



## **Pampa - Appendix**

### **Collection 10**

### **Version 2**

#### **General coordinator**

Heinrich Hasenack

#### **Team**

Eduardo Vélez-Martin

Eliseu José Weber

Juliano Schirmbeck

Grazieli Rodigheri

## 1 OVERVIEW OF CLASSIFICATION METHOD

Collection 10 contains annual land use and land cover maps (LULC) for the period of 1985-2024 and represents the process of continuous improvements throughout the successive collections (**Table 1**).

It was built up following a sequence of steps, similar to those used in the previous collections, including:

- 1) Building annual Landsat mosaics – using pixel values (median, minimum, amplitude and standard deviation) of the scenes within a temporal optimum period,
- 2) Definition of a feature space – set of remote sensing metrics, including original bands and indexes used as potential class predictors,
- 3) Training samples – reference samples for each class, using the temporal stable samples approach,
- 4) Classification of Landsat mosaics with training samples – using the Random Forest algorithm,
- 5) Post-classification treatment – a series of filters to remove noise, classification errors and unlikely transitions among years,
- 6) Final map assembling - integration of the classes mapped by the biome team with additional cross-cutting classes,
- 7) Validation - accuracy assessment using a set of independent multi annual validation samples and comparison with another available maps for the biome.

The classification stages for Collection 10 are similar to those of the preceding collections. The classification process was repeated for every year aiming to solve of the main commission and omission errors identified in the validation of Collection 9. In addition, some of the post-classification filters were enhanced and modified. All codes used for producing Collection 10 are available at GitHub (<https://github.com/mapbiomas-brazil/pampa>).

**Table 1:** Main characteristics of land use and land cover MapBiomias collections for the Pampa biome.

Collection	Temporal range	Number of classes	Improvements	Global Accuracy (level 3)
1.0	2008-2015	5	Empirical decision tree	*
2.0 - 2.2	2000-2016	5	Classification using a decision tree Temporal filter	*
2.3	2000-2016	5	Classification using random forest New temporal filter rules	63%
3.0	1985-2017	6	Stable training samples from coll.2.3	*
3.1	1985-2017	6	Reclassification of some charts Frequency filter	78.4%
4.0	1985-2018	7	Stable samples from coll. 3.1 Homogenous regions as classification units, new frequency and incidence filters	79.0%

Collection	Temporal range	Number of classes	Improvements	Global Accuracy (level 3)
4.1	1985-2018	7	Reclassification of some regions	80.5%
5.0	1985-2019	7	Stable samples from coll.4.1 Reference maps to validate training samples new biome limit, feature space increased, filter of outliers	88.4%
6.0	1985-2020	7	Stable samples from coll. 5, new Landsat mosaics, annually balanced training samples, adjusted frequency filters	89.5%
7.1	1985-2021	9	Stable samples by time interval from coll. 6 new Landsat Collection 2 mosaics, annually balanced training samples, adjusted post classification filters	89.3%
8	1985-2022	9	Classification in two stages: keeping the pixels with classification agreement in Collections 5, 6 and 7.1 and reclassifying only those discordant pixels. Stable samples by time interval from coll. 6, annually balanced training samples, adjusted post classification filters	89.7%
9	1985-2023	9	Stable samples by time intervals from coll. 8, new Landsat Collection 2 mosaics, annually balanced training samples, adjusted post classification filters	89.4%
10	1985-2024	9	Stable samples within four temporal intervals from coll. 8 maps, Removal of stable samples for 2023-2024 inconsistent with the classification of 2023 in collection 9. New Landsat Collection 2 mosaics, annually balanced training samples, classification within nine spatial regions, adjusted post classification filters, Rocky outcrops classified separately within occurrence zones	89.9%

\* Non evaluated

## 2 ASSESSMENT OF COLLECTION 9 TO GUIDE COLLECTION 10

The critical analysis of the Collection 9 was based on several criteria, including the temporal variation of the area of each class, the patterns of transitions among classes, the commission and omission errors for each class and the visual inspection of the Landsat mosaics in those areas of disagreement between Collections 8 and 9, especially for the dominant classes. Based on these analyses, the following main guidelines were established:

- Review and adjust mainly the classifications of the initial years of the series due to their lower accuracies, considering those classes with higher commission and omission errors.
- Improve the contours of forest patches, especially in regions 1 and 6.
- Reduce excessive false positives of Forest and Other non-vegetated areas in the eastern portion of region 6. Test the subdivision of the region 6 due to its large longitudinal extension to facilitate improvements in regional classification through sample balancing.
- Reduce the omissions of Rocky outcrop and confusion with the class Other non-vegetated areas.
- Review some post-classification filters rules that are incorrectly increasing the area of class 21.
- Add a new spatial filter to remove small and irregular patches of class 21 (mosaic of uses) within larger patches of class 12 (grassland). The effect of microrelief on grassland vegetation confuses the classifier, which incorrectly considers areas with greater soil exposure to be anthropized.

### 3 REGION OF INTEREST AND MOSAICS AND CLASSIFICATION UNITS

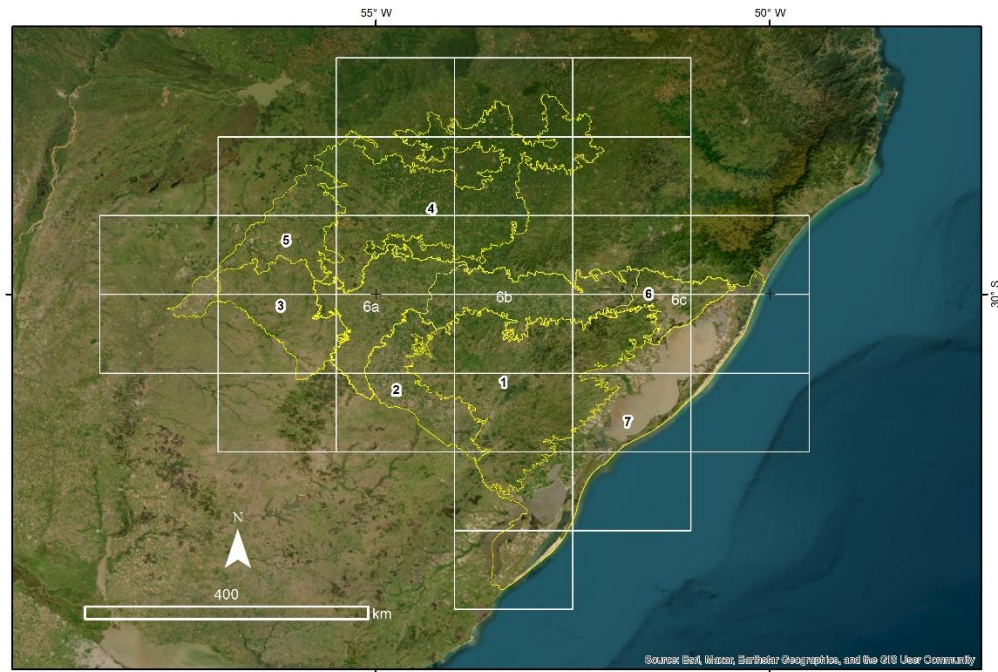
The mapped area totals 193,935 km<sup>2</sup> and corresponds to the geographical official limits of the Pampa biome (IBGE 2019). Since the first collection, the spatial units for data processing followed the subdivision of the World International Chart to the Millionth at the 1:250,000 scale, hereafter called 'charts'. Each chart defines the geographical boundaries to build up the temporal and spatial Landsat mosaics, to collect training samples, and to run the digital classification.

Up to Collection 4.1, a number of 23 charts was needed to cover the Pampa biome and the final map was generated by merging the charts. In 2019, the limits of the biome were updated (IBGE, 2019) and since Collection 5, it was necessary to merge four additional charts, totaling 27 charts to completely cover the biome extension. Landsat mosaics for these additional charts were generated for all the years between 1985 and 2019. In Collection 8, the same charts approach was adopted to build up and manage the Landsat mosaics for all the years between 1985 and 2022.

However, since Collection 5, charts remained as the spatial units to build the mosaics but were no longer used as the classification spatial units. We adopted a new scheme using seven homogeneous regions within the biome as the new geographical units for the classification. For each region, all chart Landsat mosaics were previously merged before the classification process. These regions are an adaptation of the former nine ecological systems proposed by Hasenack *et al.* (2010) for the Brazilian Pampa biome, using vegetation, relief and soils data (**Figure 1**). Thus, the 27 mosaics that cover the biome were merged and clipped to produce seven ones, according to the limits of each geographical region. In collection 10, the region six was divided into three distinct regions according to different environmental

contexts related to longitude. Because of this, the biome was independently classified into nine spatial regions (1, 2, 3, 4, 5, 6a, 6b, 6c, 7).

The use of regions instead of charts as the operational units for classification improved the results, increasing mapping accuracy and avoiding cases of edge mismatching between charts, observed in previous collections.



**Figure 1:** Geographical units of classification in the Pampa biome: the 27 charts of the World International Chart to the Millionth (1:250,000) used in Collection 9 (white lines) and the seven regions (yellow lines).

## 4 LANDSAT IMAGE MOSAICS

### 4.1 Image selection

According to the year, quality and availability, images from three Landsat satellites were used:

- 1985 to 2000 and 2003 to 2011: Landsat 5 (TM),
- 2001, 2002 and 2012: Landsat 7 (ETM+),
- 2013 to 2024: Landsat 8 (OLI).

Until Collection 5, all Landsat mosaics were built using top of atmosphere (TOA), while in Collection 6.0 surface reflectance (SR) was adopted as an attempt to improve the classification accuracy. Consequently, all annual mosaics were also recomputed. For Collection 7, all annual mosaics were also recomputed because of deprecation of Landsat Collection 1. Then, all annual mosaics used in collections 7, 8, and 9 were based on surface reflectance data from the new Landsat Collection 2 (Tier 1).

Each annual mosaic of images consists of the median, minimum, amplitude or standard deviation value for each band of each pixel, calculated over all the images available in a previously defined “optimal period” within the year. Additional details on the definition of the optimal period are presented in item **3.2**. The annual Landsat mosaics were stored in the chart format (see item **2**) and used in the classification process.

The selection of Landsat scenes to build each annual mosaic was based in the optimal period considered within the Pampa biome spring context (September-November), using a threshold of 90% of cloud cover tolerance (i.e., any available scene with up to 90% of cloud cover was accepted). In some cases, the temporal range of three months resulted in mosaics with data gaps, due to excessive cloud cover and/or lack of data. In such cases, the standard period was extended to encompass a larger number of scenes to minimize no data pixels. Whenever possible, additional months at the beginning of the period were preferred, mainly in the winter season.

In most cases, at least one month had to be added to provide enough images for the mosaic (**Table 2**), with an overall mean of 5 months. In some specific cases, it was necessary to significantly extend the temporal period, while in others it could be shortened. The temporal extension was more frequent for the years 1987, 1990, 1995, 1999, 2007 and 2010, with a mean of six months. The temporal reduction was noticeable for the year 2012, when the acceptable period for eight charts had to be reduced to only two months due to the low quality of the available Landsat 7 scenes.

In previous collections, the mosaics algorithm in Google Earth Engine included a set of procedures to produce cloud free images, combining a pixel cloud score, a temporal dark outlier mask (TDOM) and the quality band (BQA). However, the combined effect of these filters caused excessive loss of information due to confusion between clouds and the typical reflectance of sandy regions. In Collection 7, 8, and 9, we kept only BQA to filter cloudy pixels when building the new mosaics.

Another innovation included along the process of producing new mosaics since Collection 7 was to manually remove those noisy Landsat scenes within each annual chart mosaic, whenever an anomaly was visually detected when inspecting the first generated mosaic versions.

Considering the 27 charts of the Pampa biome and the 40 years of Collection 10, 1.080 chart mosaics were produced, all with satisfactory quality. Eventually, small portions of some annual mosaics remained with no data pixels, but always with a proportion lower than 5% of the chart.

#### **4.2 Definition of the optimal temporal range**

The selection of the optimal period of the year for image classification was defined to minimize the confusion between natural and cultivated vegetation, due to phenological changes, while trying to maximize the coverage by useful Landsat images after cloud removal/masking.

Unlike most of the other Brazilian biomes, the climate of the Pampa biome does not have a defined dry season; the annual variation of temperature is the main factor determining the physiological behavior of vegetation throughout the year. At the mean latitude of the biome (30° S), solar radiation in winter is nearly half that of summer, which leads to marked temperature differences between the seasons.

In the Pampa biome, the grassy ecosystems are the predominant native vegetation and tend to present a characteristic seasonal pattern. **Figure 2** presents a schematic diagram of the seasonal behavior of grasslands and the most significant summer crops in the Pampa biome, markedly paddy rice and soybean. During autumn, the photosynthetic activity of herbaceous vegetation begins to decline, reaching its lowest point in winter, when a significant portion of the leaf biomass reaches a senescent stage. From late winter and early spring on, annual species germinate, and perennial species begin to regrow, shooting new leaves and increasing progressively the photosynthetically active biomass, which will reach its peak in the summer.

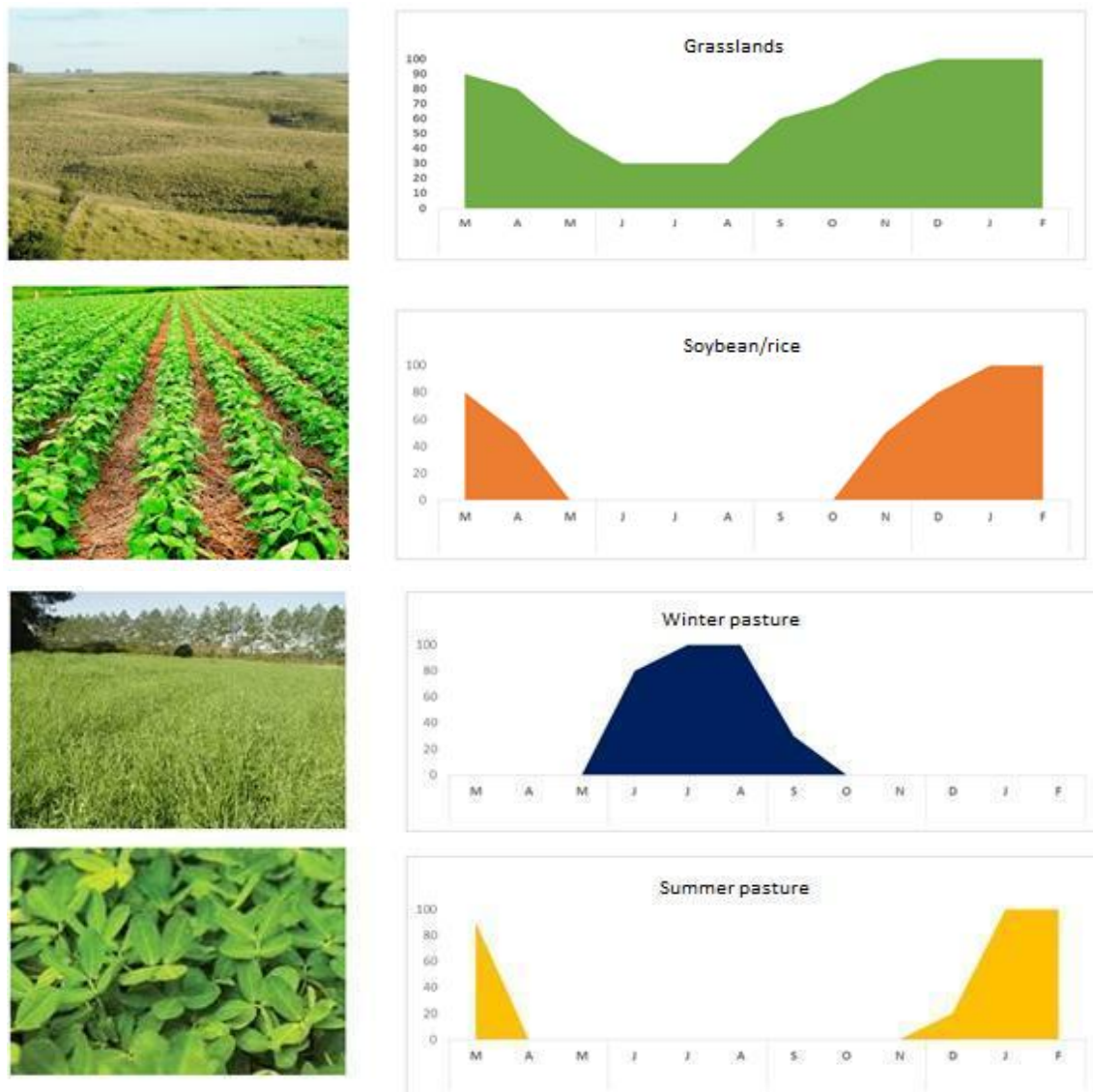
On the other hand, forests, although classified mostly as Seasonal Deciduous Forest and, to a lesser extent as Semi-Deciduous, express minor deciduousness. Only a small fraction of tree species in the forest communities lose leaves during winter, thus Pampa forests are expected to show less variation in spectral response over the year than other types of vegetation cover.

**Table 2:** Temporal range (number of months) for selection of Landsat scenes used in each chart mosaic of the Pampa biome for each year in the period 1985-2024.

	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2020/ 2024		
	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0			
	8	8	8	8	8	9	9	9	9	9	9	9	9	9	9	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1			
	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9		
SG-21-Z-D	3	4	8	3	6	5	3	3	6	4	8	4	4	4	8	5	3	5	3	3	3	4	8	3	3	4	3	2	3	3	3	3	3	3	3		
SG-22-Y-C	3	5	8	3	4	4	3	3	3	3	4	5	5	7	8	3	4	5	3	3	3	3	4	3	3	3	8	2	3	3	3	3	4	3	3		
SG-22-Y-D	3	5	8	3	4	4	3	3	3	3	4	5	5	6	3	3	4	5	3	3	3	3	4	3	3	3	3	2	3	3	3	3	4	5	3		
SH-21-V-D	8	4	3	3	3	6	3	3	3	3	7	3	3	3	4	3	3	3	3	3	6	3	8	3	3	3	3	3	3	3	6	3	3	3	3		
SH-21-X-A	3	3	3	3	3	6	3	3	8	4	3	3	4	4	3	3	4	4	3	3	4	3	8	3	3	10	3	3	3	3	6	3	3	3	3		
SH-21-X-B	3	4	8	3	6	5	3	3	6	4	8	4	4	4	5	5	3	5	3	3	3	4	8	3	3	4	3	2	3	3	3	3	3	3	3		
SH-21-X-C	4	4	3	3	3	7	3	3	8	3	8	4	5	6	4	3	4	3	3	3	4	3	9	3	3	3	3	3	3	3	6	3	3	3	3		
SH-21-X-D	5	5	3	3	3	4	3	3	8	5	8	4	5	3	3	5	4	5	3	3	3	3	7	8	3	8	9	2	5	5	3	3	3	5	3		
SH-21-Y-B	4	4	3	3	3	6	3	3	3	3	7	3	3	4	7	3	3	3	3	3	5	3	8	3	3	3	3	3	3	3	5	3	4	3	3		
SH-21-Z-A	4	3	7	3	3	7	3	3	7	3	5	3	3	4	5	3	7	3	3	4	5	3	9	3	3	3	3	3	3	3	6	3	4	3	3		
SH-21-Z-B	5	3	3	4	3	4	3	3	8	5	8	3	3	3	3	6	4	8	3	3	3	3	7	3	3	3	8	3	3	6	3	3	3	5	3		
SH-21-Z-C	7	3	3	3	3	6	7	4	5	3	4	3	3	4	4	3	3	3	3	3	3	8	3	3	3	3	3	3	3	3	3	3	3	5	3		
SH-21-Z-D	8	4	3	4	3	4	3	3	3	7	4	6	3	3	3	6	5	8	3	3	3	3	5	3	3	3	3	3	5	6	3	6	3	3	3		
SH-22-V-A	3	5	8	3	6	4	3	3	3	3	4	5	5	4	8	5	4	5	3	3	3	3	4	3	3	3	3	2	3	3	3	5	4	5	3		
SH-22-V-B	3	5	8	3	4	4	3	3	3	3	4	5	5	6	3	3	4	5	3	3	3	3	4	3	3	3	3	2	3	3	3	3	4	5	3		
SH-22-V-C	5	4	8	4	6	4	3	5	3	5	3	6	5	6	5	6	5	7	8	3	3	3	5	3	3	3	8	2	3	6	3	6	5	6	3		
SH-22-V-D	9	3	8	3	3	4	3	4	3	4	3	8	8	4	3	9	4	3	4	6	8	3	5	3	5	8	3	6	3	4	4	5	3	3	3		
SH-22-X-C	3	3	3	3	9	5	3	4	3	4	3	7	4	5	3	8	5	5	4	3	6	3	4	3	8	8	8	8	6	6	3	3	3	3	7	5	
SH-22-Y-A	5	3	7	5	3	6	3	8	3	8	3	7	4	3	8	5	4	7	5	3	3	3	8	3	3	9	8	7	3	3	3	9	3	7	3		
SH-22-Y-B	8	4	6	4	3	10	3	12	3	7	8	8	5	4	8	9	6	3	4	8	8	7	6	3	6	9	3	6	3	4	9	9	9	6	5		
SH-22-Y-C	7	3	6	4	3	8	9	3	3	8	3	8	5	3	8	3	6	3	4	3	3	4	8	7	3	9	3	6	3	5	4	9	4	4	7	3	
SH-22-Y-D	8	8	5	7	4	8	4	8	8	7	8	7	5	6	8	9	6	6	3	8	6	6	6	3	9	8	3	3	3	4	6	4	6	7	9		
SH-22-Z-A	9	3	3	7	6	7	9	9	3	5	8	8	4	8	7	7	6	7	8	6	8	6	8	4	6	9	9	3	4	8	9	5	5	3	5	5	
SH-22-Z-C	7	7	3	7	4	8	7	8	3	4	8	6	4	5	8	3	6	3	4	7	6	7	5	5	6	8	8	9	5	6	4	5	6	3	6	5	
SI-22-V-A	6	8	6	3	9	8	8	3	3	7	3	7	8	3	8	6	5	7	3	7	3	9	6	6	8	9	9	9	3	5	5	5	5	3	5	8	
SI-22-V-B	3	7	8	3	4	11	7	8	3	9	8	5	7	5	6	6	3	7	4	5	7	6	4	3	9	3	3	3	3	3	6	5	6	3	9	3	
SI-22-V-C	5	9	9	8	6	8	5	6	6	7	7	8	9	3	7	3	3	3	3	7	3	8	5	6	8	8	8	8	3	3	6	5	5	3	6	3	
Mean	5	5	6	4	4	6	4	5	5	5	6	5	5	4	6	5	4	5	4	4	4	4	6	4	4	4	6	5	4	3	4	5	4	4	4	4	

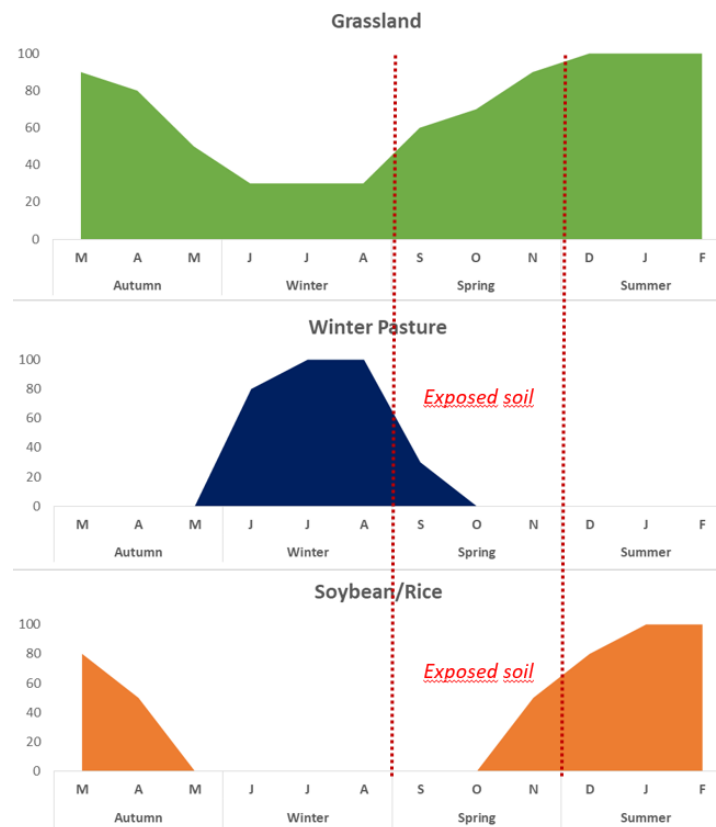


In paddy rice and soybean parcels, soil preparation takes place during late winter and early spring, which can result in exposed soil (conventional planting), dried vegetation (no-tillage) due to herbicide application, or flooded in the case of rice planting, according to the adopted agronomic management practices. It is also usual to sow winter pastures, for providing green soil cover and/or supplemental forage for the livestock. Consequently, during winter there are patches with photosynthetically active herbaceous vegetation in the landscape, contrasting with the senescent native grasslands. Summer planted pastures have less expression than winter pastures, and their peak of photosynthetic activity coincides with that of the grassland vegetation during January and February (**Figures 2 and 3**).



**Figure 2:** Scheme of typical phenological patterns of native grasslands, soybean/rice crops, winter and summer pasture in the Pampa biome. The y-axis corresponds to photosynthetic biomass production. The presented values are in percentage and are merely illustrative.

The optimal period of the year to distinguish between native vegetation, crops, pastures with exotic species and forestry through remote sensing in the Pampa biome was defined taking into account these phenological patterns. It is expected a higher contrast to be observed among these land cover types from September to November, when summer crop areas are under preparation for sowing, cultivated pastures are off their photosynthetic peak, and native grasslands are in the beginning of regrowth and development of new leaves (**Figure 3**).




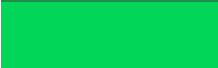








**Figure 3:** Schematic phenological pattern of grasslands, summer crops and winter pastures in the Pampa biome, indicating the temporal window with the best contrast for the purpose of satellite imagery classification.

## 5 CLASSIFICATION

### 5.1 Characterization of land cover and land use classes

The digital classification of the Landsat mosaics for the Pampa biome included nine land use and land cover (LULC) classes (or ten, considering the class Not Observed) from the complete MapBiomass Collection 10 legend (**Table 3**). After the integration with the cross-cutting themes (Forest plantation, Soybean, Rice, Other Temporary Crops, Beach, Dune and Sand Spot, Urban area and Mining), the final map encompassed a total of 16 classes, at level 4 of the hierarchical legend.

**Table 3:** Land cover and land use classes, before integration with cross-cutting themes, considered for digital classification of Landsat mosaics for the Pampa biome in the MapBiomass Collection 10.

Legend class of Collection 10	Numeric ID	Color
1.1.1. Forest Formation	3	
1.1.2. Wooded sandbank vegetation	49	
2.1. Wetland	11	
2.2. Grassland	12	
2.3 Herbaceous sandbank vegetation	50	
2.4. Rocky Outcrop	29	
3.3 Mosaic of Uses	21	
4.4. Other Non-Vegetated Areas	25	
5. River, Lake and Ocean	33	
6. Not Observed	27	

**Figure 4** shows the 2024 LULC map for Collection 10. The sixteen classes encompassed by the final map (**Figure 5**), after integration with cross-cutting themes, correspond to the following description:

1. **Forest formation** (3): woody vegetation with trees or shrubs species, with a predominance of a continuous canopy. It includes ombrophilous, deciduous and semi-deciduous forests, and part of the pioneer formations.

2. **Wooded sandbank vegetation** (49): arboreal vegetations over sandy soils along the coastal zone of the biome.

3. **Forest plantation** (9): tree species planted for commercial purposes (i.e. pine, eucalyptus, acacia).

4. **Wetland** (11): swampy areas, regionally called "banhados" or "marismas" (saltmarshes). Typically, hygrophilous vegetation, with emergent, submerged or floating aquatic plants. They occupy plains and depressions in the terrain with waterlogged soil and the shallow margins of ponds or water reservoirs.

5. **Grassland** (12): Vegetation with a predominance of an herbaceous grassy layer, together with herbaceous and subshrub dicotyledons. The botanical composition is influenced by edaphic and topographic gradients and by pastoral management (livestock). They occur in deep to shallow soils, including rocky (rupestrian grasslands) and sandy terrains (sandy or psammophilous grasslands). They occupy well-drained soils (mesic grasslands) or even soils with higher moisture content (humid grasslands - with a marked presence of sedges). In most cases, it corresponds to native vegetation, but patches of invasive exotic wild or forage species (pastures) may be present.

6. **Herbaceous sandbank vegetation** (50): herbaceous vegetations over sandy soils along the coastal zone of biome.

7. **Mosaic of Uses** (21): areas of agricultural use, where it was not possible to distinguish between pasture and agriculture. May include cropland, winter or summer pastures and horticulture. Includes areas left unsown for a period between agricultural crops (fallow).

8. **Beach, Dune and Sand Spot** (23): sandy ridges, bright white colored, where there is no predominance of vegetation of any kind.

9. **Urban area** (24): urbanized areas with a predominance of non-vegetated surfaces, including roads and buildings.

10. **Other Non-Vegetated Areas** (25): mixed class that includes natural and anthropic areas. Natural areas include sandy surfaces such as river beaches and sandy terrains. Anthropic areas include areas of exposed soil and non-permeable surfaces (infrastructure, urban sprawl or mining).

11. **Rocky Outcrop** (29): rocks naturally exposed, often with partial presence of rupicolous vegetation

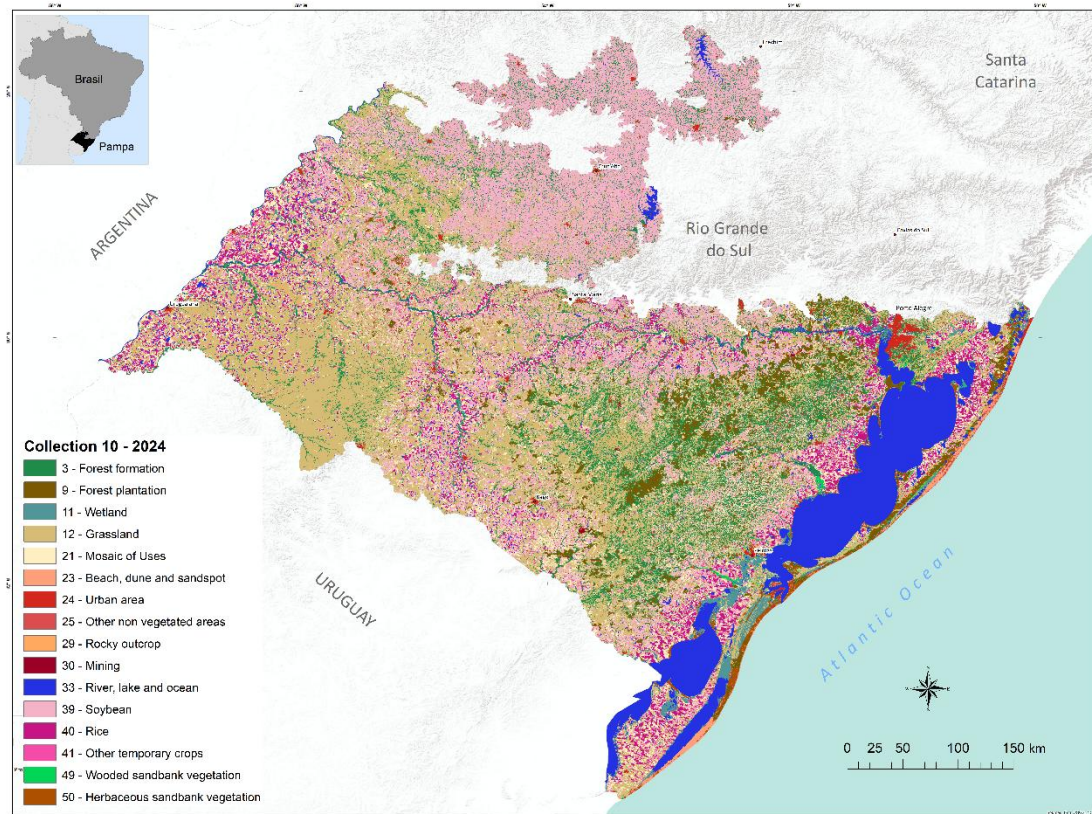
12. **Mining** (30): areas with large-scale mineral extraction, with a clear exposure of the subsoil.

13. **River, Lake and Ocean** (33): Areas with the presence of permanent surface water. Includes streams, rivers, ocean, lagoons, natural and artificial lakes.

14. **Soybean** (39): parcels cultivated with soybean.

15. **Rice** (40): parcels cultivated with paddy rice

16. **Other Temporary Crops** (41): parcels cultivated with other agricultural crops, mainly of annual cycle.



**Figure 4.** 2024 LULC map in the Pampa biome.

## 5.2 Feature space

The feature space used in the digital classification comprises a set of 93 variables (**Table 4**), which is the complete feature space available in the mosaics (90 layers) plus another three variables (see variables 90-92 in **Table 4**). These variables include statistical reducers from the original Landsat reflectance bands, as well as spectral indexes, spectral mixture modeling-derived variables (fractions), fractions indexes, terrain morphometry (slope), and a spatial texture measure. All variables stored in the mosaics were previously standardized and converted to 16 bytes format.

In many cases, the same variable was calculated considering different temporal ranges within each year, and using different metrics, before storing as different layers in the annual mosaics and according to the following possibilities:

1. Optimal period range: considering only the scenes within the months previously selected for each year as the best period for LULC classification (see **item 3.2**). Within this range the metrics included: median, minimum, amplitude (maximum-minimum) and standard deviation.

2. Dry season: this range was defined using the 25<sup>th</sup> lowest NDVI values from all scenes available within each year as a proxy to select those scenes considered to belong to a dry season. Then, within this subset of scenes the median metric was the only one calculated.





**Figure 5:** Classes mapped in the Pampa biome, Collection 10.

**Table 4:** Feature space (93 variables) used in the digital classification of Landsat image mosaics in the MapBiomass Collection 10 (1985-2024) for the Pampa biome.

ID	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
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

ID	Variable	Description	Statistics	Temporal range	Script acronym	Group
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
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
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	Landsat band	median	year -first quartile	swir2_median_dry	Landsat band



ID	Variable	Description	Statistics	Temporal range	Script acronym	Group
	Infrared (SWIR) 2 dry					
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
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

ID	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	
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

3. Wet season: this range was defined using the 25<sup>th</sup> highest NDVI values from all scenes available within each year as a proxy to select those scenes considered to belong to a wet season. Then, within this subset of scenes the median was the only one metric calculated.

As the Pampa biome rainfall is well distributed throughout the year, the terms “dry season” and “wet season” are not adequate, however they were kept to fit the general MapBiomass terminology. They in fact refer, respectively, to minimum and maximum NDVI periods within the year, which are in fact most influenced by temperature in the Pampa biome, expressing contrasting behaviors of vegetation cover in moments of low and high photosynthetic efficiency.

### **5.3 Classification algorithm, training samples and parameters**

Digital classification was performed region by region, 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 region were defined following a strategy of using pixels for which the LULC remained in the same class over the years in Collection 8, hereafter called “stable samples”.

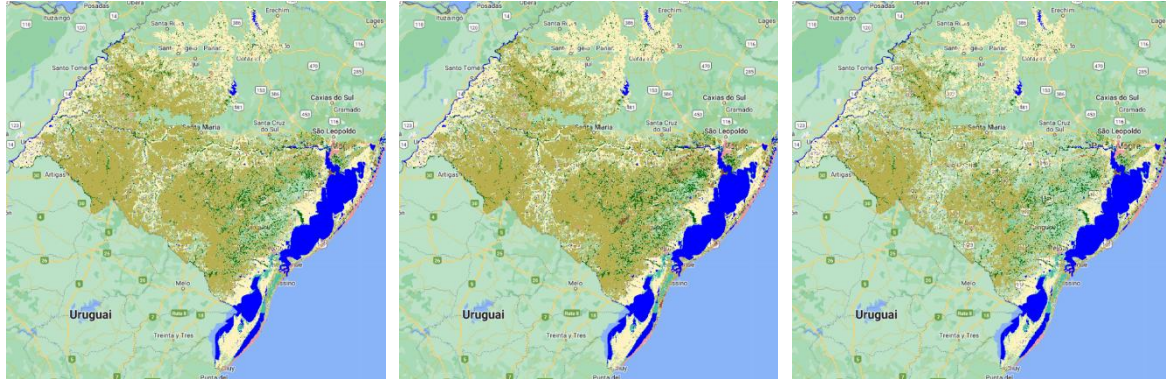
#### **5.3.1 Selection of stable samples**

We used only samples generated from places with invariant classification over time, named as stable samples. The extraction of these stable samples was based on the maps from the previous Collection 8, following several steps aiming to ensure their confidence for use to train the classification algorithm. We decided to remain using the stable samples based on the maps from collection 8, instead of collection 9, for two reasons: 1) because of the greater accuracy obtained in collection 8 and, 2) to take advantage of the annual balancing of samples from collection 9 as a starting point and make fine adjustments based on it aimed at resolving the main errors of commission and omission detected by class and for specific years.

The first step was to produce maps considering only those pixels with the same class in all annual maps. Until Collection 6, a unique layer of pixels with a stable classification throughout the years was generated, producing a “stable areas map” for all classes. However, since Collection 7.1 we adopted an alternative approach by generating three temporal stable area maps. In Collections 9 and 10 we used the following time intervals: 1 (1985-1998), 2 (1999-2012) and 3 (2013-2022) (**Figure 6**).

These stable area maps were used to randomly select a set of independent samples, used to classify only within those correspondent years for each time interval. An additional step was added before the extraction of training samples for the second and third temporal intervals. The stable areas maps from Collection 8 were masked for classes 12 (Grassland) and 21 (Mosaic of Uses). The mask was built using only the stable pixels of grasslands and agriculture within the Pampa reference maps for the years 2002 (Hasenack et al., 2015), 2009 (Weber et al., 2016) and 2015 (Hofmann et al., 2018). Then the mask was intersected with

the stable areas map. All the matches between similar classes of the two maps were considered as a validation of the stable status for those classes and the corresponding pixels kept in the stable areas map. All non-matches were used to remove pixels from the map of stable areas in an attempt to reduce uncertainty in the sample dataset. This procedure was applied only to the stable areas of the temporal intervals 2 and 3 and limited to these two classes, once they are prevalent in the biome.



**Figure 6:** Stable classes in the Pampa biome from Collection 8 used to extract a random set of stable training samples used in Collection 10. Colored pixels are the stable classes, and colorless pixels are the places that changed along each time interval: (1: 1985-1998, 2: 1999-2012, and 3: 2013-2022 from left to right).

After masking the stable areas map, we extracted a random dataset of 2,000 points for each one of the seven classes, within each one of the nine classification regions. These points were used to extract values for all variables contained in the 40 annual mosaics (1985-2024) and stored as a training sample dataset. It is important to clarify that not all of these samples were necessarily used in the classification process.

The stable samples from the third time interval (2013-2022) were also used to classify the years 2023 and 2024. However, as a precaution, all stable samples that changed class in the map for the year 2023 (collection 9) were removed before classifying these two years.

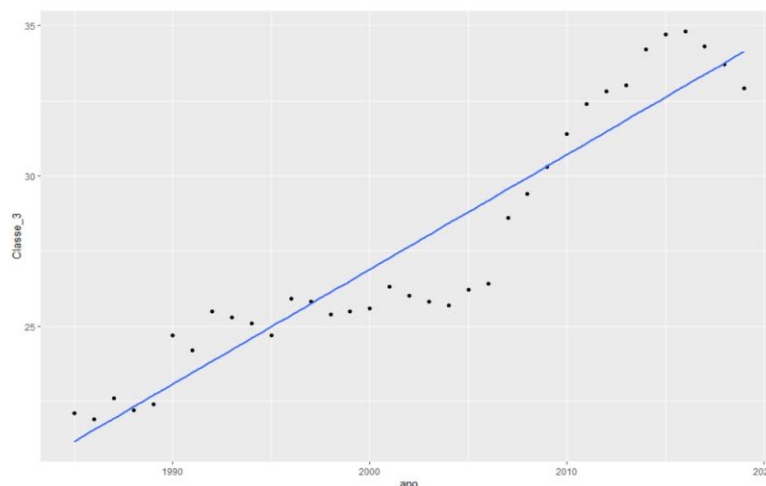
At last, we performed a data cleaning process at the training dataset. For each year, and within each training class, we searched for outliers in all variables. An outlier was defined as any value of a specific variable lower or higher than 1.5 times the interquartile range (the first quartile value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

### 5.3.2 Balancing training samples

The definition of the number of samples of each class to be used in the classification process was established through a weighted balancing. The general idea is that the weight of each class samples must be proportional to the area occupied by the class  $i$  in the year  $j$ , as a starting point for classification.

Each class weight was calculated as proportion and multiplied by the number of 2,000 points available in the training data set to establish the final number of samples to use in the classification. So, each weight was calculated for each class, for each one of the years, and for each one of the seven regions.

To calculate these annual weights to establish the proportion of samples to use for each class, we first converted the class area values observed in Collection 8 for each year to relative proportions of the region of interest. Then, we fitted a linear regression, for each class, considering the relative proportions ( $y$ ) within the region of interest along the 38 years ( $x$ ) and extracted the intercept ( $b_0$ ) and the slope of the regression line ( $b_1$ ). Therefore, for the year to be classified, each class weight (dependent variable) was calculated using the year as the independent variable (**Figure 7**). These weights correspond to proportions for each class that were multiplied by the total available samples to define the final number of samples to use. Additionally, we set a minimum sample size of 50-100 training samples, according to the region, to ensure sufficient representation for those classes with low area proportion within each region.



**Figure 7:** Example of the sample weighting for the class Forest formation (class 3) in region 1 of the Pampa biome, using relative proportions along the years computed from Collection 6. The coefficients ( $b_0$ ,  $b_1$ ) of the fitted linear regression were used to calculate the weight of this class in each year to classify the region along all years in the Collections 6, 7.1, 8, 9 and 10.

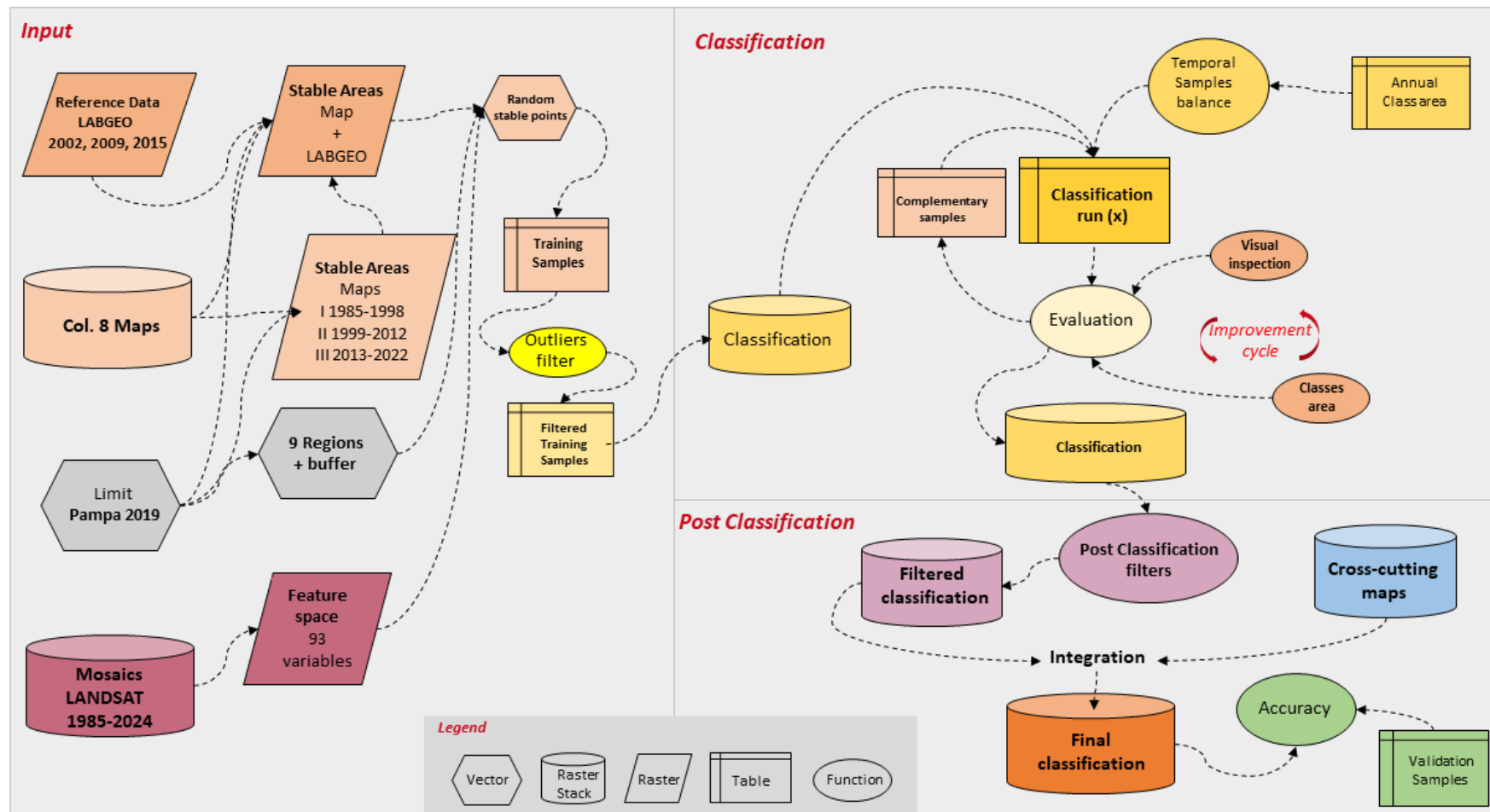
### 5.3.3 Supplementary samples

In addition to the stable samples used in the classification, a set of supplementary samples were also used to correct specific omission and commission errors. These samples were collected directly in the Google Earth Engine platform environment, as geometries (points) for some classes. Supplementary samples were added after the inspection of preliminary classification results. The classification for specific years were inspected through visual inspection and comparison with the Landsat mosaics to search for classification errors for each class. During this process, those class patches identified as commission or omission errors were used to set additional samples. Some specific pixels within misclassified patches were visually identified, assigned to the correct class, and stored as a sample point after checking its temporal stability. This verification was done after checking if the class was the same in the years 1985 and 2024 in the corresponding mosaics. Addition of these supplementary samples was conducted in batches of 5 to 10 samples, followed by new classifications to check if they were sufficient to correct the incorrect classifications. In cases where errors persisted, new batches of samples were added.

### 5.3.4 Final classification

**Figure 8** presents a general flowchart of the whole classification process. The classification was performed for all regions and years using the balanced annually defined training sample sizes as a starting point. Improved versions of the classification were subsequently produced by adjusting the sample sizes of each class, within an iterative process, through inspection of partial results (mapped area of each class, visual inspection, etc.) and identification of necessary sample size adjustments followed by a new classification. In some cases, the sample size adjustment was not sufficient, then we added a set of stable complementary samples to correct persistent errors of commission or omission.

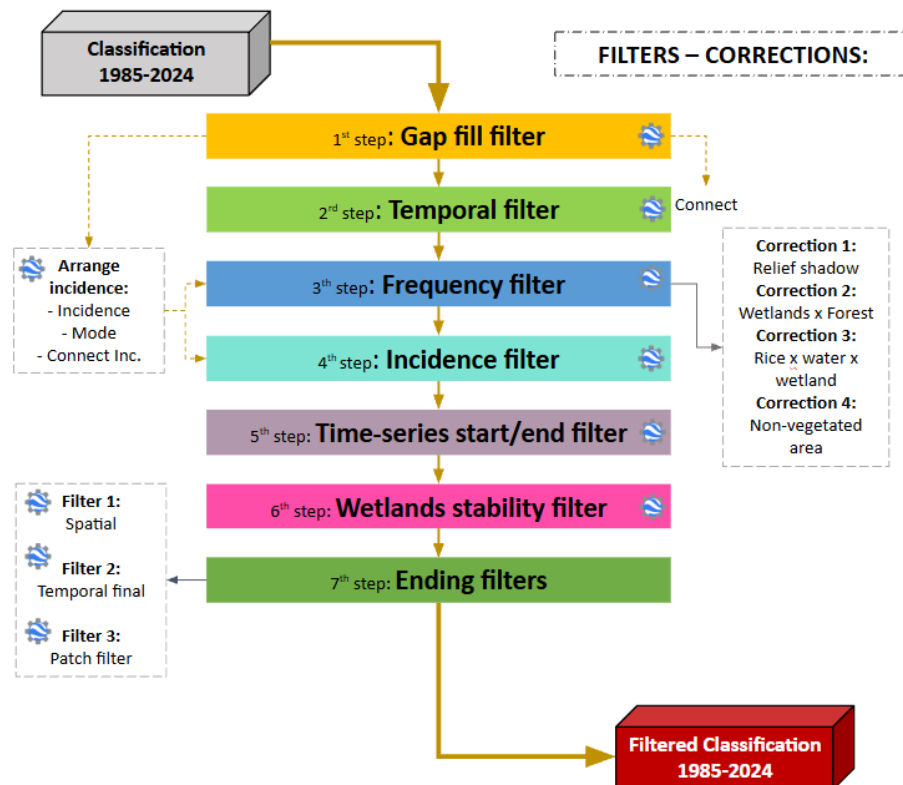
The final classification of Collection 10 produced 40 new maps, one for each year of the 1985-2024 series, each one with six classes: forest formation, wetland, grassland, mosaic of uses, other non-vegetated area, and river, lake, and ocean. The rocky outcrop class was classified separately, only within rocky outcrop zones, for the year 2022. This year was selected because their Landsat mosaics performed better in the classification of this class than those of 2023 and 2024. The annual maps containing the six classes were subjected to post-classification filters, and at the end of this process, the rocky outcrop class was overlaid on each of the classified years.



**Figure 8:** Classification steps used in Collection 10 in the Pampa biome.

## 6 POST-CLASSIFICATION

The classification results were post-processed through a sequence of 7 filters conceived to correct remaining classification issues (**Figure 11**), including: pixels with no data value due to image noise removal, salt-and-pepper noise on classification, unlikely/invalid transitions, false transitions in pixels at the border of regions with different classes, abrupt transitions from natural areas to anthropic areas and vice versa, unlikely transition between stable and variable classes, and special misclassifications cases.



**Figure 11:** Post-classification filters used in the classification of the Pampa biome, Collection 10.

### 6.1 Gap fill filter

This filter uses information of previous years to replace pixels classified as Non-observed in a given year through a forward procedure. Then, it is complemented by the information of later years, through a backward procedure, to replace pixels that remained as Non-observed.

### 6.2 Temporal filter

The temporal filter uses the information from previous and subsequent years to identify and correct pixel misclassifications for a given year, assuming a set of invalid transitions.



The rules differ for the first, the last and intermediate years of the collection. The process starts looking at the first three years, comparing the class in the year 1985 with the two following. Whenever a pixel in the first year differs from a native vegetation class (3, 11, 12, 29) and is assigned to one in the two next years (1986 and 1987), then it is replaced by the corresponding natural class. For the last three years, the year 2024 is compared with the two precedents and, whenever a pixel is classified as 21 (Mosaic of Uses) in both but is different in the last year, then it is replaced by class 21. Both procedures aim to avoid cases of false positives of regeneration.

The last step applies a 3-year moving window to correct the remaining intermediate years. Whenever the first and the third year of the window have the same class and the middle year is different, it is replaced by their class. This procedure has the purpose of fixing abrupt transitions that are unlikely to happen. The filter was applied, step by step, respecting the following sequence of classes: [29, 22, 21, 11, 3, 12, 33].

### 6.3 Frequency filter

Frequency filters were applied to use the information available for each pixel over the years to correct cases of false positives.

The general logic of the frequency filters is to search for each pixel a specific combination of classes throughout the 40 years producing a subset of pixels considered eligible for correction. Then the filter detects and overwrites only those years where cases considered as probable false positives 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 only well delimited cases. Four different variations of the frequency filters were employed in Collection 10, as described below:

- *Frequency filter 1 – Relief shadow x water and wetland*

The first frequency filter aims to correct false positives of water and wetland in shaded slopes in regions with wavy relief. It was applied only in regions 1, 2, 3, 4 and 5.

The filter selects all pixels classified as water at least in one year but in less than 34 years (<95%), or as wetland at least in one year but in less than 32 years (<90%), whenever occurring in areas of the the class 2, from a specific combination of slope data (SRTM derived) with HAND (*Height Above the Nearest Drainage*) database. This class corresponds to sloping areas relatively far from water bodies, 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.

- *Frequency filter 2 – Wetlands x other natural classes*

The second frequency filter was used to fix misclassifications of wetland in other classes. It was applied in all regions and contains three sets of rules:

The first rule selects all pixels classified as wetland in more than 25 years (>70%) and in natural classes (grassland, forest formation and water) the remnant years. The filter uses incidence (Pontius *et al.*, 2017) as ancillary information. The rule considers that whenever a pixel shifts between these classes with an incidence greater than 1 (2 or more shifts) this is

probably a result of misclassification, and all values of the pixel along the remnant years (30%) are assigned to the class wetland.

The second rule selects all pixels classified as wetland in less than 7 years (<20%) and with an incidence greater than 1. These cases are probably false positives of wetlands, being replaced by the modal class over the 40 years.

The third rule selects all pixels classified as wetland in 7 to 25 years and shifts among forest and water. Whenever the frequency of forest is lower than 11 years (<33%), all the years classified as forest are considered cases of false positives and are replaced by wetland. Pixels classified as water are not corrected.

- *Frequency filter 3 – Paddy rice x wetlands*

The third frequency filter was used to fix the confusion of paddy rice with temporary water or wetland, and was applied only in regions 5, 6, and 7.

The filter selects all pixels shifting among classes 11 (Wetland), 21 (Mosaic of Uses), and 33 (Water) and assigns all years to class 21, whenever the frequency of this class is greater than 33% over 36 years. This is because part of paddy rice parcels remains flooded after harvesting (fallow), sometimes with the development of aquatic plants, thus being false positives when classified as water or wetlands.

- *Frequency filter 4 – Non vegetated areas*

The first rule selects all pixels with high frequency (>80%) of class 25 (Other non-vegetated areas). All the remaining years classified differently from class 25 are then replaced by class 25. This majority criteria is an attempt to minimize the problem of misclassification.

The second rule selects all pixels with low frequency (<20%) of class 25 (Other non-vegetated areas). Then those years classified as class 25 are replaced by the modal class.

#### **6.4 Incidence filter**

The incident filter was applied to correct cases of false transitions observed in pixels placed in the border of patches belonging to different classes. The classification of pixels located at the boundary of two classes, are sensitive to be classified to one or another of these classes along the years because of unwanted effects. They may include small tilts in the acquisition of the original radiometric information by the satellite sensors, associated with issues of georeferencing precision, but also to the effect in data acquisition due to drier or wetter years. The consequence is an increased level of false transitions for these pixels over the years.

The filter selects all pixels presenting more than six incidents over the 40 years (Pontius *et al.*, 2017) and belonging to patches of less than six pixels and replaces them by the modal class. It is assumed that small patches with high incident rates are indicative of confusions that are common in edge pixels.

#### **6.5 Time-series start/end filter**

This filter smoothes the transitions between the first and second years and between

the penultimate and last year of the time series. In previous collections, 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 of 2023, as well as pixels from the first year receive the same classification of 1986.

## **6.6 Wetland stability filter**

Total area of wetlands in the Pampa biome is expected to have a temporal stability for the period 1985-2023, which the classifier fails to adequately capture. This is mainly due to a detection bias in Landsat 8, which outperforms both Landsat 5 and Landsat 7. This artifact generates larger areas of this class for the most recent years when compared to previous years, erroneously indicating the expansion of the class area in the territory. To deal with this problem we generated a mask based on the classifications of the most recent years (2013-2024) and equivalent to the Landsat 8 mosaics. The mask was composed of all pixels classified as wetlands for more than seven years in the 2013-2024 interval. Additionally, all pixels that had been classified at least once as forestry in any of these years were removed from the mask. This procedure was necessary due to the confusion between these classes, especially in small patches. At last, the mask was overlaid retroactively over the classifications from 1985 to 2012.

## **6.7 Ending filters**

The spatial filter uses a mask to change isolated pixels or very small patches (less than six pixels) of a class by replacing each one with the most frequent value in their corresponding eight neighbors. The filter uses the "connectedPixelCount" function in Google Earth Engine and produces a result where the minimum mapped area is a patch with at least six pixels of the same class (~0.54 ha).

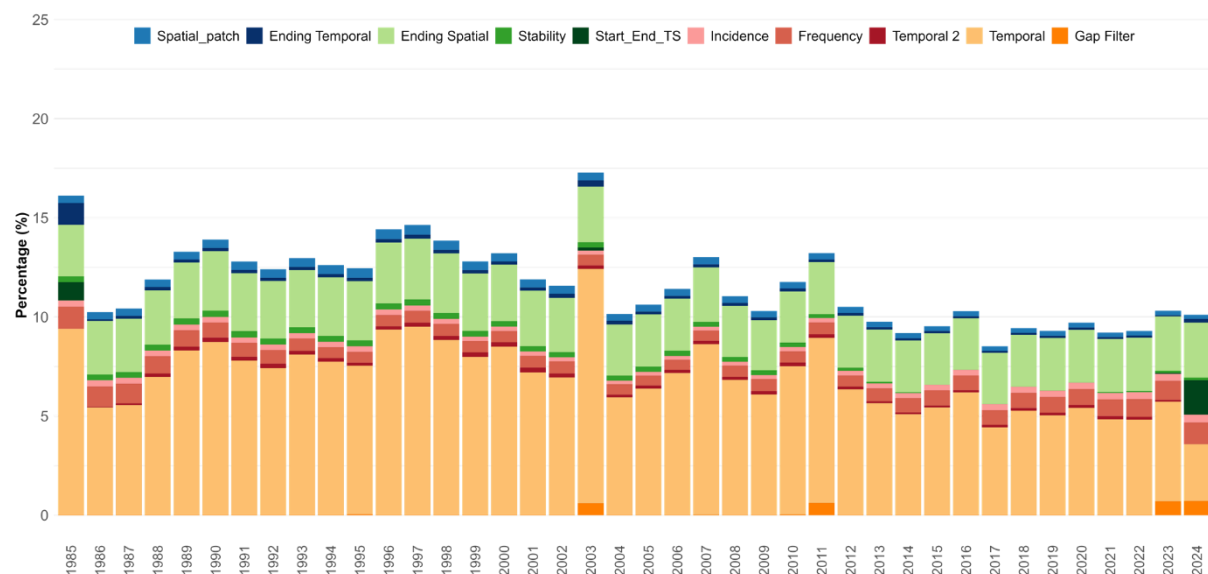
The same spatial and temporal 3-years filters, previously described, were applied again to remove incidental and unwanted effects in the filtered classification resulting from the combined application of filters.

At last, a patch spatial filter was applied to remove small patches of the class Mosaic of uses, incorrectly classified within larger patches, mainly of grasslands, but also of forest, wetlands, and water classes. This is a classification error predominantly found in grassland dominated landscapes. The microrelief variation, in interaction with the herbaceous vegetation layer, sometimes confuses the classifier, especially in places where the soil is shallower or the vegetation has less ground cover. In some cases, these locations are incorrectly classified as mosaic of uses, particularly in the form of small patches with very irregular perimeters, although they are areas where native vegetation has never been

removed. The filter was established by assigning a unique identifier to each patch of class Mosaic of uses (21), considering patches below 2ha. Then, for each of these small patches, we calculated the proportion of neighboring pixels classified as forest, wetland, grassland, or water. Only the subset of these patches with 100% of neighboring pixels belonging to one of these classes were selected and considered eligible for classification correction. Then, for each patch, all pixels of class 21 were corrected to the predominant surrounding class (forest, wetland, grassland, or water). To avoid undue corrections, an exclusion mask was applied on the biome. The mask was generated based on the Rural Environmental Registry (*CAR – Cadastro Ambiental Rural*, in portuguese), a database containing the boundaries of each rural property. All small properties (up to 80 hectares) were selected to integrate the mask. All small patches of class Mosaic of uses intersecting with this mask were considered as true positives and were not filtered.

## 6.8 Quantitative effect of filters

**Figure 12** presents the annual effect of each filter expressed as a percentage of the biome area. On average, the filters adjusted 11.6% of the annual classifications, ranging from 8.5% to 17.3%. The filters with the most significant effect were the temporal filter (average of 6.9%) and the spatial filter (average of 2.5%). All other filters had effects below 1%.



**Figure 12:** Annual quantitative effect of post classification filters in the Pampa biome, Collection 10.

## 7 CLASSIFICATION FINAL ADJUSTMENTS

After the post-classification process, three additional classes are added to the annual maps. The first corresponds to the rocky outcrop class, classified separately from the other classes (see item 5.3.3).

The two other classes correspond to wooded sandbank vegetation and herbaceous sandbank vegetation. Both were included after the final classification through a

reclassification process. The class Wooded sandbank vegetation was defined as reclassifying all the patches classified as Forest within the limits of the Brazilian marine and coastal system (IBGE, 2019). The class Herbaceous sandbank vegetation was defined as reclassifying all the patches classified as Grasslands within the limits of the Brazilian marine and coastal system (IBGE, 2019) and belonging to Regosols patches from a soil classification map (IBGE, 2018)

## 8 INTEGRATION WITH CROSS-CUTTING THEMES

The annual maps resulting from the post-classification filters for each one of the 39 years (1985-2023) were integrated with the cross-cutting themes. This assembling is performed using hierarchical prevalence rules (**Table 5**). The output is a final set of integrated LULC maps for the Pampa biome for 39 years.

Additionally, some exceptions to these prevalence rules were applied due to particularities observed in the Pampa biome, including:

1. Class 12 (grassland) intersects most of the cross-cutting class 15 (Pasture). The class pasture is classified without distinction between grasslands and planted pastures and is almost totally overlapped in the integration by class 12. Most of the remaining pixels classified as pasture correspond to small patches of annual grasses planted as a winter land cover in alternation with summer crops. Considering this particularity, the remnant pixels of class 15 are remapped to class 21 (Mosaic of Uses).

2. Class 11 (wetland) prevails over the agriculture classes 39 (soybean), class 40 (rice) and class 41 (other temporary crops). This rule was adopted to avoid false positives of these anthropic classes over areas that are recognizably wetlands.

3. Class 33 (water) prevails over classes 39, 40 and 41 to avoid false positives of these anthropic classes.

**Table 5:** Hierarchical prevalence rules for combining the output of the Pampa classification with the cross-cutting themes in Collection 10.

Class	Pixel value	Prevalence
4.4. Mining	30	1
4.1. Beach, Dune and Sand Spot	23	2
1.1.3. Mangrove*	5	3
5.2.1. Aquaculture*	31	4
2.3. Salt Flat*	32	5
4.2. Urban area	24	6
3.3 Forest plantation	9	7
3.2.1.2. Sugar Cane*	20	8
3.2.1.1. Soybean	39	9
3.2.1.3. Rice	40	10
3.2.1.4. Other Temporary Crops	41	11
3.2.1. Perennial Crop*	36	12
3.2.2.1. Coffee (beta)*	46	13
3.2.2.2. Citrus (beta)*	47	14
3.2.2.3. Other Perennial Crops*	48	15
3.2.2. Temporary Crop*	19	16
4.5. Rocky Outcrop	29	17
4.3. Other Non Vegetated Areas	25	18
5.1. River, Lake and Ocean	33	19
1.1.1. Forest Formation	3	23
1.1.2. Savanna Formation*	4	24

Class	Pixel value	Prevalence
2.1. Wetland	11	26
2.2. Grassland	12	27
3.1. Pasture	15	29
3.3 Mosaic of Uses	21	30

\* *Absent classes in the Pampa biome maps*

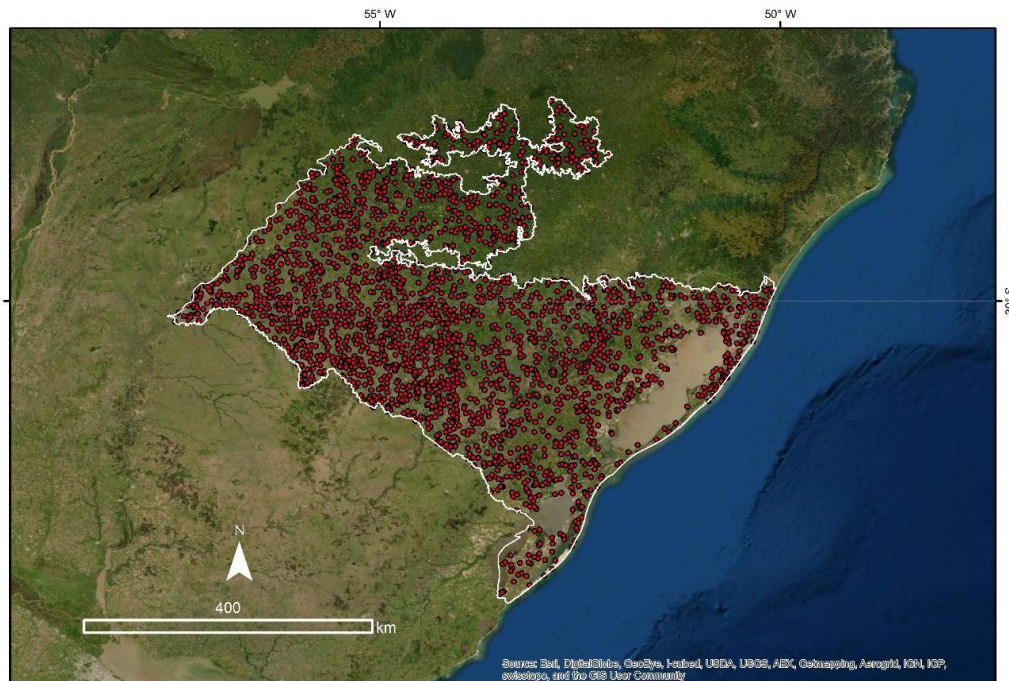
## 9 VALIDATION STRATEGIES

### 9.1 Accuracy assessment

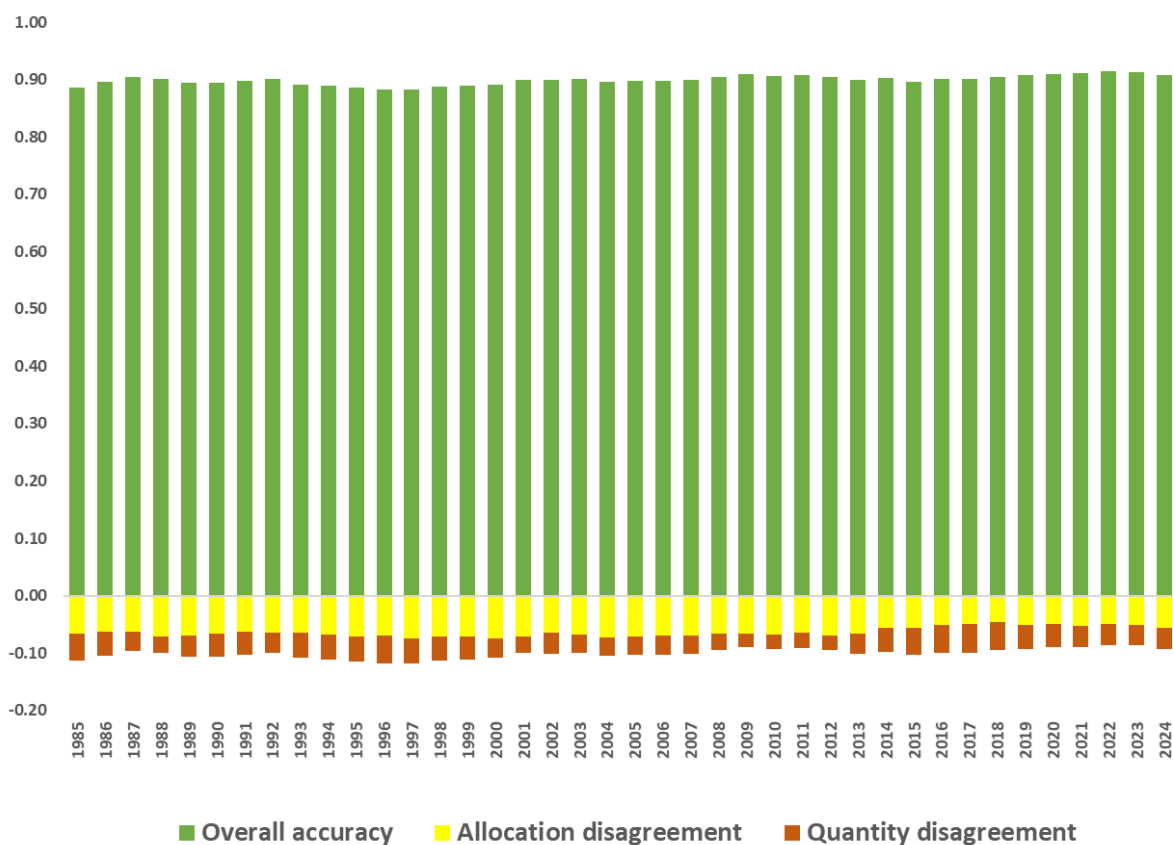
A dataset of 2.568 independent validation points provided by Lapig (*Laboratório de Processamento de Imagens e Geoprocessamento - UFG*) was used to perform accuracy analysis in the Pampa biome (**Figure 13**). This dataset allowed us to calculate the confusion matrix and extract several metrics, including: overall accuracy, quantity and allocation disagreements, as well as classes accuracy and their associated omission and commission errors.

The mean global accuracy (1985-2024) of Collection 10 was 89,9% at levels 2 and 3 of the legend (**Figures 14 and 15**). The mean allocation disagreement was 7.3% and the quantity disagreement 2.9%. These results are the best achieved across all collections, considering the average overall accuracy combined with the lowest interannual variability. The lowest accuracy recorded was 88.18%, in 1996, while the highest was 91.43%, in 2022. The lower accuracy in the early years of the series, which was a chronic issue in previous collections, was better addressed in Collection 10.

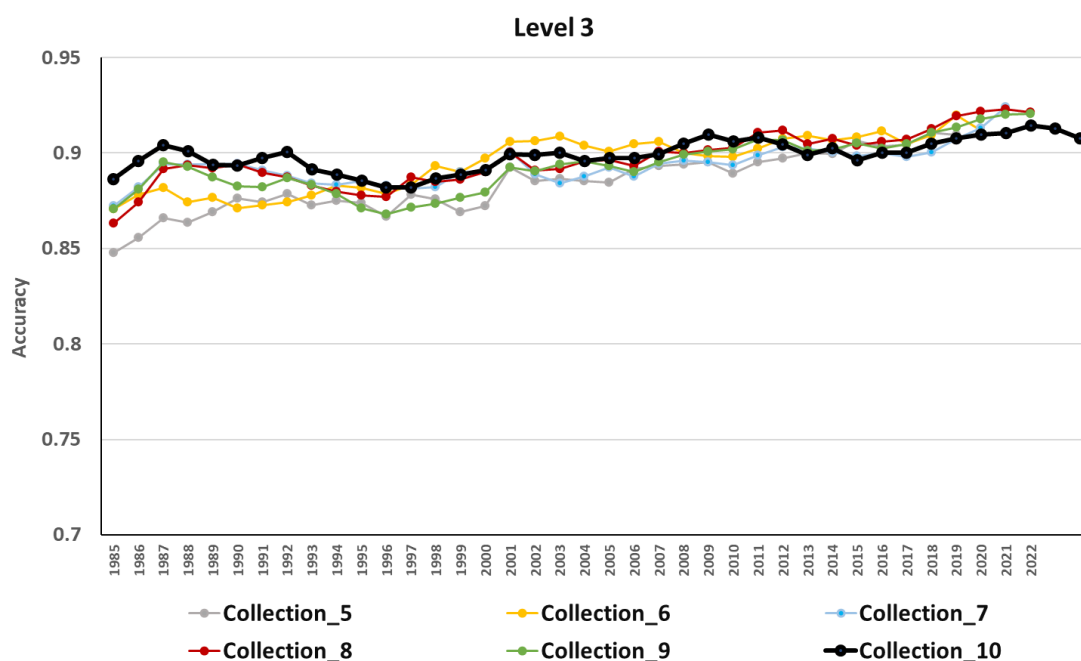
**Figure 16** shows the commission and omission errors for eight mapped classes. The forest class represents the sum of classes 3 and 49, the grassland class represents the sum of classes 12 and 50, and the temporary crops class is equivalent to the sum of classes 39, 40, and 41. The results indicate that the classes Forest, Grassland, Urban areas, and Water achieved high accuracies, with low commission and omission errors. Conversely, the worst performances were obtained by Mosaic of uses and Wetland classes. However, Mosaic of uses performed better than in collection 10, despite continuing to have high commission values. On the other hand, it is necessary to be cautious when interpreting commission and omission errors for the Wetland class. Sampling insufficiency and classification problems were detected in the reference samples, so the true accuracy of this class is significantly underestimated.



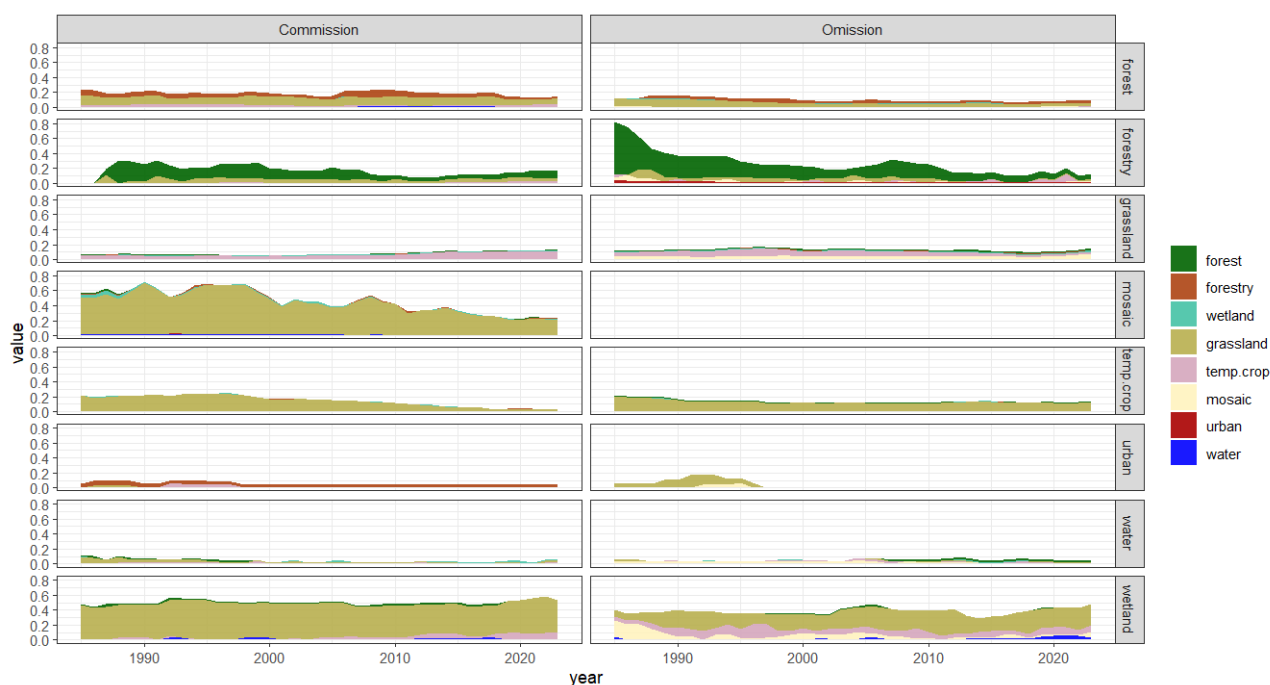
**Figure 13:** Dataset with reference samples used in the accuracy assessment of Collection 10 in the Pampa biome.



**Figure 14:** Overall accuracy (green), quantity disagreement (red) and allocation disagreement (yellow) along the years of Collection10 maps, Pampa biome.



**Figure 15:** Global overall accuracy of Collections 5, 6, 7.1, 8, 9 and 10 in the Pampa biome, considering the classes at level 3 of the legend.



**Figure 16:** Commission and Omission errors of Collection 10 in the Pampa biome, considering the classes at level 3 of the legend.



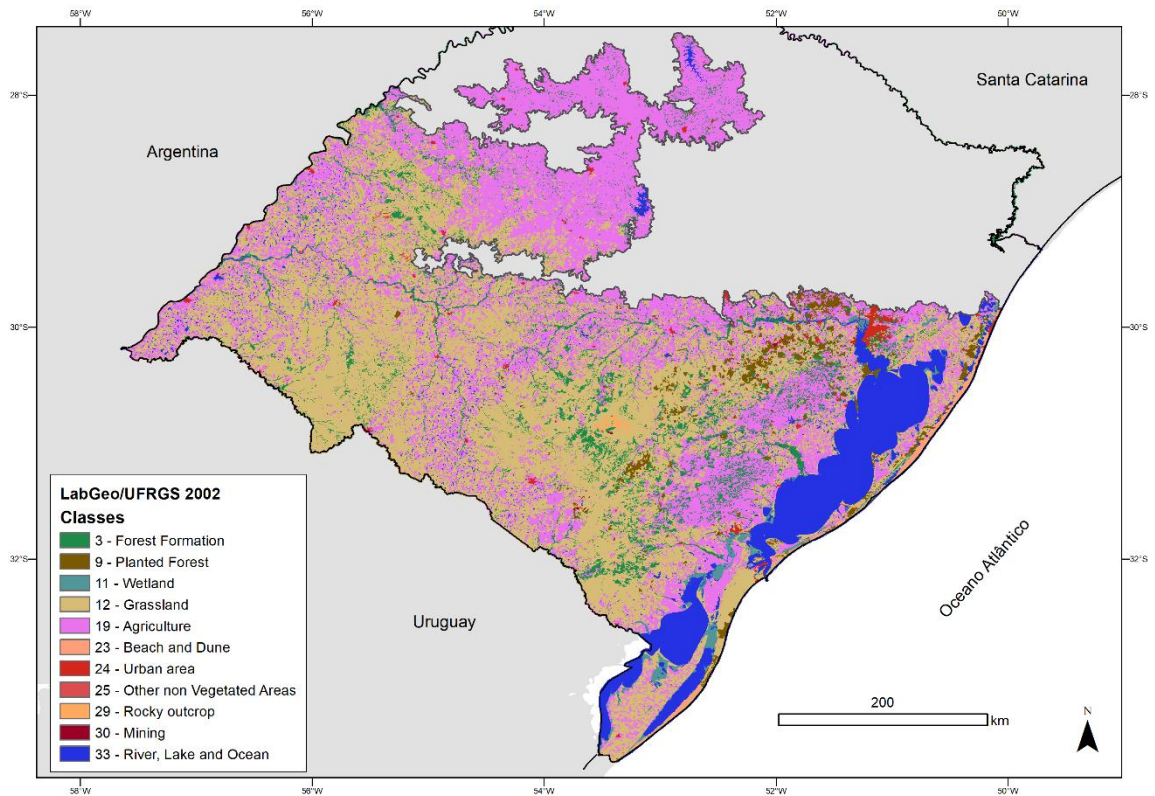
## 9.2 Agreement assessment

Overall, there have been few previous initiatives on LULC mapping the Pampa biome with spatial and thematic detail compatible with the MapBiomass Project. Basically, there are three maps that depict the years 2002 (Hasenack *et al.*, 2015), 2009 (Weber *et al.*, 2016) and 2015 (Hofmann *et al.*, 2018) for the state of Rio Grande do Sul, which includes the Pampa biome.

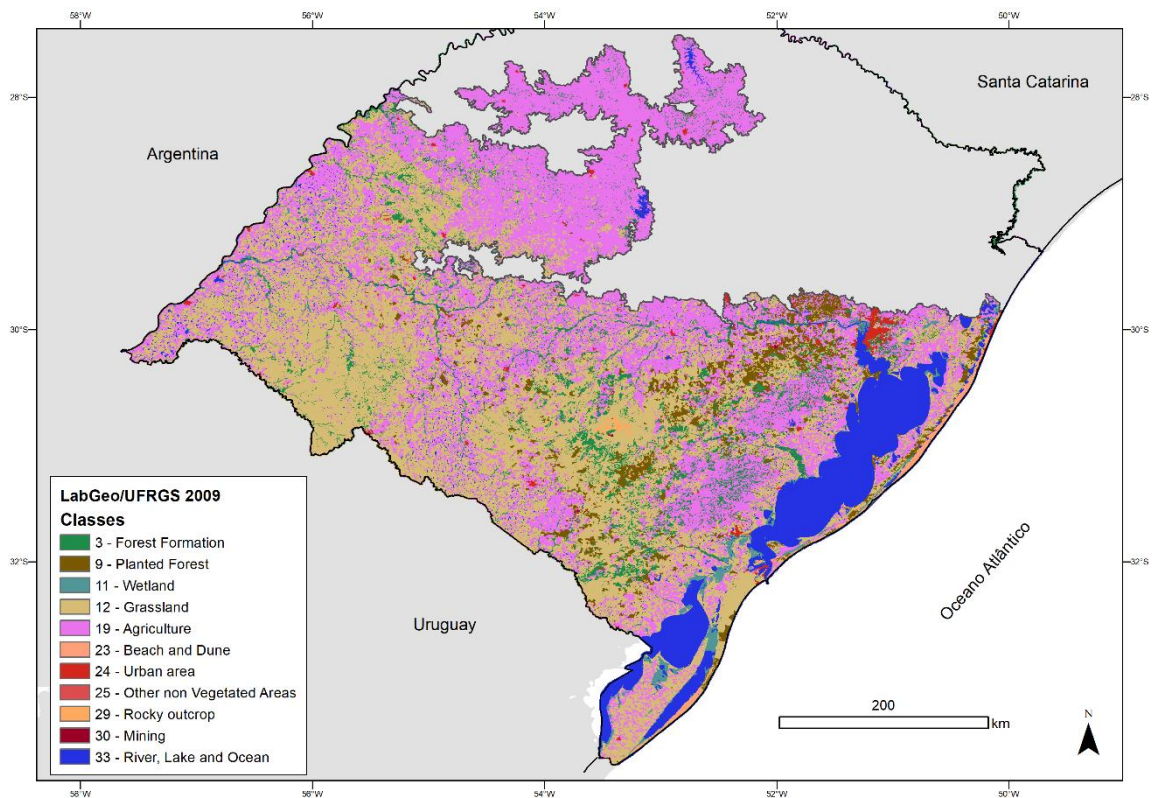
All of the mentioned maps were produced through visual interpretation of Landsat imagery, therefore being vector polygon maps aiming to ensure a level of spatial detail compatible with cartographic scale 1:250,000. Their thematic richness, on the other hand, comprises a number of categories of natural vegetation cover and anthropic uses that is larger than those of MapBiomass Collection 9. Thus, for validation purposes, the three available vector maps were first rasterized with a spatial resolution of 30 meters, like the MapBiomass maps, then reclassified according to the MapBiomass legend (**Figures 17, 18, 19**), and finally used as reference for an analysis of agreement with the digital classification maps of the same years. Different from accuracy assessment, based on sampling points, here the whole surface of each MapBiomass map, pixel by pixel, was compared with the reference map of the corresponding year.

The overall agreement between the MapBiomass LULC maps and the reclassified reference maps was 70.94% for the year 2002, 71.21% for 2009 and 72.52% for 2015. Allocation disagreement was the major component of disagreement (Pontius and Millones, 2011), reaching 23.35% in 2002, 21.32% in 2009 and 19.97% in 2015. This reflects the different nature of both sets of maps: the reference maps are inherently more generalized due to the manual polygon drawing aiming a specific cartographic scale, while the MapBiomass' digital classification maps are pixel-based, so that a single polygon delineating a unique class in the former can encompass several pixels of different classes in the latter.

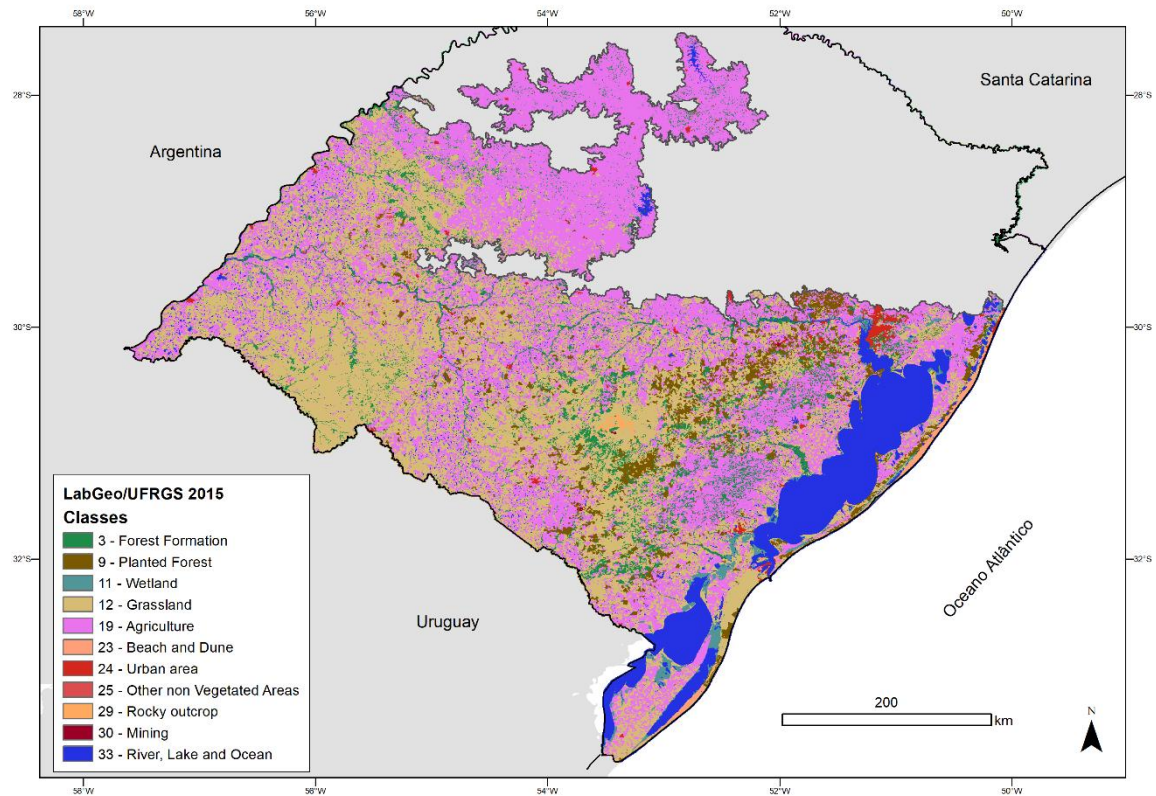
Average value for the three years was 71.56% for overall agreement, 22.55% for allocation disagreement and 6.89% for quantity disagreement. Summing overall agreement and allocation disagreement gives an indicator of area agreement, with an average of 93.11%.



**Figure 17.** Reference map for the Pampa biome, year 2002 (Hasenack *et al.*, 2015), reclassified to the legend of Collection 10 of the MapBiomas project.



**Figure 18.** Reference map for the Pampa biome, year 2009 (Weber *et al.*, 2016), reclassified to the legend of Collection 10 of the MapBiomas project.



**Figure 19.** Reference map for the Pampa biome, year 2015 (Hofmann *et al.*, 2018), reclassified to the legend of Collection 10 of the MapBiomas project.

## 10 REFERENCES

- Breiman, L. Random forests. **Machine learning**, v. 45, n. 1, p. 5-32, 2001.
- Hasenack, H.; Cordeiro, J.L.P; Weber, E.J. (Org.). 2015. **Uso e cobertura vegetal do Estado do Rio Grande do Sul – situação em 2002**. Porto Alegre: UFRGS IB Centro de Ecologia. 1ª ed.
- Hasenack, H.; Weber, E.; Boldrini, I. I. Trevisan, R. 2010. **Mapa de sistemas ecológicos das Savanas Uruguaias em escala 1:500.000 ou superior**. Porto Alegre: Centro de Ecologia. Relatório técnico Projeto UFRGS/TNC. 18 p.
- Hofmann, G.S.; Weber, E.J; Hasenack, H. (Org.). **Uso e cobertura vegetal do Estado do Rio Grande do Sul – situação em 2015**. Porto Alegre: UFRGS IB Centro de Ecologia, 2018. 1a ed.
- IBGE. 2018. **Base Temática Vetorial 1:250.000 – Brasil - Pedologia**. Rio de Janeiro: IBGE, 2018.
- IBGE. 2019. **Biomass e sistema costeiro-marinho do Brasil: compatível com a escala 1:250.000**. Rio de Janeiro: IBGE, 2019. 168 p. (Relatórios metodológicos, v. 45)
- Pontius, R.G., Millones, M., 2011. Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. **Int. J. Remote Sens.** v. 32, p. 4407–4429. <http://dx.doi.org/10.1080/01431161.2011.552923>.
- Pontius, R.G.; Krithivasan, R.; Sauls, L.; Yan, Y.; Zhang, Y.; Methods to summarize change among land categories across time intervals. **Journal of Land Use Science**, v.12, n.4, p. 218-230. 2017.
- Weber, E.J.; Hofmann, G.S.; Oliveira, C.V.; Hasenack, H. (Org.). **Uso e cobertura vegetal do Estado do Rio Grande do Sul – situação em 2009**. Porto Alegre: UFRGS IB Centro de Ecologia, 2016. 1a ed. ISBN 978-85-63843-20-3.