat sensors Thematic Mapper (TM), Enhanced Thematic Mapper Plus (ETM+), Operational Land Imager and Thermal Infrared Sensor (OLI-TIRS and OLI2-TIRS), on board of Landsat 5 (LANDSAT/LT05/C02/T1_L2), Landsat 7 (LANDSAT/LE07/C02/T1_L2), Landsat 8 (LANDSAT/LC08/C02/T1_L2) and Landsat 9 (LANDSAT/LC09/C02/T1_L2), respectively. The Landsat imagery collections with 30 m pixel resolution were accessible via Google Earth Engine (GEE) and were provided by NASA and USGS. The MapBiomas DRC Collection 1 used Collection 2, Tier 1, and Level 2 Landsat Surface Reflectance products from USGS, which underwent radiometric calibration and orthorectification based on ground control points and a digital elevation model to account for pixel co-registration and correct displacement errors.

According to the year and the quality of available images, a specific Landsat collection was selected and combined

- from 2000 to 2011: Landsat 5 and Landsat 7,
- year 2012: only Landsat 7,
- from 2013 to 2022: only Landsat 8,
- from 2023 to 2025: Landsat 8 and 9

2.2 Definition of the temporal period

The yearly mosaics were formed by compositing pixels from each set of images, considering all the Landsat images available from 1st January to 31st December. This increases the probability of getting at least one cloud-free observation even in regions with high cloud cover. This limit was established based on a visual analysis after many trials observing the results of the cloud removing/masking algorithm. Finally, the yearly mosaics were manually inspected by experts for each region to remove noisy scenes from the mosaic generation and to ensure the quality of the input data used in the classification.

3. Classification

3.1 Overview of the methodological process

The methodological procedures of MapBiomas DRC Collection 1 included several steps (Figure 2). The first step was to generate annual Landsat image mosaics for each year. The second step was to generate a training dataset of stable areas encompassing the mapped classes (Table 1) for the period 2000 to 2025. This procedure was collaboratively carried out by experts in each classification region. A final review of the training samples was made by at least one senior analyst in each classification region.

Then, the spectral feature inputs derived from the yearly Landsat mosaics were extracted and associated with each sample point. Once samples for each LULC class were selected within each classification region, the training dataset was adjusted to meet its statistical needs. The number of training samples per class was determined empirically based on reference maps and the MapBiomas DRC teams' regional knowledge. This step was repeated multiple times in each classification region, with improvements implemented based on the spatio-temporal evaluation of preliminary results. Additionally, complementary samples were generated, defining georeferenced points for different classes through visual interpretation of historical satellite images (high- and very high-resolution) and time series of vegetation indices.

Using the adjusted training dataset, supervised classification with the Random Forest algorithm was performed for each classification region and year. Following that, gap, spatial, temporal, and frequency filters were applied to remove classification noise and stabilize the classification. The LULC maps of each region were integrated to generate the final map of Collection 1. The MapBiomas DRC annual LULC maps were used to derive the transition analysis and statistics.

The statistical analysis covered different spatial territories, such as the country, provinces, territories, and biomes.

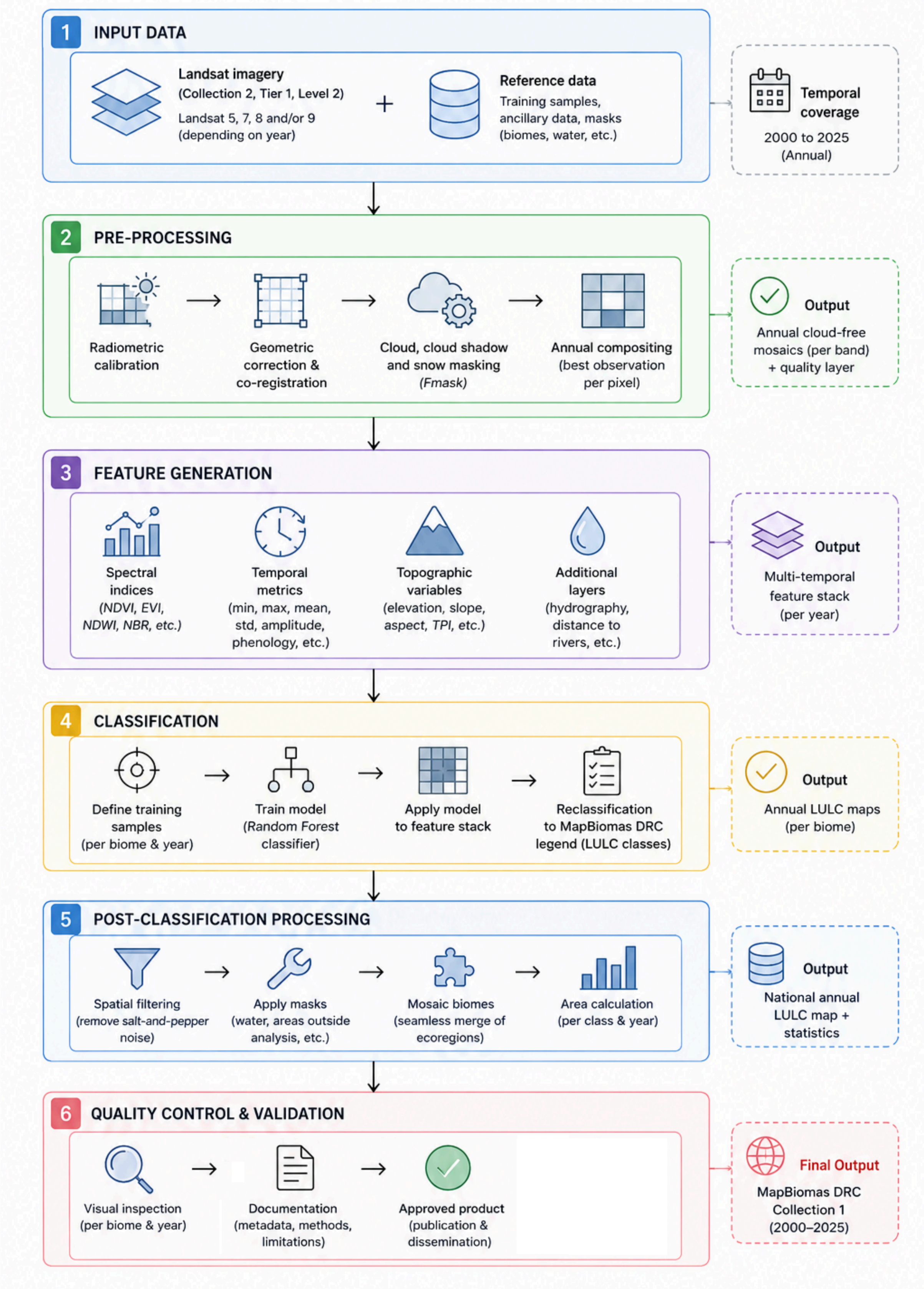

Figure 2. Flowchart of the classification process of Collection 1 in the MapBiomas Democratic Republic of the Congo (DRC) for the period 2000-2025.

3.2 Map legend

The classification for the MapBiomas Democratic Republic of the Congo (DRC) initiative using Landsat mosaics included eleven LULC classes (Table 1): Forest Formation (3), Woodland (4) Mangrove (5), Shrubland (66), Pasture and Agriculture (21), Forest Plantation (9), Urban Area (24), Other Non Vegetated Area (25), River or Lake (33), and Non Observed (27). A full description of the legend is provided in the document *Legend Description – Collection 1*, available on the country website.

Table 1. Land cover and land use classes considered in the MapBiomas Democratic Republic of the Congo (DRC) in Collection 1.

Class Name	ID	Hexacode	Color
1.1 Forest Formation	3	#1f8d49	
1.2 Woodland	4	#7dc975	
1.3 Mangrove	5	#04381d	
2.1 Shrubland	66	#a89358	
3.1. Pasture and Agriculture	21	#ffefc3	
3.2. Forest plantation	9	#7a6c00	
4.1. Urban area	24	#d4271e	
4.2 Other non-vegetated area	25	#db4d4f	
5.1 River or lake	33	#2532e4	
6. Not observed	27	#ffffff	

3.3 Annual Mosaics

The total available bands of the MapBiomas Democratic Republic of the Congo (DRC) feature space are composed of 119 input variables, including the original Landsat bands (Blue, Green, Red, NIR, SWIR 1 and 2), and spectral indexes, fractional, and textural information derived from these bands (Table 2). Per-pixel statistical reducers were used to generate yearly and seasonal temporal features, such as:

- Median: median of the pixel values in a given year
- Median_dry: median of the quartile of pixels with the lowest NDVI values
- Median_wet: median of the quartile of pixels with the highest NDVI values
- Amplitude: amplitude of variation considering all the images of the year
- stdDev: standard deviation considering all the images of the year
- Min: lower annual value of the pixels considering all the images of the year

Table 2. List of the variables included in the feature space used in the classification processes of the MapBiomas Democratic Republic of the Congo (DRC) Collection 1 (2000-2025).

Type	Name	Formula / Description	Statistics	Reference
Landsat Band	Blue, Green, Red, NIR, SWIR1, SWIR2	Original reflectance bands	Median, Median_dry, Median_wet, Minimum, stdDev	USGS
Spectral Index	NDVI Normalized Difference Vegetation Index	(NIR - Red) / (NIR + Red)	Median, Median_dry, Median_wet, Amplitude, stdDev	Rouse et al., 1974
	EVI2 Enhanced Vegetation Index 2	2.5 × (NIR - Red) / (NIR + 2.4 × Red + 1)	Median, Median_dry, Median_wet, Amplitude, stdDev	Parente et al., 2018
	SAVI		Median_dry, Median_wet, Median, stdDev	
	GCVI Green Chlorophyll Vegetation Index	(NIR / Green - 1)	Median, Median_dry, Median_wet, stdDev,	Burke et al., 2017
	PRI Photochemical Reflectance Index	(Blue - Green) / (Blue + Green)	Median, Median_dry, Median_wet	Gamon et al., 1992
	CAI Cellulose Absorption Index	SWIR2 / SWIR1	Median, Median_dry, stdDev	Nagler et al., 2003
Vegetation Structure	Hall Cover	(- Red × 0.017 - NIR × 0.007 - SWIR2 × 0.079 + 5.22)	Median, stdDev	Hall et al., 2006
Fraction	GV Green Vegetation Fraction	Fractional abundance of green vegetation within the pixel	Median, Median_dry, Median_wet, stdDev	Souza et al., 2005
	GVS Green Vegetation Shade Fraction	GV / (GV + NPV + Soil + Cloud)	Median, Median_dry, Median_wet, Minimum, Maximum, Amplitude, stdDev	Housman et al., 2018

Type	Name	Formula / Description	Statistics	Reference
	NDFI Normalized Difference Fraction Index	(GVS - (NPV + Soil)) / (GVS + (NPV + Soil))	Median, Median_dry, Median_wet, Minimum, Maximum, Amplitude, stdDev	Souza et al., 2005
	NPV Non-photosynthetic Vegetation Fraction	Fractional abundance of non-photosynthetic vegetation within the pixel	Median, Median_dry, Median_wet, Minimum, Maximum, Amplitude, stdDev	Souza et al., 2005
	SEFI Savanna Ecosystem Fraction Index	(GV + NPV_S - Soil) / (GV + NPV_S + Soil)	Median, Median_dry, stdDev	Alencar et al., 2020
	Shade Fraction	100 - (GV + NPV + Soil + Cloud)	Median, Median_dry, Median_wet, Maximum, Minimum,, Amplitude, stdDev	Housman et al., 2018
	Soil Fraction	Fractional abundance of soil within the pixel	Median, Median_dry, Median_wet, Minimum, Maximum, Amplitude, stdDev	Souza et al., 2005
	WEFI Wetland Ecosystem Fraction Index	((GV + NPV) − (Soil + Shade)) / ((GV + NPV) + (Soil + Shade))	Median, Median_wet, Amplitude, stdDev	Rosa, 2020
	Slope			

3.4 Classification algorithm, training samples and parameters

Classification was performed by year and classification region, using the Random Forest algorithm (Breiman, 2001) available in the Google Earth Engine function *ee.Classifier.smileRandomForest()*. We used 200/300 trees per iteration and standardized the number of variables per split to the square root of the number of prediction variables. Training samples for each class were manually collected by the expert team for each classification region, using stable areas from 2000 to 2025.

Identifying stable areas for collecting manual samples involves drawing polygons around “stable samples” based on a minimum temporal frequency criterion, aiming to ensure confidence in using them as training samples across all years of the time series. Then, the collected samples per region were submitted to meticulous evaluation by at least one senior expert from each team to remove noise and misinterpretations. Finally, the samples were automatically filtered by region and class based on the distribution of median_ndvi, removing values more than 3 standard deviations from the mean.

This set of samples was used to generate preliminary maps of land cover and land use for the Democratic Republic of Congo from 2000 to 2025. Then, we used the preliminary classification result to perform a second round of classification, known in the literature as “step-wise classification”, by using the stable pixels from the preliminary collection as new training samples for the algorithm — that is, pixels that were classified with the same LULC class throughout the entire time series. A layer of pixels with a stable classification for each period was then generated. From the resulting layer of stable samples, a subset of 8,000 samples for each classification region was randomly generated, with a minimum number of 800 samples (10% of the total) for the rare classes. These samples were manually balanced per class based on local knowledge and reference maps (see the next step).

3.4.1 Sample size balance

We generated a variable number of samples for each class and classification region. These balancing values were based on local knowledge about the proportion of each LULC class in the region and on the interpretation of reference maps. Whenever the classification resulted in over- or underestimation of certain classes relative to supplemental information (local knowledge and reference maps), this proportion was adjusted to correct the bias. The regional balancing strategy is presented in Table 3.

Table 3. Number of training samples used for each year, class, and classification region in the MapBiomas Democratic Republic of the Congo Collection 1.

Region/ Class ID	3	4	5	66	21	9	24	25	33
Humid Zone	54.733	x	x	5.777	514	x	772	x	9.523
Miombo	x	254	x	72	105	x	101	184	197
Mosaic	19.538	20.295	3.535	62.980	44.427	2.395	12.597	165	19.709
Mountain	8.646	x	x	2.066	2.555	x	4.232	612	5.733
Terra Firma	8.872	9.880	x	9.190	10.076	831	5.698	1.526	4.397
Total	**91.789**	**30.429**	**3.535**	**80.085**	**57.677**	**3.226**	**23.400**	**2.487**	**39.559**

3.4.2 Complementary samples

The need to add complementary samples was evaluated by visually inspecting the output of a preliminary classification, using both Landsat and high-resolution images available in Google Earth Engine, as well as time series of vegetation indices (NDVI). Complementary sample collection was also done manually using new polygons in the Google Earth Engine Code Editor. All false-color images from the 26-year (2000-2025) Landsat mosaics and the vegetation index time series were checked against the selected new samples. Based on knowledge of each classification region, samples for different classes were collected to eliminate underrepresentation and overestimation of the mapped LULC classes.

3.4.3 Final classification

The final classification was performed for all classification regions and years, combining stable samples (derived from the preliminary classification based on manual samples) and complementary samples. For some years, the classification output produced anomalous results for certain classes. Therefore, it was necessary to improve the classification by applying post-processing spatio-temporal filters.

4. Post-classification

The results of the final classification were improved through a sequence of filters that corrected missing data, “salt-and-pepper” classification errors, and, especially, misclassifications. Temporal filters were applied to generate a more stable classification pattern over time, avoiding unexpected class variations over short time intervals.

4.1 Gap-fill filter

A filter to fill no-data pixels (“gaps”) was applied. Because, theoretically, no-data values are not allowed, they are replaced by the temporally nearest valid classification. In this procedure, if no “future” valid position was available, then the no-data value was replaced by its previous valid class. Therefore, gaps should only exist if a given pixel has been permanently classified as no-data throughout the entire temporal domain due to persistent cloud cover or noise.

4.2 Transition filter

A spatio-temporal transition filter was applied to reduce isolated and potentially spurious class changes between consecutive years. The filter identifies A–B–A transitions within a three-year temporal window, where a pixel changes from one broad land-cover group to another and returns to the original group in the following year. For this procedure, the land-cover classes were aggregated into native vegetation, anthropic or non-vegetated, and water.

Only transitions occurring in small spatial patches of up to six connected pixels (approximately 0.54 ha at 30 m spatial resolution) were corrected. In these cases, the intermediate-year classification was replaced by the original class observed in the previous year. This procedure reduces small, short-term temporal inconsistencies while preserving larger or spatially coherent land-cover changes. The first and last years of the time series were not modified because a complete three-year temporal window was not available.

4.3 Temporal filter

A temporal filter was applied to reduce short-term inconsistencies in the annual classification time series. The procedure combined general and class-specific temporal rules. First, a four-year moving window was used to identify A–B–B–A patterns, in which the classification remained in a different class for two consecutive years between equal classes. When this pattern was detected, the first intermediate year was replaced by the class observed in the previous year.

Subsequently, a three-year class-specific filter was applied to remove isolated one-year occurrences of selected classes. For each target class, when the class occurred only in the intermediate year and was absent in both the previous and following years, the intermediate classification was replaced by the class of the previous year. This correction was applied sequentially following the class priority order: River or Lake (33), Other Non-Vegetated Area (25), Urban Area (24), Forest Plantation (9), Pasture and Agriculture (21), Mangrove (5), Shrubland (66), Woodland (4), and Forest Formation (3).

Finally, a specific rule was applied to the first year of the time series. When the first year differed from a target class but the following two years were consistently classified as that class, the first year was replaced by the class observed in those two subsequent years. These procedures aimed to reduce short-duration classification fluctuations while maintaining temporally persistent land-cover changes.

4.4 Spatial filter

A spatial filter was applied independently to each annual classification to reduce small isolated patches and residual “salt-and-pepper” effects. Connected patches of the same class containing up to six pixels (approximately 0.54 ha at 30 m spatial resolution) were identified using eight-neighbor connectivity, in which pixels connected by either edges or corners were considered part of the same patch.

Pixels in these small patches were replaced by the modal class of their local 3 × 3 neighborhood. The River or Lake class (33) was excluded from this correction and preserved in its original classification to avoid the removal or alteration of small water features. No-data pixels were excluded from the final classification after the spatial operations. The filter was applied separately to each year of the time series and therefore did not modify classifications based on temporal information.

5. Map integration

The independently produced and post-processed annual land cover and land use classifications of the Miombo, Mosaic, Humid Zone, Mountain, and Terra Firma regions were combined to produce a set of annual maps for the Democratic Republic of Congo from 2000 to 2025. For each year, the regional maps were merged to produce an initial national integration.

Because each classification region was processed independently, the contact among regional maps could produce artificial thematic discontinuities along some regional boundaries. This was especially evident along the interface between the Miombo and Mosaic regions, where spatial patterns of vegetation classes change abruptly in the initial years of the mapping.

To reduce this effect, an additional localized Random Forest classification was performed along the Miombo–Mosaic interface. The procedure was designed to preserve the original regional classifications away from their boundary while allowing a data-driven transition to be reconstructed inside a spatial corridor surrounding the regional contact.

5.1 Miombo-Mosaic ecotone corridor

A transition corridor was created around the boundary between the Mosaic and Miombo regions. The corridor was defined as 60 km wide on each side of the boundary between the regions. Thus, the theoretical maximum width of this transition corridor was approximately 120 km, corresponding to 60 km toward the interior of the Mosaic region and 60 km toward the Miombo region. The resulting geometry is referred to here as the Miombo–Mosaic transition corridor. It represents the only area in which the additional Random Forest transition classification was allowed to influence the regional integration. The original classifications remained unchanged outside this corridor.

5.1.1 Definition of training areas on both sides of the ecotone corridor

Training samples were not collected directly inside the transition corridor. This was done to avoid using pixels located near the artificial regional boundary as reference information for the classifier. A 3-km exclusion zone was therefore created around the complete transition corridor. Pixels located within the corridor or within this additional 3-km buffer were excluded from the stable-pixel training dataset.

Outside this exclusion zone, local training areas were established independently on both sides of the interface. These areas extended approximately 150 km beyond the exclusion zone, providing a broad representation of the environmental and spectral conditions within the Miombo and Mosaic regions. The resulting training geometry therefore consisted of two spatially separated belts:

- a Miombo training belt, containing representative pixels from the Miombo side of the ecotone; and
- a Mosaic training belt, containing representative pixels from the Mosaic side.

This configuration was designed to expose the transition classifier to the spectral characteristics represented in both regional classifications without directly training it from the pixels that were subsequently reclassified within the ecotone corridor.

5.1.2 Selection of temporally stable reference pixels

The training dataset for the ecotone Miombo-Mosaic classification was derived automatically from temporally stable pixels in the final regional classifications. For each of the Miombo and Mosaic regions, the complete annual classification series from 2000 to 2025 was evaluated. A pixel was considered stable when two conditions were simultaneously satisfied:

1. valid classification information was available for all 26 years, and
2. only one unique LULC class occurred at that pixel throughout the complete 2000–2025 period.

The class assigned to these temporally stable pixels was used as the reference label for the ecotone classifier. Stable pixels from the Miombo and Mosaic regions were subsequently combined into a common training reference. Only stable pixels located within the previously defined training belts were retained. This procedure represents a second, localized classification stage conceptually similar to the step-wise strategy used during the main classification process. Instead of relying on a new manual interpretation of the transition zone, the method uses temporally consistent pixels from the already reviewed regional classifications as reference observations for an additional set of Random Forest models.

5.1.3 Annual Mosaic-Miombo ecotone classification

For each year, the corresponding Landsat mosaic and its predictor variables were associated with the stable training samples collected from both sides of the transition. A Random Forest classifier was subsequently trained using these annual predictor values and the stable LULC class as the response variable. In the ecotone integration step, the implemented Random Forest contained 200 decision trees. The trained model was then applied exclusively inside the Miombo–Mosaic transition corridor. This generated an annual transition classification, representing the class expected from the pixel's spectral characteristics when information from both Miombo and Mosaic stable areas was considered simultaneously. Importantly, this Random Forest classification did not replace the complete regional classification. It was used only as an alternative classification candidate inside the transition corridor.

5.1.4 Classes considered in the ecotone harmonization

Although the Random Forest classifier was generated from the locally occurring classes, the subsequent harmonization was restricted to the vegetation classes associated with the main thematic discontinuity between the Miombo and Mosaic regional products. The classes considered for replacement were:

- Forest Formation (3)
- Woodland (4)
- Shrubland (66)

An ecotone prediction was therefore considered a candidate for replacing the original regional classification only when both the original and predicted classes belonged to this vegetation group. Other classes, including anthropic, water, and non-vegetated classes, remained defined by the original regional integration. This restriction reduced the possibility that the regional integration procedure would introduce unrelated land-cover changes while addressing the specific vegetation discontinuity observed between Miombo and Mosaic.

5.2 Final annual integration

For each year between 2000 and 2025, the complete integration procedure consisted of:

1. integrating the independently classified regional maps;
2. generating the Miombo–Mosaic preliminary integration;
3. identifying the ecotone corridor;
4. extracting temporally stable reference pixels from the Miombo and Mosaic training belts;
5. associating the annual Landsat predictor variables with these stable samples;
6. training an annual Random Forest classifier;
7. predicting LULC classes inside the ecotone corridor;
8. replacing eligible pixels in the national integration where the ecotone classification provided sufficient evidence.

6. Data and code access

All the produced maps and statistics are free and available on the MapBiomas Democratic Republic of the Congo platform (https://www.plataforma.mapbiomas.org/projects/mapbiomas/democratic-republic-of-congo). You can freely use the data under the CC-BY license, just cite the main source as: "MapBiomas Project – Collection 1 of the Annual Land Use and Land Cover Maps of the Democratic Republic of Congo, accessed on [date] through the link: [LINK]". The source codes used to generate the maps are also public and freely available in the GitHub repository (https://github.com/mapbiomas/drc-all-initiatives).

7. Known limitations

The first limitation concerns the difficulty of discriminating between certain land cover classes. Forest formations, mangroves, and wetlands sometimes exhibit irregular trends in some regions, while bare soils can be confused with urban areas due to similar spectral signatures. Another limitation lies in the "other non-vegetated areas" class, which groups together distinct categories of non-vegetated lands, such as mining sites, airports, and urban zones. This grouping can mask specific dynamics. Finally, the presence of demarcation lines between Landsat WRS paths/rows during yearly mosaic creation can reduce image quality and affect classification interpretation.

8. Future developments

To improve the quality of land cover and land use (LCLU) maps and reduce inconsistencies observed in some classes, several development avenues will be explored in future versions of the product, particularly within the framework of Collection 2.

Improve discrimination between land cover classes: efforts will be made to strengthen the separation between classes with similar spectral signatures, such as natural forests, planted forests, mangroves, and wetlands, as well as between bare soils and built-up areas.

Refinement of the class nomenclature: The "other non-vegetated areas" class will be reassessed to explicitly distinguish certain specific categories, such as mining sites, airport infrastructure, sand strips, and low-density urban areas. This disaggregation will allow for a better representation of land-use dynamics and facilitate sectoral analyses.

Enhance the training data quality: The methodology for selecting stable pixels will be improved to limit biases introduced by temporal and spatial filters. Increasing the number of samples, improving their spatial distribution, and conducting more rigorous validation of reference data should enhance the robustness of the classification models.

Optimization of temporal processing: More advanced approaches for detecting and correcting temporal anomalies will be evaluated to better represent changes in land cover and land use to reduce fluctuations observed in some classes over the years.

Improve Landsat mosaics: The methods for creating image mosaics will be enhanced to reduce seams between scenes and improve radiometric homogeneity. This improvement will enhance the visual quality of the products and enable a more reliable interpretation.

Strengthen the validation and analysis of inconsistencies: The inconsistencies identified in the classes of water bodies, planted forests, and mangroves will undergo specific analyses. The results of these investigations will document the causes of the observed discrepancies and allow for the implementation of necessary corrections in Collection 2, with the aim of improving the temporal consistency and overall accuracy of the product.

Reorganization of the team into thematic groups: This step allows tasks to be divided among developers into thematic groups. Each group will be responsible for a single land cover class per biome, which helps to improve the accuracy and consistency of the results.

For Collection 1, all classes are grouped into a single map per year and biome. During the Collection 2 development, the classification will be refined by grouping it into thematic categories.

9. References

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

Breiman, L. (2001). Random forests. Machine Learning, 45, 5–32. https://doi.org/10.1023/A:1010933404324

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

Devred, R., 1960. La cartographie de la végétation au Congo belge. Extrait du Vol LI, No. 2 du Bulletin Agricole du Congo Belge et du Rwanda - Urundi – Bruxelles

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

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

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

Lebrun, J. et Gilbert, G., 1954. Une classification écologique des forêts du Congo. Publications de l'INEAC, Série scientifique No. 63 – Bruxelles

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

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

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

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

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