Amazon - Appendix

Collection 10

Version 1

General Coordinator
Luis Oliveira, *Msc*

Technical Coordinator
Bruno Ferreira

Team
Jailson Soares
João Victor Siqueira
Larissa Amorim, *Msc*
Camila Damasceno
Ives Brandão
Stefany Pinheiro
Raissa Ferreira
Bianca Nunes
Manoela Athaíde

1 Overview

The mapping of Amazon in the MapBiomas Project has evolved since the first Collection launched in 2015 (Table 1). Initially, the method used decision trees for image classification. From Collection 3/3.1 onwards, the Random Forest Classifier (RFC) was applied to build land use and land cover maps for the Amazon biome. Wetlands have been included since Collection 6 as a new class, utilizing a post-classification approach. Until Collection 3.1, we performed the classification using annual Landsat mosaics. Since Collection 4/4.1, we classified all available Landsat scenes (according to the established criteria) and then integrated the results to obtain the annual maps. This methodological change allowed us to assess each pixel's spectral variations within each year. For Collection 10, we made some corrections previously identified during the assessment of Collection 9 and improved our Rocky Outcrop and Floodable classes mapping to reconstruct the land use and land cover time series from 1985 to 2024 for the Amazon biome.

Table 1. The evolution of the Amazon mapping collections in the MapBiomas Project, its periods, mapped classes, a brief methodological description, and global accuracy in Levels 1 and 2.

Collection	Period	Mapped classes	Method/ Mapping Unit	Global Accuracy
Beta & 1	8 years 2008-2015	Forest; Non-Forest; Water Mask and Cloud Mask	Empirical Decision Tree / Annual Landsat Mosaic	
2.0 & 2.3	16 years 2000-2016	Non observed; Dense Forest; Inundated Forest, Degraded Forest; Secondary Forest; Nature Non-Forest Formations; Agriculture and Pasture; Non-Vegetated Areas; Water Surface; Unobserved	Empirical Decision Tree Random Forest (2.3) / Annual Landsat Mosaic	
3.0 & 3.1	33 years 1985-2017	Non observed; Forest Formation; Other Nature Non-Forest Formation; Mosaic of Agriculture and Pasture; Other Non-Vegetated Area; River, Lake and Ocean.	Random Forest / Annual Landsat Mosaic	Level 1: 95.1% Level 2: 95%
4.0 & 4.1	34 years 1985-2018	Non observed; Forest Formation; Other Non-Forest Natural Formation; Pasture; Agriculture; River, Lake and Ocean	Random Forest / All Selected Landsat Scenes	Level 1: 95.9% Level 2: 95.8%
5.0	35 years 1985-2019	Non observed; Forest Formation; Savanna Formation; Grassland Formation; Pasture; Agriculture; River, Lake and Ocean	Random Forest / All Selected Landsat Scenes	Level 1: 97.8% Level 2: 97.5%

6.0	36 years 1985 - 2020	Non observed; Forest Formation; Savanna Formation; Wetland; Grassland Formation; Pasture; Agriculture; River, Lake and Ocean	Random Forest / All Selected Landsat Scenes	Level 1: 97.4% Level 2: 96.7%
7.0	37 years 1985 - 2021	Non observed; Forest Formation; Savanna Formation; Wetland; Grassland Formation; Pasture; Agriculture; River, Lake and Ocean	Random Forest / All Selected Landsat Scenes	Level 1: 97% Level 2: 96.5%
7.1	37 years 1985 - 2021	Non observed; Forest Formation; Savanna Formation; Wetland; Grassland Formation; Pasture; Agriculture; River, Lake and Ocean	Random Forest / All Selected Landsat Scenes	Level 1: 97.6% Level 2: 96.8%
8	38 years 1985 - 2022	Non observed; Forest Formation; Floodable Forest; Savanna Formation; Wetland; Grassland Formation; Pasture; Agriculture; Rocky Outcrop; River, Lake and Ocean	Random Forest / All Selected Landsat Scenes and Annual Landsat Mosaic	Level 1: 97.7% Level 2: 96.9%
9	39 years 1985 - 2023	Non observed; Forest Formation; Floodable Forest; Savanna Formation; Wetland; Grassland Formation; Pasture; Agriculture; Rocky Outcrop; River, Lake and Ocean	Random Forest / All Selected Landsat Scenes and Annual Landsat Mosaic	Level 1: 97.7% Level 2: 96.8%
10	40 years 1985 - 2024	Non observed; Forest Formation; Floodable Forest; Savanna Formation; Wetland; Grassland Formation; Pasture; Agriculture; Rocky Outcrop; River, Lake and Ocean	Random Forest / All Selected Landsat Scenes and Annual Landsat Mosaic	Level 1: 97.5% Level 2: 96.5%

2 Landsat images

The MapBiomas Collection 10 generated annual Land Use and Land Cover maps for 40 years (1985 to 2024). All Landsat images available for this period (Landsat 5 [L5], Landsat [L7], Landsat 8 [L8], and Landsat 9 [L9]) were used with Cloud Cover (CC) less than or equal to 50%. The mapping unit for this collection is the Landsat path-row. Figure 1 shows the distribution of Landsat WRS-2 path-rows in the Amazon biome. The classification results were later integrated to fit the biome limits, mapping units used by the MapBiomas Initiative (Figure 1).

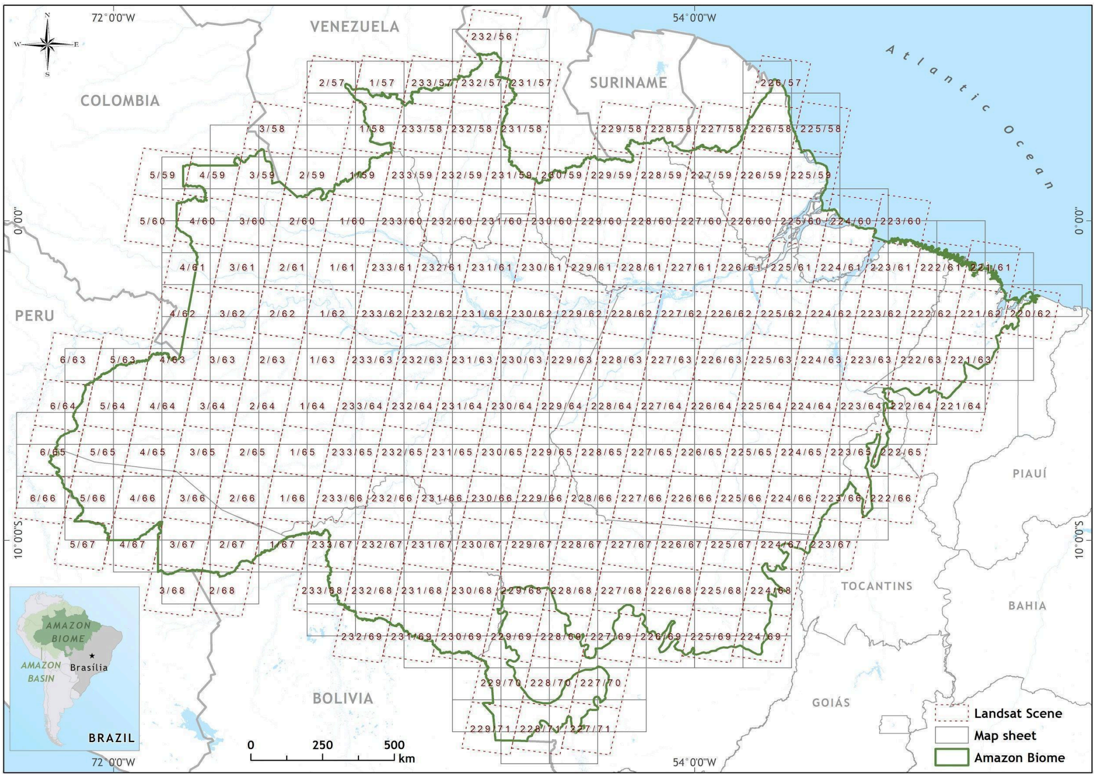

Figure 1. Distribution of Landsat path-rows for the MapBiomas Amazon biome.

A total of 201 path-rows cover the entire Amazon biome, representing over 97,000 Landsat images, considering all years of the time series. Figure 2 shows the number of images used each year by Landsat sensors for the Amazon biome.

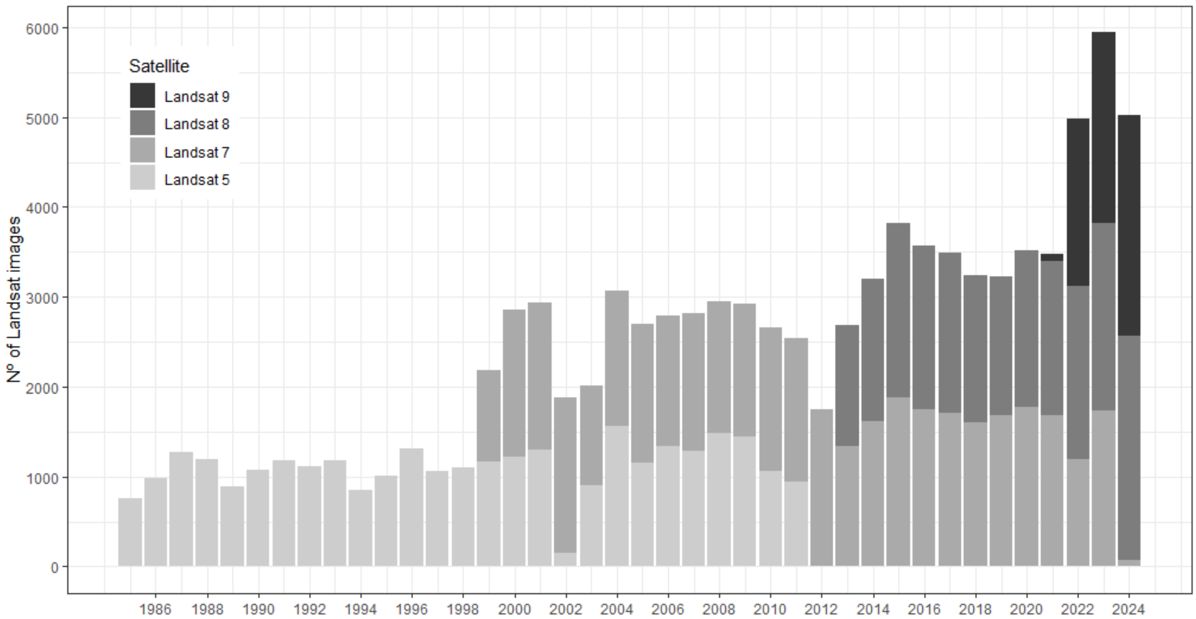

Figure 2. The number of Landsat images used per year and by Landsat sensors in the Amazon biome in Collection 10.

We also used Annual Median Landsat Mosaics to map the layers of Rocky Outcrop and Floodable Areas for the Amazon biome in the post-classification step described in the Classification session.

2.1 List of Landsat images removed from the database

We created a list of Landsat scenes that could contaminate the classification results using a Google Earth Engine (GEE) App that showed us a preview of the classification results for each image. The team selected the scenes to be removed visually. Images were removed for several reasons, including cloud cover, haze, no data, and Landsat 7 stripes.

3 Classification

The Collection 10 method had three main steps:

1) **Image Selection and Cloud/Shadow Masking**: We selected the Landsat 5, 7, 8, and 9 scenes filtered by the sensor, date range, and cloud cover. We applied the Temporal Dark Outlier Mask (TDOM) algorithm and the Band Quality Assessment (BQA) band available in the Landsat Collection.
2) **Random Forest Calibration, Training, and Image Classification**: We analyzed the best parameters to generate an optimized RFC in this step. We trained the RFC using the samples produced by LAPIG/UFG, plus a new pool of samples created to increase the number in areas with low sample density. We implemented a segmentation approach in the Landsat images to generate these samples and crossed the segments with the samples from LAPIG/UFG. For details about this new pool of samples, see section 3.3. Finally, we integrated the classification results in each path-row to generate the annual Land Use and Land Cover (LULC) maps;
3) **Post-classification**: Rocky Outcrops and Floodable Areas (Floodable Forest and Wetlands) were mapped and integrated into annual LULC maps to generate the final annual classification. Temporal and Frequency filters were applied to the annual maps. The last step was to integrate them with the cross-cutting themes and run the accuracy analysis.

Figure 3 shows the workflow to produce MapBiomas Collection 10 for the Amazon biome.

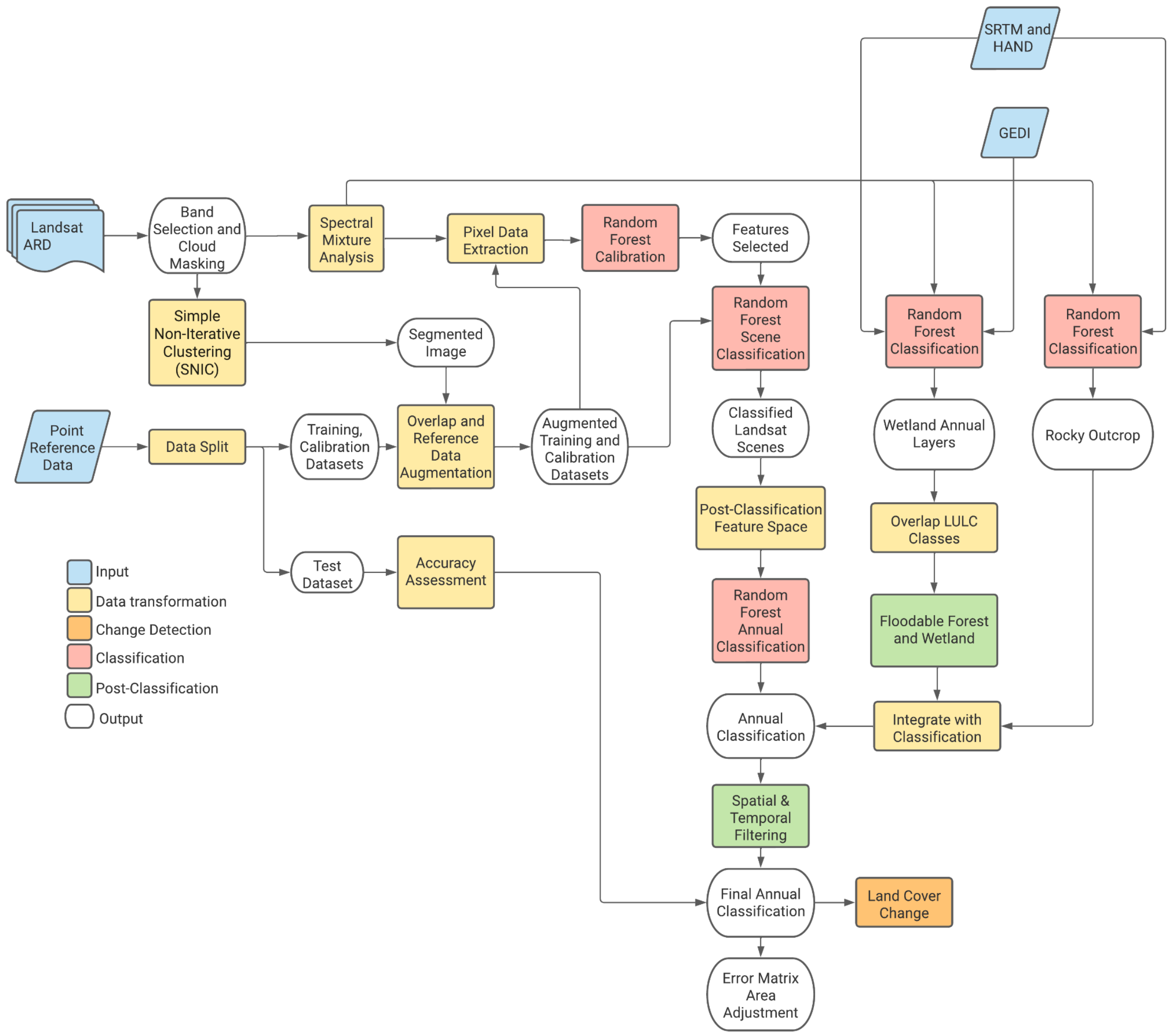

Figure 3. Classification process of Collection 10 in the Amazon biome.

3.1 Classification scheme

We mapped the same classes as the previous collection. Table 2 shows the ten classes mapped for Collection 10 in the Amazon biome.

Table 2. Classification scheme of Collection 10 for the Amazon biome (adapted from Souza Jr et al., 2023).

Value	Color	Color code	Class	Description
3		#1f8d49	Forest Formation	Vegetation types have a predominance of tree species with a high-density continuous canopy. It also includes mangrove and secondary regrowth and planted forests
4		#7dc975	Savanna Formation	Vegetation types with a tree layer varying in density, distributed over a continuous shrub-herbaceous layer
6		#026975	Floodable Forest	Forest areas permanently or temporarily flooded

11		#519799	Wetland	Shrubland and nature grassland permanently or temporarily covered by water
12		#d6bc74	Grassland Formation	Herbaceous vegetation, including patches with a well-developed shrub-herbaceous stratum
15		#edde8e	Pasture	Areas of natural or planted forest converted to farming activity
19		#E974ED	Agriculture	Areas are predominantly occupied by annual crops, and some regions are with perennial crops
25		#db4d4f	Other non Vegetated Areas	Impervious surface areas or exposed soil not mapped in other classes
29		#ffaa5f	Rocky Outcrop	Naturally exposed rocks in the terrestrial surface without soil cover, often with the partial presence of rock vegetation and high slope
33		#2532e4	River, Lake, and Ocean	Rivers, lakes, dams, reservoirs, and other water bodies

These classes are a subset of the MapBiomas classification system and were the primary input for classification integration with other cross-cutting themes and biomes (discussed in this document in the following sections).

For Collection 5, the class Other Non-Forest Formation (ONFF) was replaced by Savanna Formation (SF) and Grassland Formation (GF). We classified the Landsat images, including SF and GF samples, to map these classes in the Amazon/Cerrado ecotone. In areas outside of the Amazon/Cerrado ecotone, the class ONFF was replaced by GF, the most prevalent native vegetation class in these areas previously mapped as ONFF.

For Collection 6, we revisited the ONFF samples to separate SF from GF samples; this effort enabled the mapping of SF and GF classes for the entire biome. The 2020 LULC map was built using the updated samples and added to Collection 6. The next step was to select the path rows with pixels classified as ONFF (replaced by GF) from 1985 to 2019 in Collection 5 for reclassification using the updated samples.

For Collections 7 and 7.1, we used LULC maps from Collection 6 (1985 to 2020), reclassifying 10 path rows to improve the classification results. The 2021 LULC map was built and added to the other annual maps to complete the Amazon biome time series for the new collection.

For Collection 8, we used LULC mapping from the previous collection. In the post-classification step, we added the 2022 LULC map and integrated it with Rocky Outcrop and Floodable Forest classes as a cross-cutting theme.

For Collection 9, we revised the previous collection and added the 2023 LULC map. The main goal was to improve the mapping of the Savanna and Floodable classes in the Amazon biome. We also made some corrections in native vegetation areas with the commission from anthropic classes.

For Collection 10, we revised the previous collection, adjusting classification mistakes previously found by the Amazon team and other problems reported by users. We also added the 2024 LULC map to Collection 10. We expanded the Floodable classes in the Amazon biome and made some corrections in native vegetation areas with commission from anthropic classes. Figure 4 shows the 2024 LULC map from Collection 10:

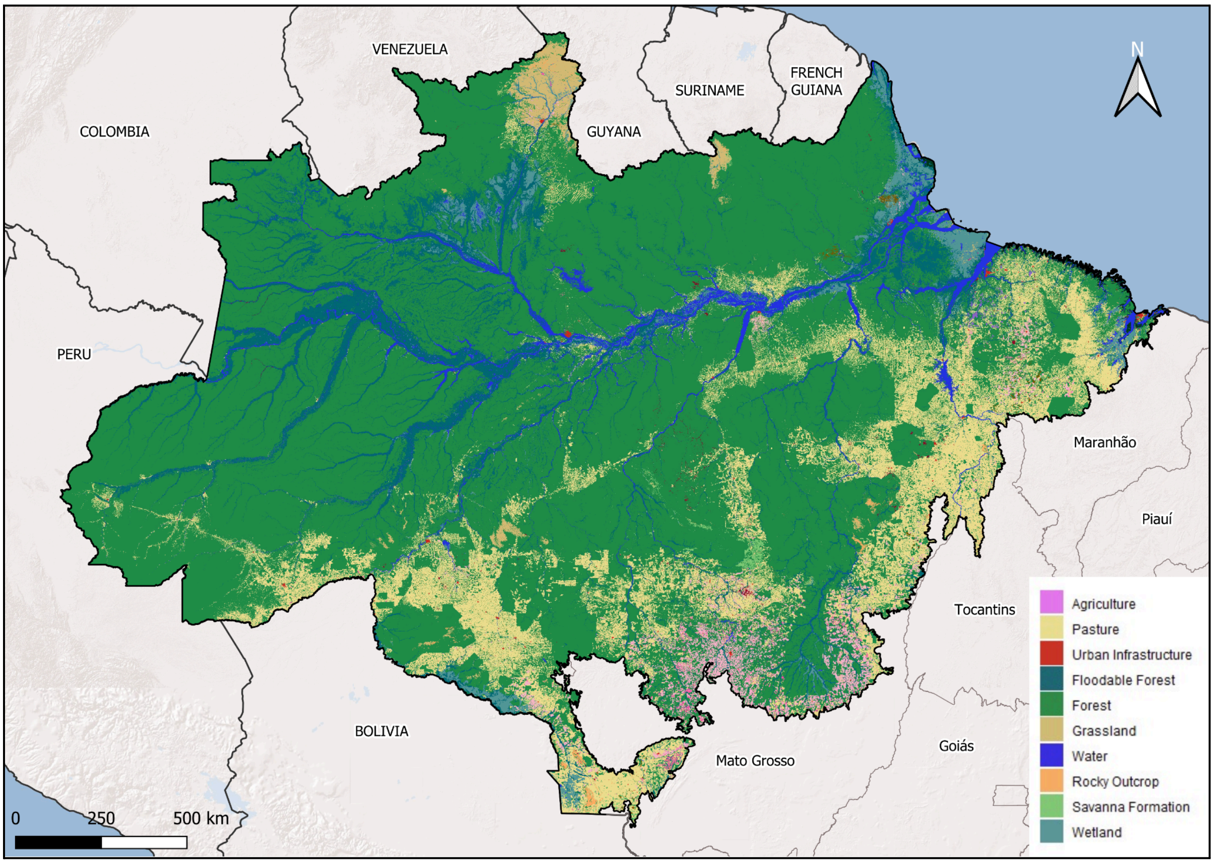

Figure 4. 2024 LULC map in the Amazon biome.

3.2 Feature space, classification algorithm, and training samples

The entire feature space produced for the MapBiomas Collection 10 was analyzed using 35,000 random points for the Amazon biome distributed over 40 years, obtained from the reference dataset provided by LAPIG/UFG in three phases:

- Phase I - 10,000 random samples used for algorithm training/calibration;
- Phase II - 10,000 random samples used for accuracy assessment;
- Phase III - 15,000 random samples used for accuracy assessment.

Statistical analysis defined the minimum number of samples to estimate the accuracy assessment of all Level 2 classes in the Amazon biome. Therefore, the reference dataset from LAPIG/UFG was split into two sets: training/calibration of the RFC (10k Phase I) and accuracy assessment (~25k Phase II + Phase III). The objective was to use the points collected in Phase I to identify the optimal features to be used in the RFC to reduce computational cost and allow a better understanding of the response of the spectral features to map the target classes.

The feature selection process was conducted in the R Language because GEE does not have specialized statistical libraries. We included products from Landsat images, such as the reflectance bands, spectral indices, and fractions from Spectral Mixed Analysis (SMA). Looking for the top results, we decided to use some highly important fractions and indices at the subpixel level.

The final feature space had eight variables, including Green Vegetation (GV), Non-Photosynthetic Vegetation (NPV), Soil, Cloud, Green Vegetation Shade (GVS), Normalized Difference Fraction Index (NDFI), Shade, and Canopy Shade Fraction (CSFI). These features were selected using the feature importance algorithm available in the R Language RFA implementation (Table 3). The metric used was the *Mean Decrease in Accuracy*, the default in the package.

Table 3. The feature space subset was used to classify the Amazon biome in Collection 10.

ID	Variable	Description
1	GV	gv fraction
2	NPV	npv fraction
3	SOIL	soil fraction
4	CLOUD	cloud fraction
5	GVS	gv normalized fraction
6	NDFI	normalized difference fraction index
7	SHADE	shade fraction
8	CSFI	canopy shade fraction index

3.3 Additional samples for Collection 10

In addition to the 10,000 samples produced by LAPIG/UFG used as a reference dataset in the RFC, we added new samples in the classification using the following approach:

Regionalized samples for the Amazon biome

To increase the number of samples in each Landsat scene and improve the RFA's training, we applied a segmentation technique called SNIC (Simple Non-Iterative Clustering) in all 97k images using six Landsat bands (red, green, blue, NIR, SWIR-1, SWIR-2). As a result, we have segmented images that were later crossed with the samples from LAPIG/UFG. The segment touched by the reference dataset was considered to belong to the same class and used to sort new samples (regional) randomly. We estimated the number of samples needed per class in each Landsat scene per year according to the class's proportion. The goal was to guarantee that RFA's training uses the maximum number of regional samples. If we don't reach the quantity necessary to classify the Landsat image, we also use the reference data from LAPIG. This approach was applied in the entire Amazon biome over the time series. Figure 5 shows how we add new regional samples to improve the RFA training.

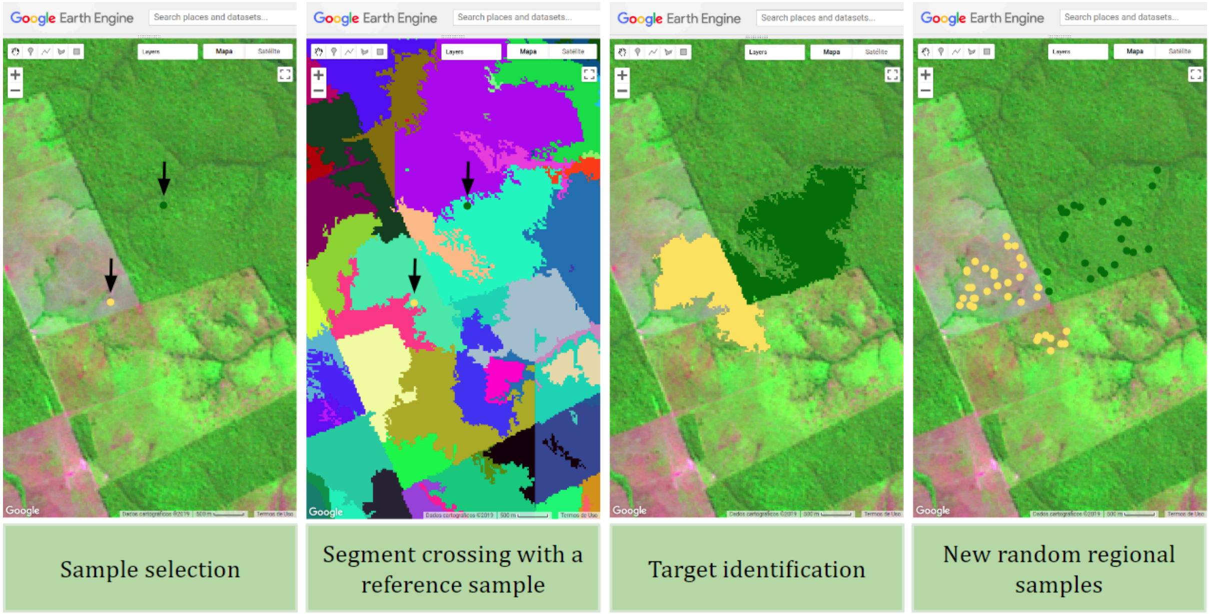

Figure 5. Four steps to create new random regional samples.

Additional Samples for Wetland and Rocky Outcrop Mapping

We mapped these two classes as a cross-cutting theme and overlapped them with the LULC mapping in the post-classification step. To map the moisture areas in the Amazon biome, we randomly sorted stratified samples (Wetland and Non-Wetland), using reference maps (Hess et al., 2015; Gumbricht et al., 2017; Tootchi et al., 2018) to indicate permanent or temporarily flooded areas and not flooded every year. For Rocky Outcrop mapping, we applied a visual inspection. Using visual analysis, we collected stratified samples (with Rocky Outcrop and Non-Outcrop) for one year. Rocky Outcrop has a static behavior over time, so we used a unique pool of samples for all years to train and calibrate a Random Forest model to map Rocky Outcrop annually.

3.4 Accuracy sensitivity to inspected parameters

A sensitivity analysis was run to evaluate the effect of input parameters of the RFA on per-class users' and producers' accuracies of the classification outputs. The results indicated that these metrics had low sensitivity to input parameters. Three parameters were used for the RFA: *ntree* (number of trees to be estimated), *mtry* (number of variables in each tree), and *nodesize* (size of the tree). The user's and producer's accuracies were estimated for each parameter to define values that optimize the computation time and accuracy. As a result, we defined a set of parameters that reduces the computational cost and increases the efficiency of the RFA. This analysis shows that the optimal values for the parameters were: *ntree* = 50, *mtry* = 7, and *nodesize* = 25.

3.5 Classification algorithm and training samples

The optimized version of RFA was implemented to produce Collection 10 using Google Earth Engine. The classifier's training dataset used 10,000 random samples from LAPIG/UFG, plus the additional samples described in section 3.3, collected for the Amazon biome. All the selected Landsat scenes were classified based on the RFA. Each year in the time series has 201 Landsat path-rows, and each Landsat path-row can have from 0 to 56 Landsat scenes, according to Landsat sensors overlapping, and 0 to 23 when only one is in operation (Figure 2).

3.6 Path-row integration and annual maps

For Collections 4 and 5, the annual classification for each path-row was defined using a statistical measure of central tendency named *mode* (the most frequent value in the observations) for each pixel. We also identified a set of post-classification rules (see Amazon ATBD Collection 5) to deal with some transitions not captured by mode in the time series. The union of all Landsat path-rows (mode product + post-classification rules) in the same year represents the LULC annual map.

For Collections 6, 7, 7.1, 8, 9, and 10, we calculate some metrics:

- Mode;
- Mode from the Wet season;
- Total Transitions: number of all class changes in the time series;
- Transitions per Year: number of class changes in each year;
- Total Distinct: number of different class changes in the time series;
- Distinct per Year: number of different class changes in each year;
- Grassland, Savanna, Agriculture, and Water Total Occurrence: occurrence of these classes in the time series;
- Forest, Grassland, Savanna, Pasture, Agriculture, and Water Occurrence per Year: The occurrence of these classes in each year.

Initially, the metrics were calculated to improve the post-classification rules, but at some point, these rules became so complex that new adjustments brought new challenges to the mapping. Therefore, we chose to use these metrics for training another round of RFA to integrate the classification results for each year and let the algorithm decide based on these metrics which class will prevail in the final map. This approach allows us to automate this step in the Amazon mapping classification process, avoiding subjectivity from post-classification rules in the results integration.

4 Post-classification

4.1 Moisture Areas and Floodable Forest Mapping

Using a post-classification approach, we added the Floodable Forest and Wetland for collection 10. We aimed to extract the Maximum Flooded Area (MFA) in the time series for the Amazon biome. First, we created annual mosaics and used reference maps to stratify samples of two classes: Wetlands and Non-Wetland. We used the Global Ecosystem Dynamics Investigation (GEDI), Shuttle Radar Topography Mission (SRTM), Height Above the Nearest Drainage (HAND), Canopy Height, and SMA fraction imagery dataset to train and calibrate an RF to map the Flooded area. The sampled pixels were automatically classified as a binary map, Wetland, and Non-Wetland. We used the trained and calibrated samples to classify the 40 annual mosaics. Finally, we analyze all 40 annual layers classified as wetlands and apply a maximum reducer to synthesize the layers and define the MFA in the time series for the Amazon biome. Every year, we cross the LULC map with the MFA layer. When the pixel agrees with Forest Formation and MFA, we remapped it as Floodable Forest. When the pixel agrees with the Savanna Formation or the Grassland Formation and MFA, we remapped it as Wetlands.

4.2 Rocky Outcrop Mapping

To map the Rocky Outcrops from the Amazon biome, we used the same annual mosaics described in section 4.1, plus random stratified samples (with Rocky Outcrop and Non-Outcrop) to train and calibrate an RF model to map Rocky Outcrop. The steep altitudes, slopes, escarpments, hills, and predominantly exposed soil give a unique spectro-temporal behavior to outcrops. To represent such features, we used fractions derived from spectral mixing models, such as Soil, NPV, and GV. Morphological characteristics of the terrain were represented using data such as SRTM and HAND.

4.3 Temporal filter

The temporal filter is a set of rules for non-allowed transitions applied to each image classified in a given year. That way, it was possible to remove clouds and correct non-allowed transitions. In the 3-year filter, when the initial and final year are in the same class and the intermediate year is in another class, the intermediate class is replaced by the class of the extremes. In the 5-year filter, when the intermediate year has a class that differs from the previous and two subsequent years, and all have the same class, it applies to the intermediate year. At the beginning and end of the time series, the filter is only applied in the case of non-observation, replacing this information with the class of the two subsequent years, or the two previous years, when it is the same class, respectively.

Fifty rules, distributed in three groups, were used: a) rules for cases not observed in the first year (RP); (b) rules for cases not observed in the final year (RU); (c) rules for examples of implausible transitions or not observed for intermediate years (RG) (Table 4).

Table 4. Temporal filter rules applied to Amazon Collection 10, RG = General Rule, RP = First-Year Rule, RU = Last Year Rule. Land Use and Land Cover classes, FF = Forest Formation, SF = Savanna Formation, GF = Grassland Formation, P = Pasture, AG = Agriculture, NO = Non-Observed, and W = Water.

rule	type	kernel	active	tminus2	tminus1	t	tplus1	tplus2	result
RG01	RP	3	1	null	NO	FF	FF	null	FF
RG02	RP	3	1	null	NO	SF	SF	null	SF
RG03	RP	3	1	null	NO	GF	GF	null	GF
RG04	RP	3	1	null	NO	P	P	null	P
RG05	RP	3	1	null	NO	AG	AG	null	AG
RG06	RP	3	1	null	NO	W	W	null	W
RG07	RU	3	1	null	FF	FF	NO	null	FF
RG08	RU	3	1	null	SF	SF	NO	null	SF
RG09	RU	3	1	null	GF	GF	NO	null	GF
RG10	RU	3	1	null	P	P	NO	null	P
RG11	RU	3	1	null	AG	AG	NO	null	AG
RG12	RU	3	1	null	W	W	NO	null	W
RG13	RG	3	1	null	FF	NO	FF	null	FF
RG14	RG	3	1	null	SF	NO	SF	null	SF
RG15	RG	3	1	null	GF	NO	GF	null	GF
RG16	RG	3	1	null	P	NO	P	null	P
RG17	RG	3	1	null	AG	NO	AG	null	AG
RG18	RG	3	1	null	W	NO	W	null	W
RG19	RG	3	1	null	FF	SF	FF	null	FF
RG20	RG	3	1	null	FF	GF	FF	null	FF
RG21	RG	3	1	null	FF	P	FF	null	FF
RG22	RG	3	1	null	FF	AG	FF	null	FF
RG23	RG	3	1	null	FF	W	FF	null	FF
RG24	RG	3	1	null	SF	FF	SF	null	SF
RG25	RG	3	1	null	SF	GF	SF	null	SF
RG26	RG	3	1	null	SF	P	SF	null	SF
RG27	RG	3	1	null	SF	AG	SF	null	SF
RG28	RG	3	1	null	SF	W	SF	null	SF
RG29	RG	3	1	null	GF	FF	GF	null	GF
RG30	RG	3	1	null	GF	SF	GF	null	GF
RG31	RG	3	1	null	GF	P	GF	null	GF
RG32	RG	3	1	null	GF	AG	GF	null	GF
RG33	RG	3	1	null	GF	W	GF	null	GF
RG34	RG	3	1	null	P	FF	P	null	P
RG35	RG	3	1	null	P	SF	P	null	P
RG36	RG	3	1	null	P	GF	P	null	P
RG37	RG	3	1	null	P	AG	P	null	P
RG38	RG	3	1	null	P	W	P	null	P
RG39	RG	3	1	null	AG	FF	AG	null	AG
RG40	RG	3	1	null	AG	SF	AG	null	AG
RG41	RG	3	1	null	AG	GF	AG	null	AG
RG42	RG	3	1	null	AG	P	AG	null	AG
RG43	RG	3	1	null	AG	W	AG	null	AG
RG44	RG	3	1	null	W	FF	W	null	W
RG45	RG	3	1	null	W	SF	W	null	W
RG46	RG	3	1	null	W	GF	W	null	W
RG47	RG	3	1	null	W	P	W	null	W
RG48	RG	3	1	null	W	AG	W	null	W
RG49	RG	5	1	FF	FF	SF	P	P	P
RG50	RG	5	1	FF	FF	GF	P	P	P

4.4 Frequency filter for native classes

A frequency filter was applied for the Amazon biome exclusively for the native vegetation classes: Forest Formation (FF), Savanna Formation (SF), and Grassland Formation (GF). If a pixel varied between these classes during the time series, the most frequent class would prevail, changing the classification in the years when that pixel was not classified as the most frequent class. The objective of the filter was a classification with more stable behavior between native classes. Other classes that may appear during the time series were not changed.

4.4 Additional filter

The 2024 map received special attention for errors in areas with consistent behavior along the time series. Savanna, Agriculture, and Pasture were targets for an additional filter that corrected the decrease of these classes in the last year. This filter uses the frequency of these classes in the time series to identify pixels with high frequency (more than 75%) and maintain the classification, avoiding undesirable transitions in the 2024 map.

4.5 Integration with cross-cutting themes

After applying the temporal filter, the products of digital classification for each of the 40 years from 1985 to 2024 were integrated with the cross-cutting themes by applying a set of specific hierarchical prevalence rules (Table 5). As the output of this step, a final land cover and land use map was obtained for each chart of the Amazon biome for each year; these annual maps may have up to 22 different classes.

Table 5. Prevalence rules for combining the output of digital classification with the cross-cutting themes in the Amazon biome in Collection 10.

Order	Class	Class ID	Source
1	Photovoltaic Power Plant	75	Cross-cutting Theme
2	Mining	30	Cross-cutting Theme
3	Beach and Dune	23	Cross-cutting Theme
4	Mangrove	5	Cross-cutting Theme
5	Aquaculture	31	Cross-cutting Theme
6	Salt Flat	32	Cross-cutting Theme
7	Water (Work Group)	33	Cross-cutting Theme
8	Urban Infrastructure	24	Cross-cutting Theme
9	Sugar Cane	20	Cross-cutting Theme
10	Soybean	39	Cross-cutting Theme
11	Rice	40	Cross-cutting Theme

12	Cotton	62	Cross-cutting Theme
13	Other Temporary Crops	41	Cross-cutting Theme
14	Perennial Crops	36	Cross-cutting Theme
15	Coffee	46	Cross-cutting Theme
16	Citrus	47	Cross-cutting Theme
17	Other Perennial Crops	48	Cross-cutting Theme
18	Temporary Crops	19	Cross-cutting Theme
19	Forest Plantation	9	Cross-cutting Theme
20	Rocky Outcrop	29	Biome
21	Other non Vegetated Area	25	Biome
22	River, Lakes and Ocean	33	Biome
23	Forest Formation	3	Biome
24	Floodable Forest	6	Biome
25	Savanna Formation	4	Biome
26	Wooded Restinga	49	Biome
27	Wetland	11	Biome
28	Grassland Formation	12	Biome
29	Herbaceous Sandbank Vegetation	50	Biome
30	Pasture	15	Cross-cutting Theme

5 Validation strategies

5.1 Accuracy Analysis

The second dataset, 25,000 reference samples collected by LAPIG/UFG, was used as a validation dataset. For validation, we calculated and reported confusion matrices, users', producers', and overall accuracies, the post-stratification class area estimates, and 95% confidence intervals for each statistic.

Global accuracy analysis has variations for MapBiomas Collections in the Amazon biome over time, but always with values above 97% at level 1 and 96% at level 2. Collection 5 (which first mapped the Savanna Formation and the Grassland for the entire biome) had the highest accuracy among the Amazon biome versions. Figure 6 shows the behavior of the accuracy analysis since Collection 5 for the Amazon biome integrated maps.

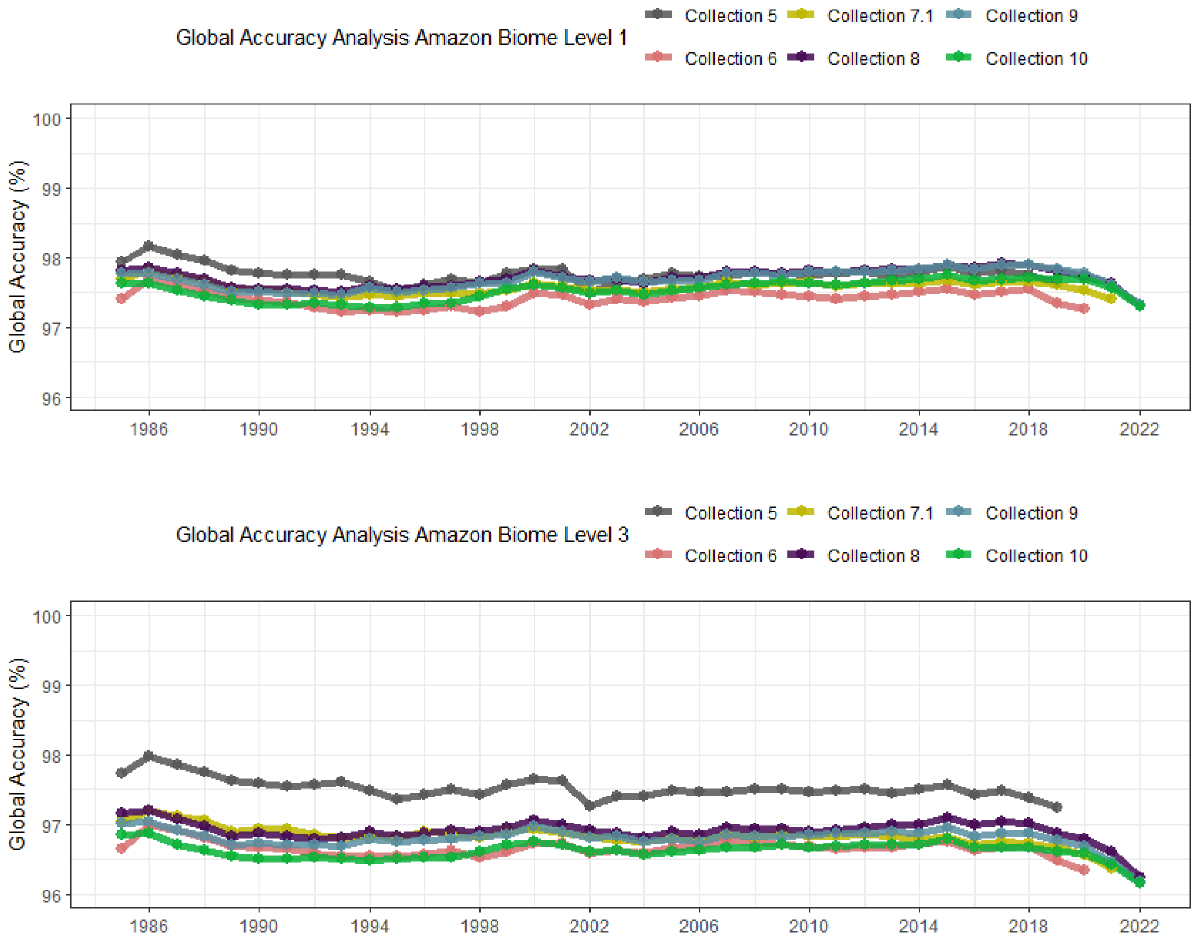

Figure 6. Accuracy analysis since Collection 5 for the Amazon biome (Level 1 and 3).

6 References

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

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

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

Housman, I.; Chastain, R.; Finco, M. 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 Sens. 2018, 10, 1184.

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

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

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

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