

MapBiomas Uruguay

Collection 4

Algorithm Theoretical Basis Document (ATBD)

2026

General Coordinator
Tasso Azevedo

Country coordinators
Santiago Baeza
Federico Gallego

Uruguay team
Andrea Barbieri
Laura Bruzzone
Sebastián Ramos
Gonzalo Rama
Pablo Baldassini

1 INTRODUCTION

1.1 Scope and content of the document

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 Uruguay from 1985 to 2025 (Collection 4). The document presents a general description of the satellite image processing, the feature inputs and the process step by step applied to obtain the annual classifications.

1.2 Region of Interest

MapBiomas Uruguay initiative was created to produce LULC annual maps for the Uruguay territory (**Figure 1**). The total mapped area was 17,8 million hectares (Mha)

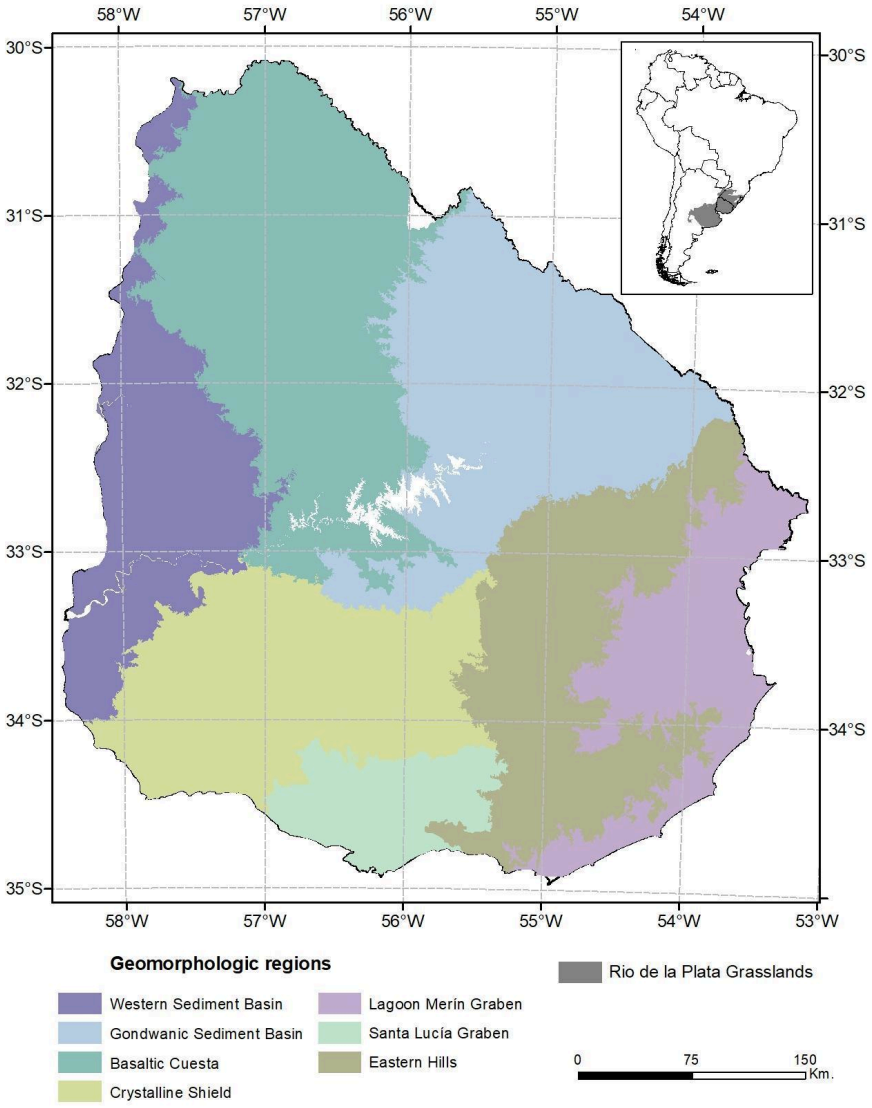


Figure 1. Location of Uruguay within the Rio de la Plata grasslands biome and the regionalisation used in the classifications (Geomorphological regions proposed by Panario et al., 2014).

2 GEOGRAPHICAL UNITS OF CLASSIFICATION

The classification process was carried out in smaller and homogeneous zones spatial units. These units correspond to seven geomorphological regions (Panario et al. 2014) (Figure 1). The purpose of these geographical units of classification was to try to reduce samples and classes confusion and to allow a better balance of samples and results to improve accuracy.

3 REMOTE SENSING DATA

3.1 Landsat Collection

The imagery dataset used in the *MapBiomas Uruguay (LULC),* Collection 4 was obtained from the Landsat sensors Thematic Mapper (TM), Enhanced Thematic Mapper Plus (ETM+) and the Operational Land Imager and Thermal Infrared Sensor (OLI-TIRS), on board of Landsat 5, Landsat 7 and Landsat 8, respectively. The Landsat imagery collections with 30 m-pixel resolution were accessible via Google Earth Engine and were provided by NASA and USGS. The *MapBiomas Uruguay* Collection 4 used Collection 2, Tier 1 Landsat Surface Reflectance products from USGS, which underwent through radiometric calibration and orthorectification correction based on ground control points and digital elevation model to account for pixel co-registration and correction of displacement errors. A total of 18 scene boundaries were used to cover the entire region, where each of them is totally or partially within the area.

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

- from 1985 to 1999: Landsat 5,
- year 2000: Landsat 5 (Brazil and Uruguay) and Landsat 7 (Argentina),
- years 2001, 2002 and 2012: Landsat 7,
- from 2003 to 2011: Landsat 5,
- from 2013 to 2025: Landsat 8.

3.2 Landsat Mosaics

All Landsat scenes were merged and clipped within standardized spatial units for data processing, hereafter called ‘charts’, based on the grid of the World

International Chart to the Millionth, at the 1:250,000 scale level. A total of 19 charts were used to cover the biome (**Figure 2**). Each chart sets the geographical limits to build up the temporal and spatial Landsat mosaics and to proceed with digital classification procedures. Each geographical classification unit was generated by merging the correspondent mosaic charts.

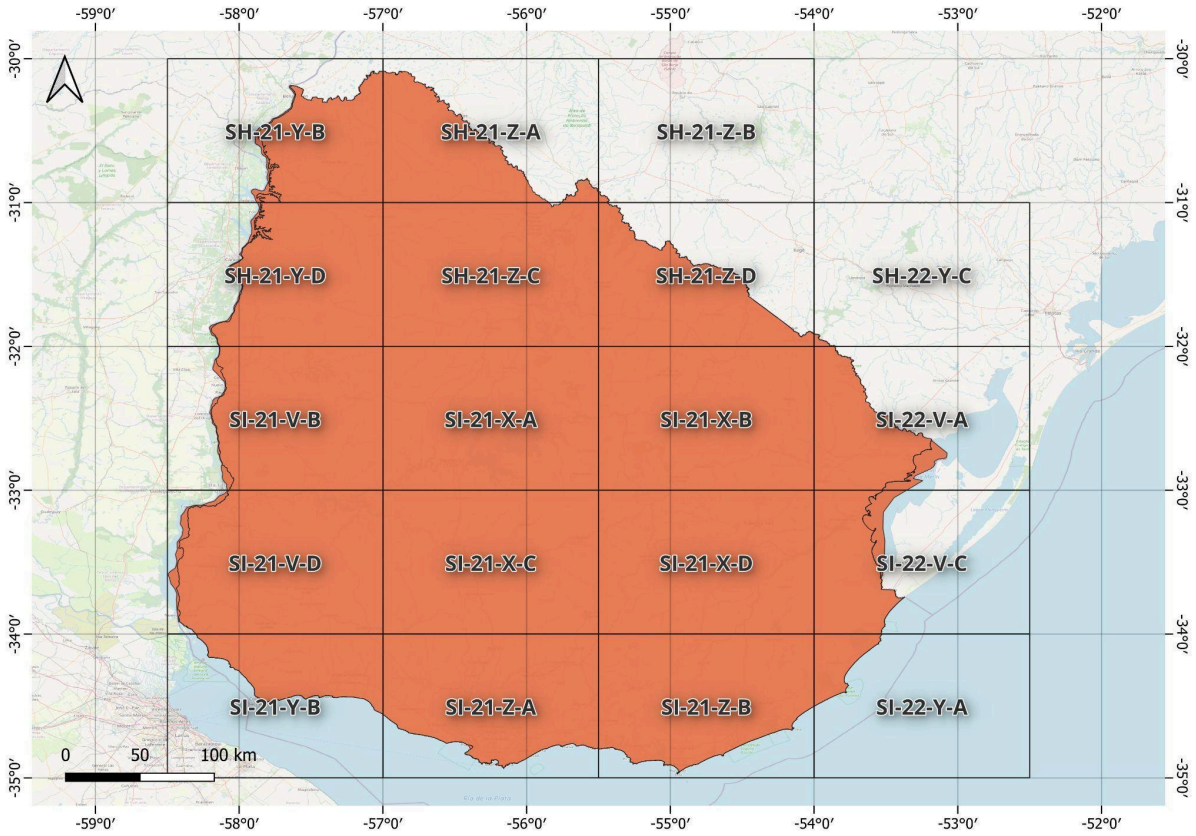


Figure 2. Charts scheme used to build up Landsat mosaics used throughout the classification process.

3.3 Definition of the temporal period

The mosaics were formed by the composition of pixels in each set of images for a certain time period. The periods of the year in which the images are selected vary by country and result from the balance between the probability of maximizing the differences in classes spectral behavior and the availability of cloud-free images. The considered period was from September to November of each year. Nevertheless, for some years this period was adapted (extended one to three months) for each chart according to the availability of cloud-free images. For example, if during the three-months period a cloud free mosaic could not be generated, the period was extended to four, five or six months to get a complete or almost complete mosaic.

For the selection of Landsat scenes a threshold of 90% of cloud cover was applied (i.e., any available scene with up to 90% of cloud cover was accepted). This limit was established based on a visual analysis, after many trials observing the results of the cloud removing/masking algorithm. Time periods were extended for some years and portions of the study area when the availability of cloud-free images was low.

4 CLASSIFICATION

4.1 Overview of methodological process

The methodological procedures of Collection 4 included several steps (**Figure 3**).

The first step was to generate annual Landsat image mosaics based on yearly periods. The second step was to generate a new selection of temporally stable samples derived from the stable areas of the maps of Collection 3. Stable areas were defined in sub-periods of near 5 years-length (1985-1989; 1990-1994;….; 2020-2024). Then, the spectral feature inputs derived from the Landsat bands were extracted and associated to each sample point. Once the samples for each LULC class were selected for each of the subregions, it was possible to adjust the training data set according to its statistical needs. The number of samples for training for each class was defined initially according to the proportion of the area of each class and its variation along the classification period (sample size balance). Additionally, to improve the classification results, complementary samples were generated, defining georeferenced points of different classes by visual interpretation of historical satellite images (high and very high-resolution images) and time series of vegetation indices. Based on the adjusted training data set, a supervised classification using the random forest algorithm was run.

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 4. The MapBiomas Uruguay annual LULC maps were used to derive the transition analysis (with an additional spatial filter application) and statistics. The statistical analysis covered different spatial territories, such as countries, state similar and municipality similar levels of each country, water basin and phytogeographic provinces.

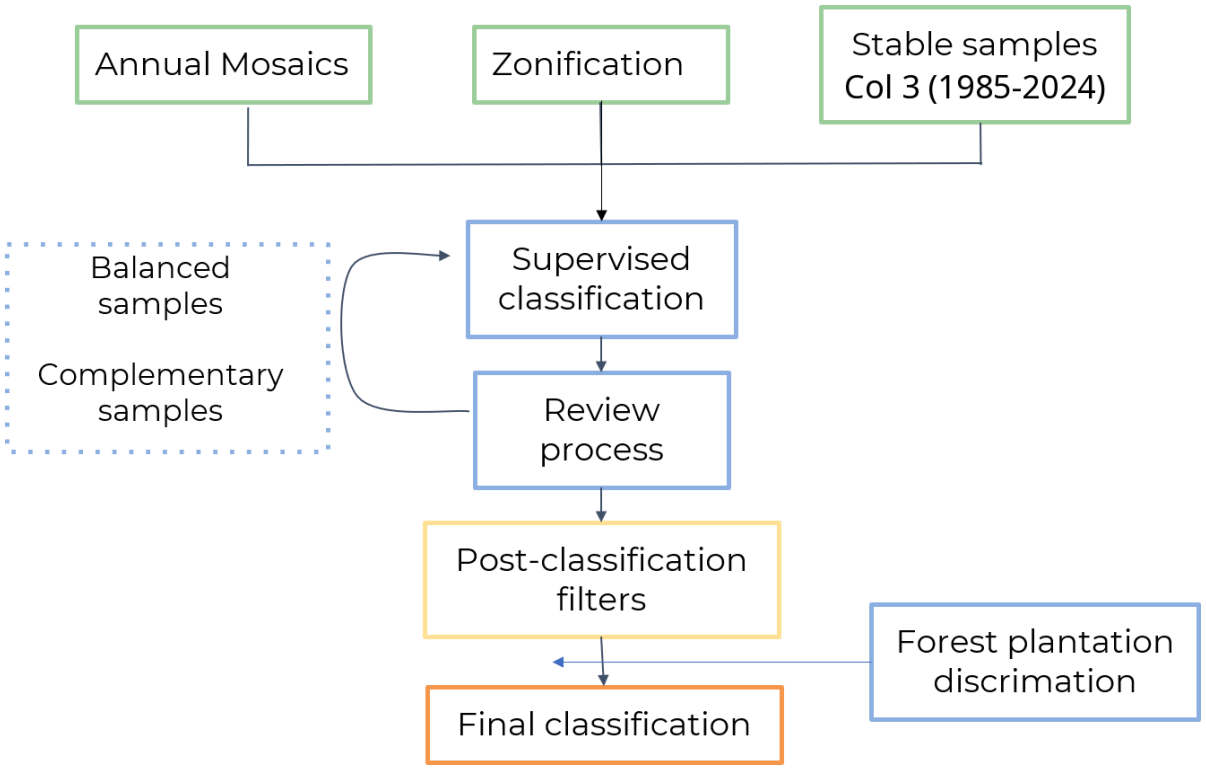


Figure 3. Classification process of Collection 4 in the *MapBiomas Uruguay* for the period 1985-2025

4.2 Map Legend

The classification for the *MapBiomas Uruguay* initiative using Landsat mosaics included eleven land use and land cover (LULC) classes (**Table 1**): Closed forest and closed shrubland (3), Flooded grassland and swampy areas (11), Grassland (12), Pasture (15), Agriculture (18), Pinus plantation (79), Eucalyptus plantation (80), Other forestry uses (83), Urban areas (24), Other non vegetated area (25), River, lake or ocean (33) and Not observed (27). A full description of the legend is described in the document Legend Description.

Table 1. Land cover and land use classes considered for digital classification of Landsat mosaics for the MapBiomas Uruguay initiative - Collection 4.

COLLECTION 4 - CLASSES	Numeric id	Hexadecimal code	Color
1. Natural woody vegetation	**1**	**#1f8d49**	
1.1. Closed forest and closed shrubland	3	#1f8d49	
2. Natural herbaceous vegetation	**10**	**#d6bc74**	
2.1. Wetland	11	#519799	
2.2. Grassland	12	#d6bc74	
3. Agricultural and livestock area	**14**	**#ffefc3**	
3.1. Agriculture	18	#C27BA0	
3.2. Pasture	15	#edde8e	
3.3. Forest plantation	9	#7a5900	
3.3.1 Pinus plantation	79	#69601e	
3.3.2 Eucaliptus plantation	80	#7a6c00	
3.3.1 Other types of forest plantation	83	#948618	
4. Non-vegetated area	**22**	**#d4271e**	
4.1 Urban area	24	#d4271e	
4.2 Other non-vegetated areas	25	#db4d4f	
5. Water body	**26**	**#2532e4**	
5.1. River, lake or ocean	33	#2532e4	
6. Not observed	**27**	**#ffffff**	

4.3 Annual Mosaics

The total available bands of the MapBiomas Uruguay feature space is composed of 93 input variables, including the original Landsat bands, fractional and textural information derived from these bands (**Table 2**). Reducers were used to generate temporal features such as:

● Median: median of the pixel values of the best mapping period defined by each country.

● 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 of the index considering all the images of each year.

● stdDev: standard deviation of all pixel values of all images of each year.

● Min: lower annual value of the pixels of each band.

Table 2. List of the variables included in the feature space used in the classification processes of the MapBiomas Uruguay Initiative Collection 4 (1985-2025).

D	Variable	Description	Statistics	Temporal range	Script acronym	Group
0	Evi 2	Enhanced Vegetation Index 2	amplitude	mosaic months	evi2_amp	Spectral index
1	Gv	Green vegetation fraction	amplitude	mosaic months	gv_amp	Spectral Mixture Modeling
2	Ndfi	Normalized Difference Fraction Index	amplitude	mosaic months	ndfi_amp	Spectral Mixture Modeling
3	Ndvi	Normalized Difference Vegetation Index	amplitude	mosaic months	ndvi_amp	Spectral index
4	Ndwi	Normalized Difference Water Index	amplitude	mosaic months	ndwi_amp	Water Index
5	Soil	Soil fraction	amplitude	mosaic months	soil_amp	Spectral Mixture Modeling
6	Wefi	Woodland ecosystem fraction index	amplitude	mosaic months	wefi_amp	Fraction index
7	Blue	Landsat band	median	mosaic months	blue_median	Landsat band
8	Blue dry	Landsat band	median	year -first quartile	blue_median_dry	Landsat band
9	Blue wet	Landsat band	median	year – fourth quartile	blue_median_wet	Landsat band
10	Cai	Cellulose Absorption Index	median	mosaic months	cai_median	Spectral index
11	Cai dry	Cellulose Absorption Index	median	year -first quartile	cai_median_dry	Spectral index
12	Cloud	Cloud fraction	median	mosaic months	cloud_median	Spectral Mixture Modeling
13	Evi 2	Enhanced Vegetation Index 2	median	mosaic months	evi2_median	Spectral index
14	Evi 2 dry	Enhanced Vegetation Index 2	median	year -first quartile	evi2_median_dry	Spectral index
15	Evi 2 wet	Enhanced Vegetation Index 2	median	year – fourth quartile	evi2_median_wet	Spectral index
16	Gcvi	(nir/green − 1)	median	mosaic months	gcvi_median	Spectral index

D	Variable	Description	Statistics	Temporal range	Script acronym	Group
17	Gcvi dry	(nir/green − 1)	median	year -first quartile	gcvi_median_dry	Spectral index
18	Gcvi wet	(nir/green − 1)	median	year – fourth quartile	gcvi_median_wet	Spectral index
19	Green	Landsat band	median	mosaic months	green_median	Landsat band
20	Green dry	Landsat band	median	year -first quartile	green_median_dry	Landsat band
21	Green wet	Landsat band	median	year – fourth quartile	green_median_wet	Landsat band
22	Gv	Green vegetation fraction	median	mosaic months	gv_median	Spectral Mixture Modeling
23	Gvs	GV / (100 - shade)	median	mosaic months	gvs_median	Spectral Mixture Modeling
24	Gvs dry	GV / (100 - shade)	median	year -first quartile	gvs_median_dry	Spectral Mixture Modeling
25	Gvs wet	GV / (100 - shade)	median	year – fourth quartile	gvs_median_wet	Spectral Mixture Modeling
26	Hallcover	Hall cover vegetation index	median	mosaic months	hallcover_median	Spectral index
27	Ndfi	Normalized Difference Fraction Index	median	mosaic months	ndfi_median	Spectral Mixture Modeling
28	Ndfi dry	Normalized Difference Fraction Index	median	year -first quartile	ndfi_median_dry	Spectral Mixture Modeling
29	Ndfi wet	Normalized Difference Fraction Index	median	year – fourth quartile	ndfi_median_wet	Spectral Mixture Modeling
30	Ndvi	Normalized Difference Vegetation Index	median	mosaic months	ndvi_median	Spectral index
31	Ndvi dry	Normalized Difference Vegetation Index	median	year -first quartile	ndvi_median_dry	Spectral index
32	Ndvi wet	Normalized Difference Vegetation Index	median	year – fourth quartile	ndvi_median_wet	Spectral index
33	Ndwi	Normalized Difference Water Index	median	mosaic months	ndwi_median	Water Index

D	Variable	Description	Statistics	Temporal range	Script acronym	Group
34	Ndwi dry	Normalized Difference Water Index	median	year -first quartile	ndwi_median_dry	Water Index
35	Ndwi wet	Normalized Difference Water Index	median	year – fourth quartile	ndwi_median_wet	Water Index
36	Near Infrared (NIR)	Landsat band	median	mosaic months	nir_median	Landsat band
37	Near Infrared (NIR) dry	Landsat band	median	year -first quartile	nir_median_dry	Landsat band
38	Near Infrared (NIR) wet	Landsat band	median	year – fourth quartile	nir_median_wet	Landsat band
39	Npv	Non-photosynthetic vegetation fraction	median	mosaic months	npv_median	Spectral Mixture Modeling
40	Pri	Photochemical reflectance index	median	mosaic months	pri_median	Spectral index
41	Pri dry	Photochemical reflectance index	median	year -first quartile	pri_median_dry	Spectral index
42	Pri wet	Photochemical reflectance index	median	year – fourth quartile	pri_median_wet	Spectral index
43	Red	Landsat band	median	mosaic months	red_median	Landsat band
44	Red dry	Landsat band	median	year -first quartile	red_median_dry	Landsat band
45	Red wet	Landsat band	median	year – fourth quartile	red_median_wet	Landsat band
46	Savi	Soil-adjusted Vegetation Index	median	mosaic months	savi_median	Spectral index
47	Savi dry	Soil-adjusted Vegetation Index	median	year -first quartile	savi_median_dry	Spectral index
48	Savi wet	Soil-adjusted Vegetation Index	median	year – fourth quartile	savi_median_wet	Spectral index
49	Sefi	Savanna Ecosystem Fraction Index	median	mosaic months	sefi_median	Fraction index
50	Sefi dry	Savanna Ecosystem Fraction Index	median	year -first quartile	sefi_median_dry	Fraction index

D	Variable	Description	Statistics	Temporal range	Script acronym	Group
51	Shade	Shade fraction	median	mosaic months	shade_median	Spectral Mixture Modeling
52	Soil	Soil fraction	median	mosaic months	soil_median	Spectral Mixture Modeling
53	Shortwave Infrared (SWIR) 1	Landsat band	median	mosaic months	swir1_median	Landsat band
54	Shortwave Infrared (SWIR) 1 dry	Landsat band	median	year -first quartile	swir1_median_dry	Landsat band
55	Shortwave Infrared (SWIR) 1 wet	Landsat band	median	year – fourth quartile	swir1_median_wet	Landsat band
56	Shortwave Infrared (SWIR) 2	Landsat band	median	mosaic months	swir2_median	Landsat band
57	Shortwave Infrared (SWIR) 2 dry	Landsat band	median	year -first quartile	swir2_median_dry	Landsat band
58	Shortwave Infrared (SWIR) 2 wet	Landsat band	median	year – fourth quartile	swir2_median_wet	Landsat band
59	Wefi	Woodland ecosystem fraction index	median	mosaic months	wefi_median	Fraction index
60	Wefi wet	Woodland ecosystem fraction index	median	year – fourth quartile	wefi_median_wet	Fraction index
61	Blue min	Landsat band	minimum	mosaic months	blue_min	Landsat band
62	Green min	Landsat band	minimum	mosaic months	green_min	Landsat band
63	Near Infrared (NIR) min	Landsat band	minimum	mosaic months	nir_min	Landsat band

D	Variable	Description	Statistics	Temporal range	Script acronym	Group
64	Red min	Landsat band	minimum	mosaic months	red_min	Landsat band
65	Shortwave Infrared (SWIR) 1	Landsat band	minimum	mosaic months	swir1_min	Landsat band
66	Shortwave Infrared (SWIR) 2	Landsat band	minimum	mosaic months	swir2_min	Landsat band
67	Blue	Landsat band	standard deviation	mosaic months	blue_stdDev	Landsat band
68	Cai	Cellulose Absorption Index	median	mosaic months	cai_stdDev	Spectral index
69	Cloud	Cloud fraction	standard deviation	mosaic months	cloud_stdDev	Spectral Mixture Modeling
70	Evi 2	Enhanced Vegetation Index 2	standard deviation	mosaic months	evi2_stdDev	Spectral index
71	Gcvi	(nir/green − 1)	standard deviation	mosaic months	gcvi_stdDev	Spectral index
72	Green	Landsat band	standard deviation	mosaic months	green_stdDev	Landsat band
73	Gv	Green vegetation fraction	standard deviation	mosaic months	gv_stdDev	Spectral Mixture Modeling
74	Gvs	GV / (100 - shade)	standard deviation	mosaic months	gvs_stdDev	Spectral Mixture Modeling
75	Hallcover	Hall cover vegetation index)	standard deviation	mosaic months	hallcover_stdDev	Spectral index
76	Ndfi	Normalized Difference Fraction Index	standard deviation	mosaic months	ndfi_stdDev	Spectral Mixture Modeling

D	Variable	Description	Statistics	Temporal range	Script acronym	Group
77	Ndvi	Normalized Difference Vegetation Index	standard deviation	mosaic months	ndvi_stdDev	Spectral index
78	Ndwi	Normalized Difference Water Index	standard deviation	mosaic months	ndwi_stdDev	Water Index
79	Near Infrared (NIR)	Landsat band	standard deviation	mosaic months	nir_stdDev	Landsat band
80	Red	Landsat band	standard deviation	mosaic months	red_stdDev	Landsat band
81	Savi	Soil-adjusted Vegetation Index	standard deviation	mosaic months	savi_stdDev	Spectral index
82	Sefi	Savanna Ecosystem Fraction Index	standard deviation	mosaic months	sefi_stdDev	Fraction index
83	Shade	Shade fraction	standard deviation	mosaic months	shade_stdDev	Spectral Mixture Modeling
84	Soil	soil fraction	standard deviation	mosaic months	soil_stdDev	Spectral Mixture Modeling
85	Shortwave Infrared (SWIR) 1	Landsat band	standard deviation	mosaic months	swir1_stdDev	Landsat band
86	Shortwave Infrared (SWIR) 2	Landsat band	standard deviation	mosaic months	swir2_stdDev	Landsat band
87	Wefi	Woodland ecosystem fraction index	standard deviation	mosaic months	wefi_stdDev	Fraction index
88	Slope	Terrain slope	identity	Permanent	slope	Geomorphometric
89	Green Texture	Texture from Landsat band	mean	mosaic months	green_median_texture	

D	Variable	Description	Statistics	Temporal range	Script acronym	Group
90	Latitude	Geographical coordinate	-	Permanent	Latitude	Geographic
91	Longitude	Geographical coordinate	-	Permanent	Longitude	Geographic
92	Ndvi_3years	Normalized Difference Vegetation Index	amplitude	Last 3 years mosaic months	ndvi_amp_3y	Spectral index

4.4 Classification algorithm, training samples and parameters

Classification was performed subregion by subregion, year by year, using the Random Forest algorithm (Breiman, 2001) available in Google Earth Engine, running 100 iterations (random forest trees).

Training samples for each subregion were defined following a strategy of using random pixels for which the land use and land cover remained the same (stable samples) along the maps of Collection 4 over different subperiods: 1985-1986, 1990-1994, 1995-1999, 2000-2004, 2005-2009, 2010-2014, 2015-2019 and 2020-2024, named as “stable samples”.

The identification of stable areas consists in extracting random pixels or “stable samples” based on a criterion of minimum temporal frequency aiming to ensure confidence to use them as training areas. Each pixel should be classified with the same LULC class throughout each sampling subperiod (1985-1989, 1990-1994, ….. 2020-2024). A layer of pixels with a stable classification for each subperiod was then generated. From the resulting layer of stable samples, a subset of 2,000 samples for each subregion was randomly generated for each class for each subperiod. It is important to clarify that not all of these samples were necessarily used in the classification process for each year.

In addition, a classical procedure to detect outliers was implemented. For each year, and within each training class, we searched for outliers in all variables. An outlier was defined as any value of a specific variable lower or higher than 1.5 times the interquartile range (the first quartile value subtracted from the third quartile value) considering all values of this variable within a specific class of a particular year. Samples containing values considered outliers for some variables were not discarded a priori, but fixed by replacing those values with the 5th percentile or the 95th percentile, whenever they were lower or higher than the thresholds considered, respectively. Finally, we disregarded only those samples containing simultaneously more than 20 variables of the feature space with values considered as outliers.

4.4.1 Sample size balance

We generated a fixed number of samples for each class, subregion and subperiod for classification. However we used in the classification process only a random subset based on the class area proportion within each subregion, considering each

year to be classified. To do this, we previously adjusted linear simple functions to estimate the area of each class for each year from 1985 to 2025 based on the annual class area observed along the Collection 3 dataset. These functions were used to estimate, for each year, the proportion of each class to train the classifier. Then, these annual proportions for each class were set to extract a subset of the available samples for the correspondent classification in each year. Whenever the classification resulted in overestimation or underestimation of the class after comparing with supplemental information (e.g.: Collection 3 maps, independent crop type maps, etc.) this proportion was adjusted changing the bias (intercept of linear regression model) accordingly. Notwithstanding the above, a minimum number of 50 to 100 samples per class was set for each region and year, to ensure the correct detection of the less frequent categories.

4.4.2 Complementary samples

The need for adding complementary samples was evaluated by visual inspection of the output of a preliminary classification, with both Landsat and high-resolution images available in Google Earth Engine and time series of vegetation indices, and also by comparing with the Collection 3 classification. Complementary sample collection was also done manually using points in Google Earth Engine Code Editor. All the false-color images of the 41 years (1985-2025) Landsat mosaics and the vegetation index time series were checked at the selected point. Based on the knowledge of each subregion, the samples for different classes were collected. Complementary samples previously generated for Collection 3 were also added in some regions to improve the classification when necessary.

4.4.5 Final classification

The final classification was performed for all subregions and years combining stable and complementary samples. For some years, the classification output resulted in anomalous results for some classes. Then, it was necessary to improve the classification through a new sample size balance and a specific set of complementary samples.

4.4.6 Post-classification

The results of the final classification were improved through a sequence of filters, to correct missing data, “salt-and-pepper” classification errors and, specially, cases of misclassification. Temporal filters were done with the aim to generate a more stable

classification pattern over time, avoiding unexpected class variation during short times.

4.4.6.1. Gap fill filter

A filter to fill no-data pixels (“gaps”) was applied. Because theoretically the 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.

4.4.6.2. Spatial filter

The spatial filter avoids unwanted modifications to the edges of the pixel groups; a spatial filter was built based on the "connectedPixelCount" function. Native to the Google Earth Engine platform, this function locates connected components (neighbors) that share the same pixel value. Thus, only pixels that did not share connections to a predefined number of identical neighbors were considered isolated. In this filter, at least six connected pixels were needed to reach the minimum connection value. Consequently, the minimum mapping unit is directly affected by the spatial filter applied, and it was defined as 6 pixels (~0,5 ha).

4.4.6.3. Temporal filters

The temporal filters use the information from the year before and after to identify and correct a pixel misclassification, considered as cases of invalid transitions. In a first step, the filter looks for specific cover classes (3, 11, 12, 33) that are not this class in 1985 and were kept unchanged in 1986 and 1987 and then corrects the 1985’s value to avoid any regeneration in the first year. In a second step, the filter looks at a pixel value in 2025 that for example is not 11 (wetland) but is equal to 11 in 2023 and 2024. The value in 2025 is then converted to 11 to avoid any regeneration in the last year. The third process looks in a 3-year moving window to correct any value that changed in the middle year and returns to the same class next year.

A temporal filter with a slightly different approach was applied to solve problems in forestry classification. To correct the problems related to the years with forestation cutting, interrupting a continuous series of years classified as forestry we used a

special six-year spatial filter. The rule of application checks whether two years before and two years after the class was forestation, if this is true it shifts the classification of the two middle years to silviculture.

4.4.6.4. Frequency filter

To correct classification problems associated with some classes in specific regions, frequency filters were applied to use the temporal information available for each pixel to correct cases of false positives. The general logic of the frequency filter is to search for each pixel a specific combination of classes throughout the 41 years producing a subset of pixels considered eligible for correction. Then the filter detects and overwrites only those years where cases of false positives are present using a fixed class value, that usually is the mode of classifications detected along the temporal range. This type of filter was used with parsimony to solve very well delimited cases.

4.4.6.5. Specific filters

Additional specific filters were generated to remove unexpected classification changes that remained after applying previous standard filters. In general, these filters that we applied work with frequency and incidence. Frequency is the number of years a class occurs in a pixel. The incidence is the number of times that a pixel classification changes along the entire series of years. The application of these filters was limited to fix problems of false transitions between specific classes.

We also used a filter that eliminates problems related to the shadows of the mountains. These filters use characteristics of the relief, in addition to the frequency to be applied. It corrects false positives of water and wetland in shaded slopes in regions with wavy relief. The filter selects all pixels classified as water at least in one year but in less than 38 years (<95%), or as wetland at least in one year but in less than 36 years (<90%), whenever occurring in areas of cliffs and slopes, established by a combination of slope data (SRTM derived) with HAND (Height Above the Nearest Drainage) database, to define places where it is not expected the presence of water or wetland. In such cases, both classes were replaced by the class corresponding to the pixel mode.

A filter to smooth abrupt transitions between the first and the second year (1985 - 1986) and the last and penultimate years (2023-2025) was applied It has been

observed in previous collections, that the last year of the series registered an unexpected increase in the area of anthropic classes and a decrease of natural classes, most likely corresponding to an artifact resulting from the set of applied filters. To alleviate the problem, a filter was developed to smooth this abrupt transition, avoiding all transitions from natural areas to anthropic areas, and vice versa, in patches equal to or smaller than 2 hectares. In these cases, the corresponding pixels from the last year receive the same classification as the penultimate year as well as pixels from the first year receive the same classification as the second year.

Exceptionally, the spatial effect of some filters was limited to a set of polygons, in such a way as not to modify the entire zone classification. Similarly, in some cases, filters were applied only for specific years. Examples of these filters include: a grassland filter that unifies wet and dry years, taking into account the coverage of that place and not the rainfall of a particular year. Or a rice filter that corrects sites classified as wet grasslands, only for certain years, as long as it has been previously classified as agriculture.

4.4.7 Discrimination Between Forest Plantations

In Collection 3 of MapBiomas Uruguay, a relevant methodological improvement was introduced: the disaggregation of Class 9, originally representing *forest plantations* as a single category, into three distinct classes. From that version onwards, pine plantations, eucalyptus plantations, and other forest plantations are mapped separately. This enhancement allows for a more detailed and accurate characterization of the extent and composition of forest plantations across the national territory.

4.4.7.1 Imagery

To differentiate Class 9 (forest plantations) into the category's eucalyptus, pinus, and other forest plantations throughout the temporal series (1985–2025), satellite imagery from Landsat 5, 7, 8, and 9 was used. To ensure consistency across sensors, Landsat collections were harmonized following the approach proposed by Roy et al. (2016).

For each year, spectral bands (B2, B3, B4, B5, B6, B7) and several vegetation and water indices were calculated, specifically the Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), Modified Soil-Adjusted Vegetation Index (MSAVI), Normalized Difference Water Index (NDWI), and Normalized Burn Ratio (NBR).

To enhance class separability, temporal statistical metrics—median, maximum, minimum, and standard deviation—were calculated for each spectral band and index to summarize intra-annual spectral variability. These metrics were then aggregated into a multi-band annual composite stack. In addition, topographic variables, specifically slope and the digital elevation model (DEM) derived from the Shuttle Radar Topography Mission (SRTM), were incorporated to account for terrain-related spectral variability.

4.4.7.2 Classification

Forest plantation classes were mapped using a supervised pixel-based classification based on the Random Forest algorithm (Breiman, 2001). The classification was constrained to areas identified as forest by the annual MapBiomas forest cover layers for each year of the study period, within which the three plantation categories were discriminated.

A single classification model was developed and trained using satellite data from 2020 and subsequently applied to the entire temporal series. The year 2020 was chosen because it was the most recent year with species-disaggregated cartographic reference data available at the time of the study (DGF-MGAP, 2021).

Training samples were generated using a stratified random sampling approach to ensure adequate representation across the spectral variability of all target classes. In total, 205 samples of eucalyptus plantations, 177 samples of pine plantations, and 110 samples of other forest plantation classes were collected (Figure 4).

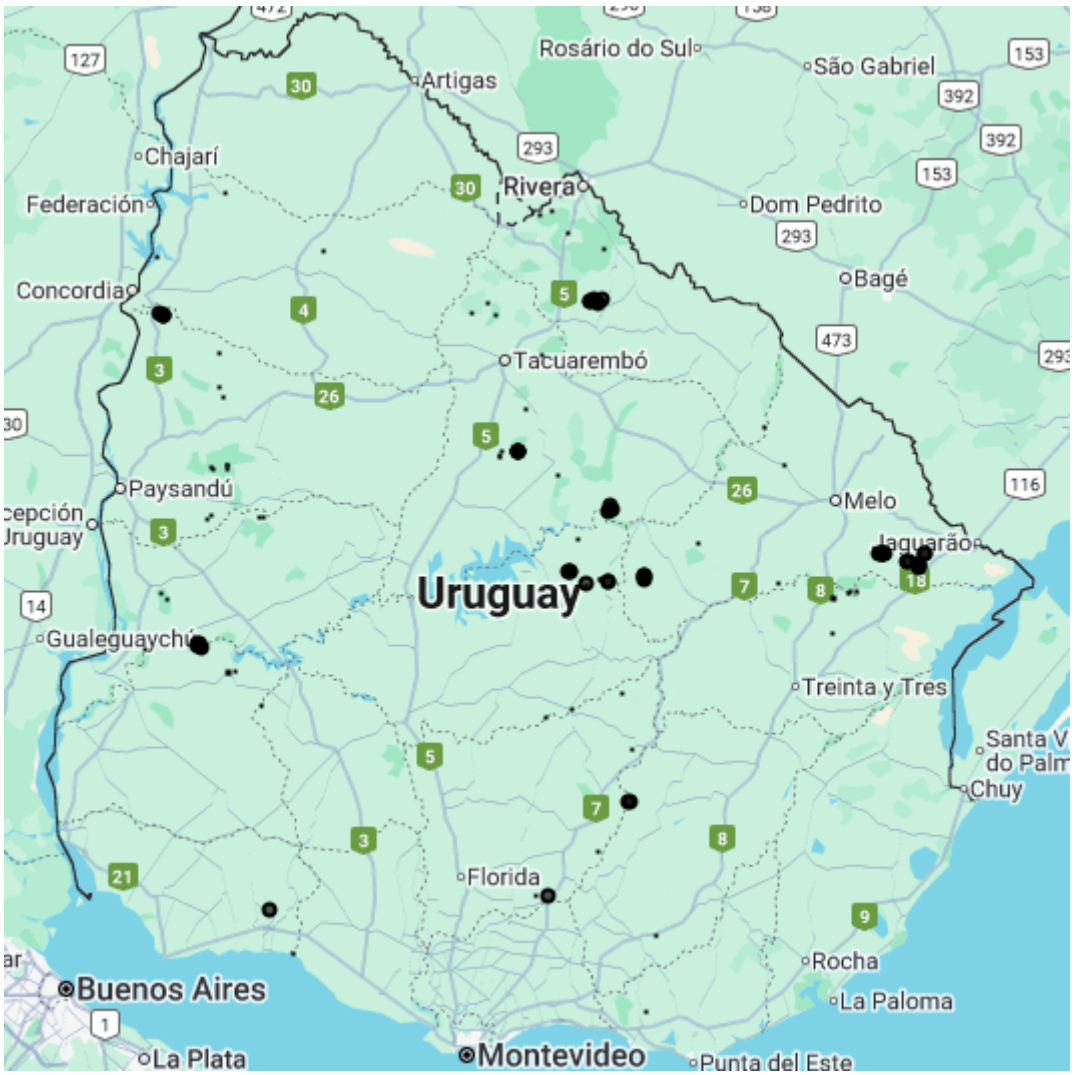


Figure 4: distribution of eucalyptus, pines and other forest plantations training samples.

4.4.7.3 Post-processing

Given the pixel-based classification and the long temporal span of the dataset, several post-classification spatial and temporal filters were applied to improve consistency and reduce noise. The correction process included gap filling, spatial filtering, and temporal filtering steps.

Gap filterling: A gap-fill filter was applied to ensure temporal continuity at the pixel level. In many cases, the annual MapBiomas masks identifying woody classes (e.g., native forest and forest plantations) were not continuous in time, and some pixels intermittently appeared as non-forest (e.g., grassland). Pixels classified at least once as *forest plantation* or *native forest* during the time series were assigned the non-forest class for all corresponding years to maintain temporal coherence.

Spatial filtering: To reduce edge noise and small misclassified patches, a mode filter (3×3 moving window) was applied, reassigning each pixel to the majority class of its neighborhood. Additionally, a connected pixel filter based on the connectedPixelCount function (native to Google Earth Engine) was implemented to remove isolated pixels and very small patches. Only areas with at least 111 connected pixels (≈10 ha) were retained, defining the minimum mapping unit. This spatial filter was applied exclusively to pixels classified as *forest plantation* at least once, or as *native forest* for fewer than 10 years in the series.

Temporal filtering:

Temporal filters were applied to correct implausible class transitions.

1. The first filter corrected early-year inconsistencies (1985) by checking class continuity in 1986 and 1987.
2. The second filter corrected the final year (2025) using information from 2023 and 2024, while avoiding removal of harvest-related non-vegetated states.
3. The third and fourth filters corrected short-term reversions between forest plantation classes within 3- to 6-year moving windows, replacing intermediate inconsistencies with stable class values before and after the transition.
4. The fifth filter addressed confusion between *forest plantation* subclasses, which commonly occurs in early growth stages. In this step, the class present at the end of the plantation cycle was propagated backward to correct the initial years, using a moving window of up to 13 years.

Together, these post-classification procedures minimized temporal noise, removed spurious transitions, and improved the overall thematic consistency of the forest plantation maps (Figure 5), without discarding legitimate transitions such as harvests or land-use changes.



Figure 5: Illustrative image of the separation of Class 9 into eucalyptus, pinus and other forest plantations.

4.4.8 Discrimination between Urban Areas and Others Non Vegetated Areas

Class 22 (Non vegetated area) is subdivided into Urban areas (24) and Other non-vegetated areas (25) (Table 1). Urban areas are defined as areas with a high density of roads and buildings and a predominance of non-vegetated surfaces, including infrastructure and open spaces. The discrimination follows the methodology of MapBiomas Argentina - Urban areas, Collection 2 (Picone et al., 2026), which draws on the MapBiomas Brasil Urban Areas Collection 10 (MapBiomas Brasil, 2025), adapted here to the characteristics of Uruguay: the national territory is processed in the seven geomorphological regions used for the LULC classification (Panario et al., 2014) instead of ecoregions, and the search areas and training samples are built from Uruguayan data sources. The procedure was developed in Google Earth Engine and is divided into input preparation (search areas, mosaics and samples), a probability classification with Random Forest, score harmonization, threshold determination, and spatial and temporal filtering. Table 3 lists the data sources and the step in which each one is used.

Table 3. Sources and data used in the discrimination between urban areas and other non-vegetated areas.

Source	Used information	Methodological step
National statistics institute (INE)]	Census urban localities and population 2023	Search area
10 m annual global LULC]	Class 7: Built area	Search area / Spatial filter
Global Human Settlement Layer, built-up surface (1985, 2005, 2020)]	Urbanized area	Samples
MapBiomas Uruguay Collection 4	Stable classes: grassland, agriculture, forest and water	Samples

Source	Used information	Methodological step
Landsat 5, 7, 8 and 9 collections	Annual mosaics (1985–2025)	Classification

4.4.8.1 Input preparation

Search area: The classification is restricted to a search area where urban areas can occur, which limits confusion with other non-vegetated surfaces in rural environments. The searches area was built from the intersection of the census urban areas, the Built area class of a 10 m global LULC product was extended with buffers of variable width according to the size of each locality, so as to include transitional and expansion areas. Buffers are applied according to the population of each locality, they range from 300 m for agglomerated rural population of fewer than 2,000 inhabitants to 2500 m for urban areas with more than one million inhabitants (e.g., Montevideo).

Mosaics: The mosaics are generated in Google Earth Engine using the Landsat Collection from satellite 5, 7, 8 and 9. Each annual mosaic is created on the fly and contains a total of 39 bands, which are a combination of Indexes, Fractions and Bands. The complete list is presented in Table 4

Table 4: List of Bands, Indexes and Fraction that are created to generate the annual Mosaic.

N°	BAND / INDEX / FRACTION	DESCRIPTION	REDUCTION METHOD	SCRIPT ACRONIC
1	BLUE	Landsat band	Median	BLUE_median
2	GREEN	Landsat band	Median	GREEN_median
3	RED	Landsat band	Median	RED_median
4	NIR	Landsat band	Median	NIR_median
5	SWIR1	Landsat band	Median	SWIR1_median
6	SWIR2	Landsat band	Median	SWIR2_median
7	NDVI	Normalized Difference Vegetation Index	Median	NDVI_median
8	EVI1	Enhanced Vegetation Index 1	Median	EVI_median
9	EVI1	Enhanced Vegetation Index percentiles	10p, 90p	EVI_p10, EVI_p90
10	EVI1	Enhanced Vegetation Index percentiles difference	Difference	EVI_dif9010

11	*EVI2*	*Enhanced Vegetation Index 2*	*Median*	*EVI2_median*
12	*EVI2*	*Enhanced Vegetation Index percentiles*	*10p, 90p*	*EVI2_p10, EVI2_p90*
13	*EVI2*	*Enhanced Vegetation Index percentiles difference*	*Difference*	*EVI2_dif9010*
14	*MNDWI*	*Modified Normalized Difference Water Index*	*Median*	*MNDWI_median*
15	*NDWI*	*Normalized Difference Water Index*	*Median*	*NDWI_median*
16	*NDBI*	*Normalized Difference Built-Up Index*	*Median*	*NDBI_median*
17	*NBR*	*Normalized Burn Ratio*	*Median*	*NBR_median*
18	*NDRI*	*Normalized Difference Road Index*	*Median*	*NDRI_median*
19	*BAI*	*Bare Soil Area Index*	*Median*	*BAI_median*
20	*UI*	*Urban Index*	*Median*	*UI_median*
21	*NDUI*	*Normalized Difference Urban Index*	*Median*	*NDUI_median*
22	*BSI*	*Bare-Soil Index*	*Median*	*BSI_median*
23	*BU*	*Built-up Index*	*Median*	*BU_median*
24	*GV*	*Green Vegetation Fraction*	*Median*	*GV_median*
25	*NPV*	*Non-Photosynthetic Vegetation Fraction*	*Median*	*NPV_median*
26	*SOIL*	*Soil Fraction*	*Median*	*SOIL_median*
27	*CLOUD*	*Cloud Fraction*	*Median*	*CLOUD_median*
28	*SHADE*	*Shade Fraction*	*Median*	*SHADE_median*
29	*GVS*	*Green Vegetation + Soil Fraction*	*Median*	*GVS_median*
30	*NDFI*	*Normalized Difference Fraction*	*Median*	*NDFI_median*
31	*SUBS*	*Index Substrate Fraction*	*Median*	*SUBS_median*
32	*VEG*	*Vegetation Fraction*	*Median*	*VEG_median*
33	*DARK*	*Dark Fraction*	*Median*	*DARK_median*
34	*EBBI*	*Enhanced Built-Up and Bareness Index*	*Median*	*EBBI_median*
35	*EBBI*	*Enhanced Built-Up and Bareness Index percentiles*	*25p, 90p*	*EBBI_p25, EBBI_p90*
36	*EBBI*	*Enhanced Built-Up and Bareness Index percentiles difference*	*Difference*	*EBBI_dif7525*

37	*EBBIsNeg*	*Positive part of the Enhanced Built-Up and Bareness Index*	*Median*	*EBBIsNeg_median*
38	*EBBIsNeg*	*Positive part of the Enhanced Built-Up and Bareness Index percentiles*	*25p, 75p*	*EBBIsNeg_p25, EBBIsNeg_p75*
39	*EBBIsNeg*	*Positive part of the Enhanced Built-Up and Bareness Index percentiles difference*	*Median*	*EBBIsNeg_dif7525*

Samples: To obtain samples of urbanized areas, the Built-up Surface (GHS-S) product was used (European Commission, 2024). This product, developed by the Joint Research Centre (JRC) of the European Commission, represents the physical extension of the built environment (built surfaces). This product offers us comparable multitemporal data for different times (1975, 1980, 1985, 1990, 1995, 2000, 2005, 2010, 2015 and 2020), with a spatial resolution of 100 meters. The product quantifies the proportion of built-up area in each pixel, allowing the analysis of urban sprawl, densification and urbanization patterns on a global and regional scale.

For this first version of MapBiomas Uruguay - Urban areas, data for the years 1985, 2005, 2020 (Figure 6) were selected, using only those pixels that had at least 30 % of their constructed area (>= 3000 m2). On this total number of pixels, samples were randomly extracted, using a stratified methodology to get representation of the different region of MapBiomas Uruguay and categories of urban areas.

On the other hand, for the samples of the uses and covers that surround the cities, stable samples (those that have not changed throughout all the years of classification) from collection 4 of MapBiomas Uruguay were used, these represent: water, herbaceous, cultivated areas and forests. From these samples only those located between the perimeter of the search area and the corresponding buffer area according to urban category were selected (Figure 6).

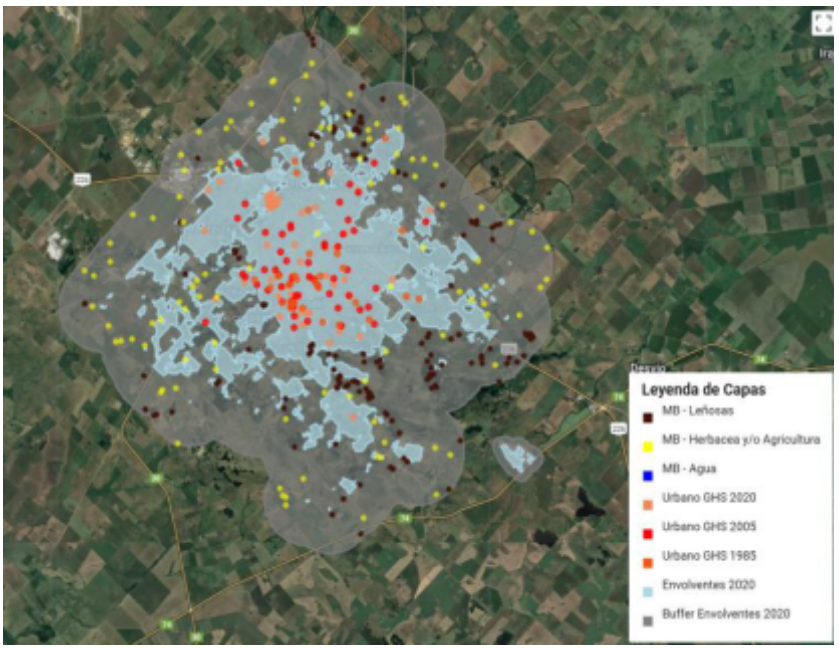


Figure 6. Example of urban and no-urban samples

4.4.8.2 Classification

Random Forest: The classification was executed independently for each region using the Random Forest algorithm (Breiman, 2001) in Google Earth Engine, with 100 decision trees and a minimum leaf population of 5. Using the annual mosaics, the search areas and the samples, the model calculates for each pixel and each year between 1985 and 2025 a score (0–100) that represents the probability of being an urbanized area.

Score harmonization: Because each year is classified independently along a 40-year series, the probability scores vary between years due to changes in soil moisture, phenology of the adjacent vegetation, atmospheric artifacts and illumination. To reduce this spectral noise, a continuous temporal window of 5 years (t-2, t-1, t, t+1, t+2) is applied to the scores as a moving average, which stabilizes the historical series by attenuating abnormal peaks and drops.

Determination of thresholds: The continuous score is transformed into a binary classification (urban / non-urban) with an adaptive threshold defined for each region and classification period (1985–2004, 2005–2019 and 2020–2025). The threshold is determined with two criteria: the ROC curve (Receiver Operating Characteristic), which identifies the point that maximizes the detection of urban areas while minimizing confusion with other land covers, and the 15th percentile (P15) of the

probability values of the urban training samples, following MapBiomas standards. Pixels of Class 22 above the threshold are assigned to Urban areas (24) and the remaining pixels to Other non-vegetated areas (25).

4.4.8.3 Post-processing

To obtain a coherent cartographic sequence from the annual binary layers, a succession of post-classification filters was applied.

Spatial filter 1: A 3 × 3 pixel moving window evaluates geographical contiguity. It eliminates salt-and-pepper noise by discarding isolated pixels and fills logical gaps within consolidated urban tissue.

Temporal filters 1 to 3: The trajectory of each pixel is evaluated with a 5-year moving window (t-2, t-1, t, t+1, t+2). The urban class is validated for a given year only if it persists in at least 3 of the 5 years. This smooths interannual inconsistencies and interpolates temporal data absences, for example those caused by clouds.

Spatial filter 2: A supermask based on external high-resolution data (Class 7 of a 10 m Sentinel-derived LULC) removes persistent false positives in peri-urban or rural contexts that survived the previous filters because of spectral similarity, such as quarries, dunes, or other exposed soils.

Temporal filter 4 (irreversibility rule): The transformation of natural or agricultural land into paved or built surfaces is assumed to be a cumulative and irreversible process. Once a pixel is validated and consolidated as urban, it cannot revert to a non-urban condition in subsequent years, which ensures progressive growth trajectories.

5 VALIDATION STRATEGIES

Validation was performed for the classifications of the years 1986, 1991, 1996, 2001, 2006, 2012, 2018 and 2022 following the good practices recommendations proposed by Olofsson et al. (2014) for area and error estimation. The accuracy assessment was designed for the entire Río de la Plata grassland biome and included a total of 2,330 samples.2,330 randomly selected samples. The number of samples for each class was proportional to the area of each class obtained from Collection 1 for the year 2010. Independent samples were raffled and class classified by visual interpretation of Landsat images, very high resolution images from Google Earth and

time series of vegetation indices. Two interpreters evaluated each of the sample points generated from the stratified random design. In those sample points where discordance in class classification was detected among interpreters, a third interpreter defined the final class assignment. When a final class could not be defined by the three interpreters (e.g. three different class assignments), a final class was agreed by a team of interpreters. More details of the validation methodology are described in Baeza et al. (2022) and Gallego et al. (2026)

A complete classification accuracy analysis comparing the different MapBiomas Pampa collections can be found here. An analysis of a subset of the data, conducted exclusively on points corresponding to the territory of Uruguay (N=490), carried out for Mapbiomas Uruguay Collection 2, can be found in Gallego et al. 2026. The results showed an overall accuracy of 81% for 1986, 88 % for 1991, 89 % for 1996, 85 % for 2001, 85 % for 2006, 91 % for 2012, 86 % for 2018 and 89% for 2022 (Gallego et al 2026)

6 REFERENCES

Baeza, S., Vélez-Martin, E., De Abelleyra, D., Banchero, S., Gallego, F., Schirmbeck, J…….. , & Hasenack, H. (2022). Two decades of land cover mapping in the Río de la Plata grassland region: the MapBiomas Pampa initiative. *Remote Sensing Applications: Society and Environment*, 28, 100834. https://doi.org/10.1016/j.rsase.2022.100834

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

Dirección General Forestal – Ministerio de Ganadería, Agricultura y Pesca (DGF-MGAP). (2021). *Cartografía Forestal Nacional 2021 – República Oriental del Uruguay (Versión 1.2, agosto 2021).* Montevideo: MGAP.

Gallego, F., Barbieri, A., Ramos, S., Bruzzone, L., Vallejos, M., Rama, G., Baldassini, P. & Baeza, S. (2026). The MapBiomas Uruguay initiative to monitor annual land use and land cover changes from 1985 to 2023. *Ecología Austral*, 031-046.

Liu, F. T., Ting, K. M., & Zhou, H. (2012). Isolation-based anomaly detection. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 6(1), 3. https://doi.org/10.1145/2133360.2133363

MapBiomas Brasil. (2025). Urban Areas – Algorithm Theoretical Basis Document (ATBD), Collection 10, Version 1. MapBiomas Brasil.

Olofsson, P., Foody, G. M., Herold, M., Stehman, S. V., Woodcock, C. E., & Wulder, M. A. (2014). Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment*, 148, 42–57. https://doi.org/10.1016/j.rse.2014.02.015

Panario, D., Gutiérrez, O., Sánchez Bettucci, L., Peel, E., Oyhantçabal, P., & Rabassa, J. (2014). Ancient landscapes of Uruguay. In J. Rabassa & C. Ollier (Eds.), *Gondwana landscapes in southern South America* (pp. 161–199). Cham: Springer.

Picone, N., Linares, S., Álvarez, M., Schteingart, L., Dieguez Gaviola, G., Eljall Qüesta, A. V., Rizzo, P., Gomez, L., Cardozo, O., & Flores, P. (2026). MapBiomas Argentina – Urban areas "Handbook": Algorithm Theoretical Basis Document (ATBD), Collection 2, Version 1. MapBiomas Argentina.

Pontius Jr, R. G., & Millones, M. (2011). Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing*, 32(15), 4407–4429.

Roy, D. P., Kovalskyy, V., Zhang, H. K., Vermote, E. F., Yan, L., Kumar, S. S., & Egorov, A. (2016). Characterization of Landsat-7 to Landsat-8 reflective wavelength and normalized difference vegetation index continuity. *Remote Sensing of Environment*, 185, 57–70. https://doi.org/10.1016/j.rse.2015.12.024