# Amazon - Appendix

## Collection 8

## Version 1


**General Coordinator**

Carlos Souza Jr., *Ph.D.*

**Team**

Luis Oliveira Jr., *Msc*
João Victor Siqueira
Jailson Soares
Bruno Ferreira
Larissa Amorim, *Msc*
Raissa Ferreira
Bianca Nunes

## 1 Overview

The mapping of the Amazon in the MapBiomas Project has been evolving through the Collections launched since 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 in the Amazon biome. Wetlands were included, since Collection 6, as a new class using a post-classification approach in the mapping. We classified all available Landsat scenes (according to the established criteria) and then integrated the results to obtain the annual maps. In previous collections, we performed the classification using annual Landsat mosaics. This methodological change allowed us to assess all the spectral variations contained in one year. For Collection 8, we added two new classes in the Amazon biome mapping: Floodable Forest and Rocky Outcrop. All scripts used to generate Collection 8 are available in [MapBiomas Amazon GitHub](https://github.com/mapbiomas-brazil/amazon).

**Table 1.** The evolution of the Amazon mapping collections in the MapBiomas Project, its periods, level and number of classes, 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; Pasture; Agriculture; River, Lake and Ocean | Random Forest / All Selected Landsat Scenes | Level 1: 97.6% Level 2: 97.5% |

| | | | | |
|---|---|---|---|---|
| 6.0 | 36 years 1985 - 2020 | Non observed; Forest Formation; Savanna Formation; Wetland; Grassland; Pasture; Agriculture; River, Lake and Ocean | Random Forest / All Selected Landsat Scenes | Level 1: 97% Level 2: 96.6% |
| 7.0 | 37 years 1985 - 2021 | Non observed; Forest Formation; Savanna Formation; Wetland; Grassland; 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; Pasture; Agriculture; River, Lake and Ocean | Random Forest / All Selected Landsat Scenes | Level 1: 96.8% Level 2: 95.9% |
| 8 | 38 years 1985 - 2022 | Non observed; Forest Formation; Floodable Forest; Savanna Formation; Wetland; Grassland; Pasture; Agriculture; Rocky Outcrop; River, Lake and Ocean | Random Forest / All Selected Landsat Scenes and Annual Landsat Mosaic | Level 1: 96.8% Level 2: 96.4% |

## 2 Landsat images

The MapBiomas Collection 8 generated annual maps of land use and land cover for 38 years (1985 to 2022). All Landsat images available for this period (Landsat 5 [L5], Landsat 7 [L7], Landsat 8 [L8], and Landsat 9 [L9]) were used with Cloud Cover (CC) less 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 with the mapping units used by the MapBiomas Initiative (Figure 1).

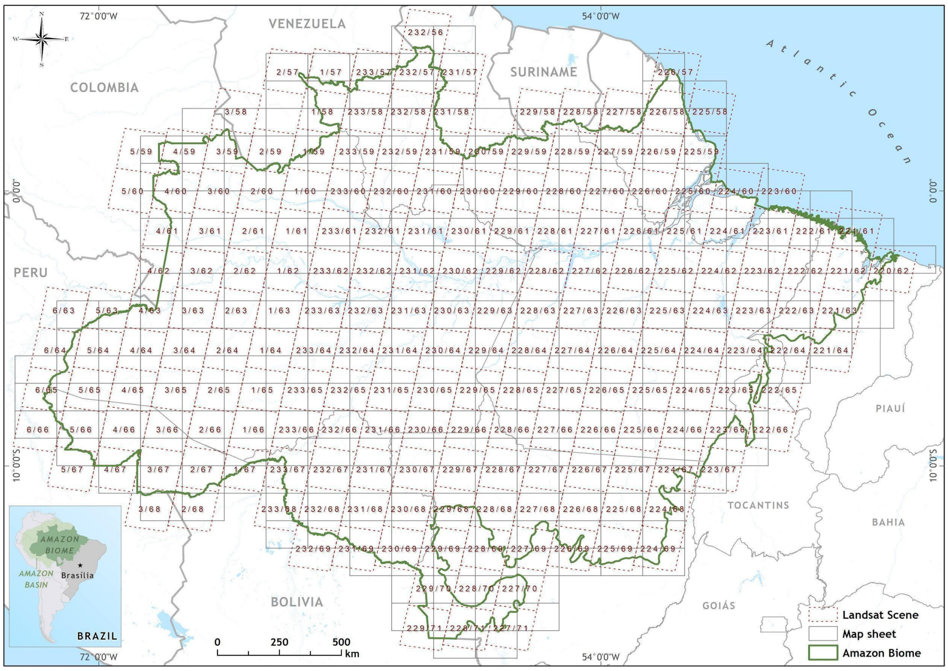


**Figure 1.** Distribution of Landsat path-rows for MapBiomas Amazon biome.

A total of 201 path-rows cover the entire Amazon biome, representing over 86,000 Landsat images in the time series. Figure 2 shows the number of images used each year by Landsat sensors for the Amazon biome.

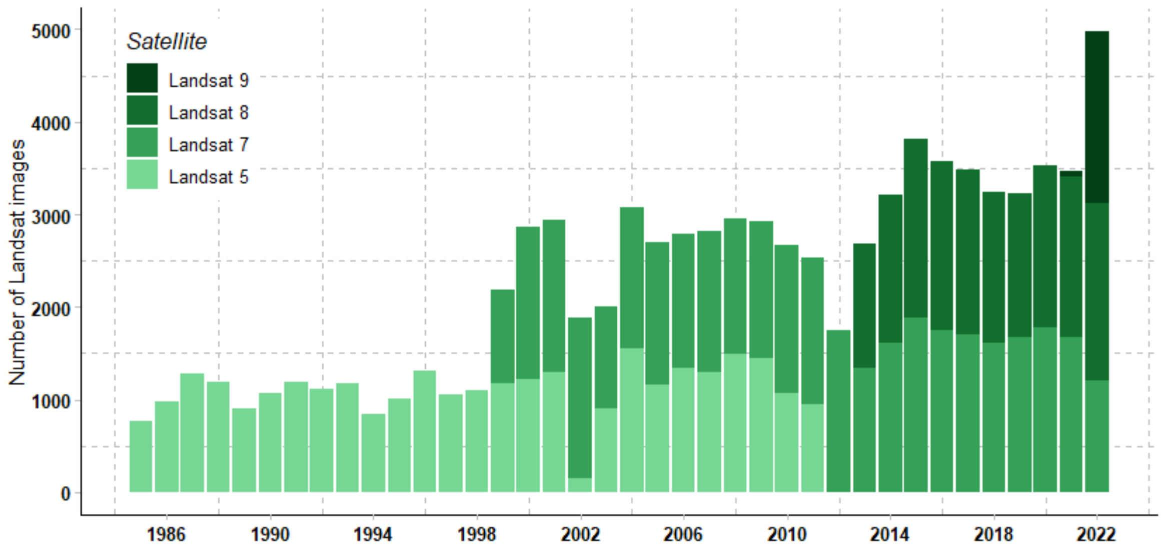


**Figure 2.** The number of Landsat images used per year and by Landsat sensors in the Amazon biome in Collection 8.

We also used Annual 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 (Gorelick, 2017) App that showed us the preview of classification results for each image. The removed scenes were selected visually by the team, the image can be removed for reasons like cloud cover, haze, no data, and Landsat 7 stripes.

# 3 Classification

The Collection 8 method had three main steps:

1) Image Selection and Cloud/Shadow Masking: We selected the Landsat 5, 7, 8, and 9 scenes filtering 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 for that purpose.
2) Random Forest Calibration, Training, and Image Classification: In that step, we ran an analysis to identify the best parameters to generate an optimized RFC. We trained the RFC using the samples produced by LAPIG/UFG plus new samples created by data augmentation analysis and classified all selected Landsat scenes. 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 Outcrop, Floodable Areas and Wetlands were mapped and integrated to annual LULC maps to generate the final annual classification. Temporal and Frequency filters were applied on 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 used to produce MapBiomas Collection 8 to Amazon biome.

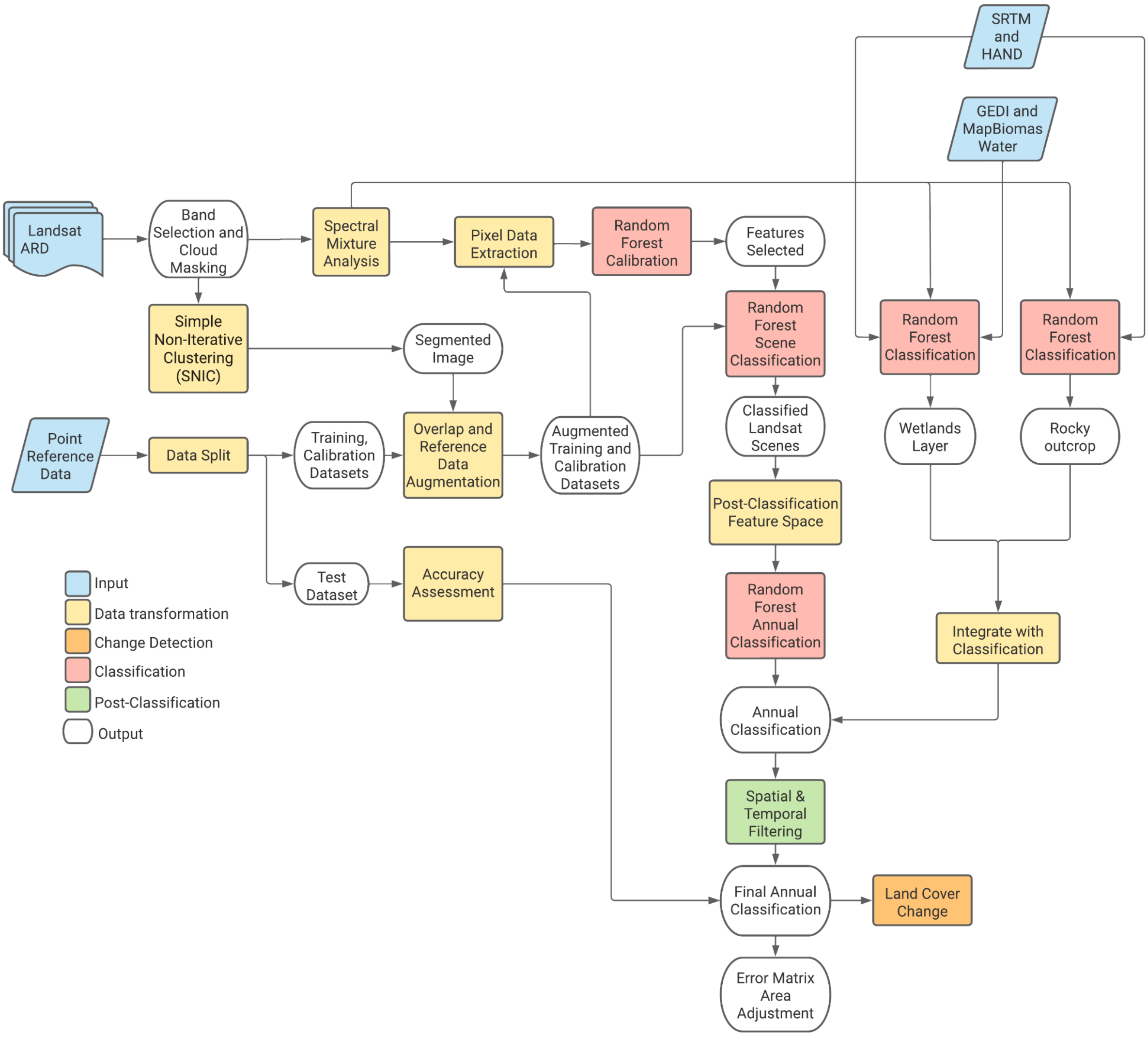


**Figure 3.** Classification process of Collection 8 in the Amazon biome.

### 3.1 Classification scheme

We mapped the same classes as the previous collection and added two new classes: Rocky Outcrop and Floodable Forest. Table 2 shows all classes mapped for Collection 8 in the Amazon biome.

**Table 2.** Classification scheme of Collection 8 for the Amazon biome.

| Value | Color | Color code | Class |
|---|---|---|---|
| 3 | | #006400 | Forest Formation |
| 4 | | #32CD32 | Savanna Formation |
| 6 | | #76A5AF | Floodable Forest |
| 11 | | #45C2A5 | Wetland |
| 12 | | #B8AF4F | Grassland |
| 15 | | #FFD966 | Pasture |
| 19 | | #E974ED | Agriculture |
| 27 | | #D5D5E5 | Non Observed |
| 27 | | #FF8C00 | Rocky Outcrop |
| 33 | | #0000FF | River, Lake, and Ocean |

These classes are a subset of the whole MapBiomas classification system and were the primary input for classification integration with other classes of cross-cutting themes and biomes (which is 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 (GF). We classified the Landsat images, including SF and GF samples to map these classes in the Amazon/Cerrado ecotone. In areas outside of Amazon/Cerrado ecotone, the class ONFF was replaced by GF, which is 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 that had pixels classified as ONFF (replaced by GF) from 1985 to 2019 in Collection 5 for reclassification using the updated samples.

For Collection 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 others annual maps to complete the Amazon biome time series for the new collection.

For Collection 8 we used LULC mapping from the previous collection adding the 2022 LULC map and integrated with Rocky Outcrop and Floodable Forest classes as a cross-cutting theme in the post-classification step. Figure 4 shows the 2022 LULC map.

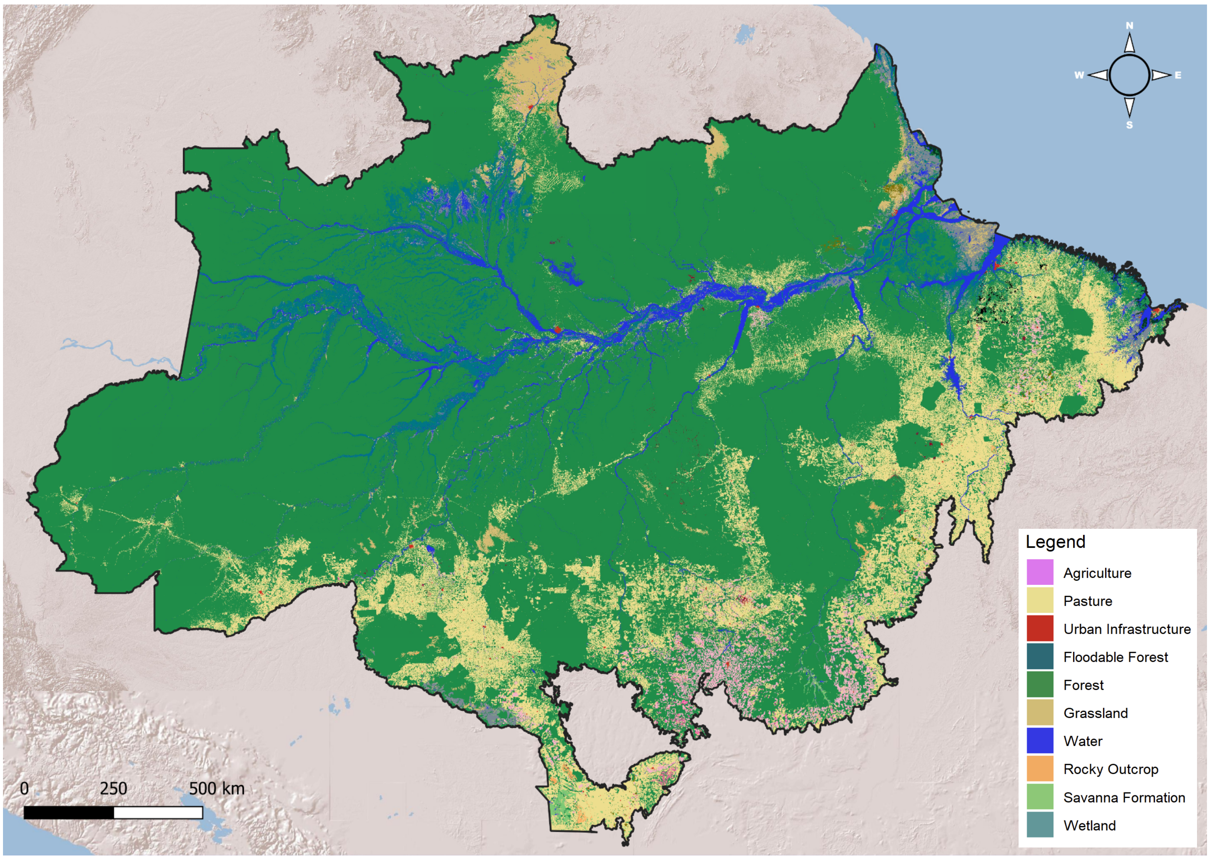


**Figure 4.** 2022 LULC map in the Amazon biome.

For more details about the description of classes mapped by MapBiomas Project see the document "Legend Description Collection 8" in MapBiomas website.

### 3.2 Feature space, classification algorithm, and training samples

The full feature space produced for the MapBiomas Collection 8 was analyzed using 35,000 random points for the Amazon biome, 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 was done to define the minimum number of samples to estimate the accuracy assessment of all Level 2 classes in the Amazon biome. Therefore, the full reference dataset from LAPIG/UFG was split by the pasture team into two sets: training/calibration of the RFC (10k Phase I), and accuracy assessment (~25k Phase II + Phase III). The objective was to identify the most 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 R Language because GEE does not have specialized statistical libraries. In this selection process, we included products from Landsat images like the reflectance bands, spectral indices, and fractions from Spectral

Mixed Analysis (SMA). Looking for the top results, we decided to use some fractions and indices that had high importance and were also in the subpixel level.

The final feature space ended up with 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 R Language RFA implementation (Table 3). The metric used was the *Mean Decrease in Accuracy*, the default in the package.

**Table 3.** Feature space subset used in the classification in the Amazon biome in the Collection 8.

| **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 8

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 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 86k images using six Landsat bands (red, green, blue, NIR, SWIR-1, SWIR-2). As a result, we have segmented images that later were crossed with the samples from LAPIG/UFG. The segment touched by the reference dataset was used to sort new samples (regional) randomly. We estimated the number of samples needed per class in each Landsat scene and we 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 use the reference data from LAPIG as well. This approach was applied in the entire Amazon biome through the time series. Figure 5 shows how we add new regional samples to improve the RFA training.

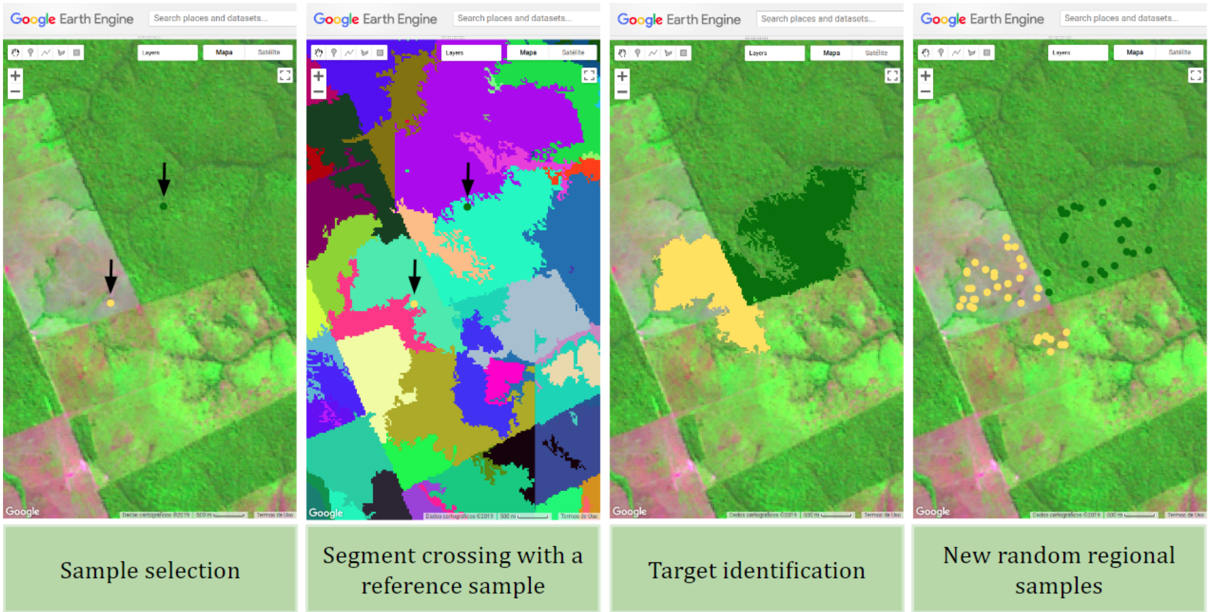


**Figure 5.** Four steps to create new random regional samples.

*Additional Samples for Wetland and Rocky Outcrop Mapping*

We mapped these two classes like a cross-cutting theme and added them to 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 to indicate permanent or temporarily flooded areas and not flooded every year. For Rocky Outcrop mapping, applying a visual inspection and selected stratified samples (with Rocky Outcrop and Non-Outcrop) 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 user's and producer's 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 of the parameters to define their 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 8 using GEE. 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 Collection 4 and 5 the annual classification for each path-row was defined using a statistical measure of central tendency named *mode* (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 Collection 6, 7, 7.1 and 8 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 differents class changes in the time series;
- Distinct per Year: number of differents 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: occurrence of these classes in each year.

Initially, the metrics were calculated to improve the post-classification rules, but at some point, these rules got so complex that new adjustments brought new challenges to the mapping. Therefore we opted to use these metrics for training another round of RFA to integrate the results of classifications 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 brought by post-classification rules in the results integration. Finally, Collection 8 shows the changes in the Amazon landscape over the past 38 years.

## 4 Post-classification

### 4.1 Moisture Areas and Floodable Forest Mapping

We added the Floodable Forest class for Collection 8 using a post-classification approach. First, we created annual mosaics and used reference maps to stratify samples. We used a monthly MapBiomas Surface Water, 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 wetland class. The sampled pixels were automatically classified as a binary map, Wetland, and Non-Wetland. We used the trained and calibrated samples to rank the 38 annual mosaics. Finally, we analyze all 38 annual layers classified as wetlands and apply a maximum reducer to synthesize the layers and define the Maximum Flooded Area (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 Savanna Formation or Grassland 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 and 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, 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. A number of 50 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 8 Land Use and Land Cover classes. RG = General Rule, RP = First-Year Rule, RU = Last Year Rule, FF = Forest Formation, SF = Savanna Formation, GF = Grassland Formation, P = Pasture, AG = Agriculture, NO = Non-Observed, 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/Cerrado ecotone region exclusively for the native vegetation classes: Forest Formation (FF), Savanna Formation (SF), and Grassland (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.

### 4.5 Additional filters

The 2022 map received special attention for errors in areas with consistent behavior along the time series. Savanna, Agriculture, and Pasture were a target for an additional filter that corrected the decrease of these classes in the last year. We also created a five year moving window to stabilize Pasture areas with incorrect oscillation. This filter was applied from 2009 to 2022 and allowed smoothing of the Pasture increment concentrated in a few years in the last decade of the mapping.

### 4.6 Integration with cross-cutting themes

After applying the temporal filter, the products of digital classification for each of the 38 years in the period 1985-2022 were then 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.

**Table 5.** Prevalence rules for combining the output of digital classification with the cross-cutting themes in the Amazon biome in Collection 8.

| Order | Class | Class ID | Source |
|---|---|---|---|
| 1 | Mining | 30 | Cross-cutting Theme |
| 2 | Beach, Dune and Sand Spot | 23 | Cross-cutting Theme |
| 3 | Mangrove | 5 | Cross-cutting Theme |
| 4 | Aquaculture | 31 | Cross-cutting Theme |
| 5 | Hypersaline Tidal Flat | 32 | Cross-cutting Theme |
| 6 | Water (Work Group) | 33 | Cross-cutting Theme |
| 7 | Urban Infrastructure | 24 | Cross-cutting Theme |
| 8 | Sugar Cane | 20 | Cross-cutting Theme |
| 9 | Soybean | 39 | Cross-cutting Theme |
| 10 | Rice | 40 | Cross-cutting Theme |

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

# 5 Validation strategies

## 5.1 Accuracy Analysis

The second dataset of ~25,000 reference samples, collected by LAPIG/UFG, was used for the validation dataset. For validation, we calculated and reported confusion matrices, user's, producer's, and overall accuracies, as well as the post-stratification class area estimates, along with 95% confidence intervals for each statistic.

The global accuracy analysis has increased over MapBiomas Collections in the Amazon biome. Collection 5 had the highest accuracy among the Amazon biome versions. Collections 6, 7, 7.1 and 8 (which first mapped Savanna Formation and Grassland for the entire biome) come next in the accuracy levels. Figure 6 shows the behavior of accuracy analysis since Collection 3.1 for the Amazon biome integrated maps.

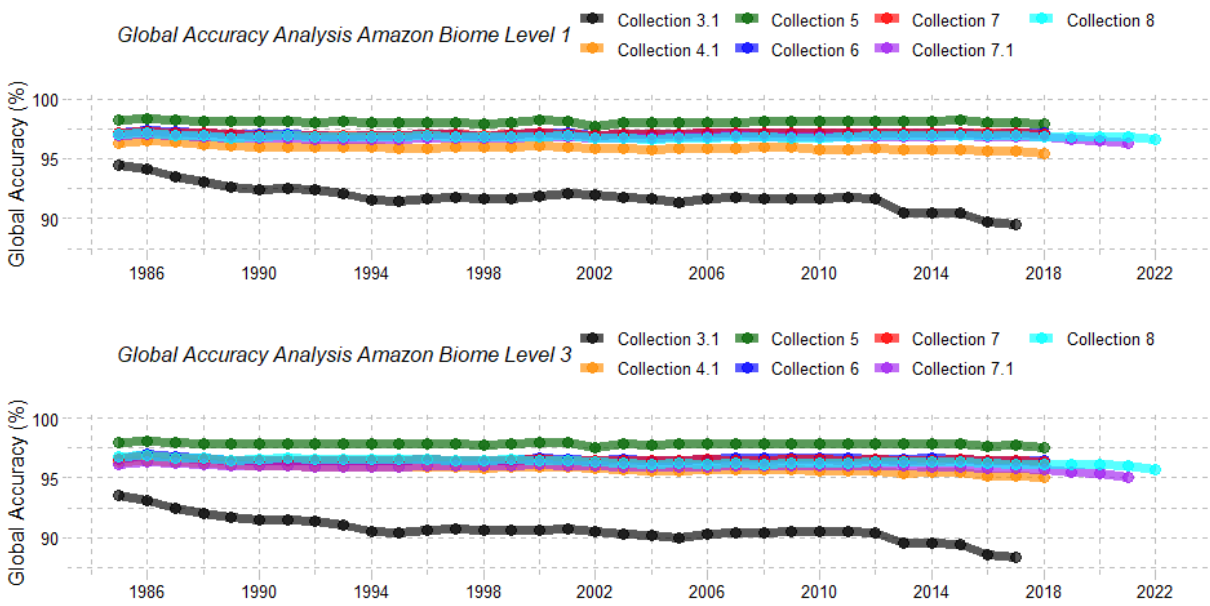


**Figure 6.** Accuracy analysis since Collection 3.1 for Amazon biome (Level 1 and 3).

## 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.

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.

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.