



Pantanal - Appendix

Collection 5

Version 1

General coordinator

Marcos Reis Rosa

Team

Eduardo Reis Rosa

Viviane Cristina Mazin

1. Overview of the Pantanal classification

The production of the Collection 5, with land cover and land use annual maps for the period of 1985-2019, followed a sequence of steps in the Pantanal biome, similar to those used in the previous Collections 3 and 4 (Figure 1). However, some improvements were added up, particularly in balance of samples, inclusion of a Probability classification of Natural Classes (forest, savanna, grassland, flooded grassland and water) and in the post classification filters. Furthermore, the geographical official limits of the biome were updated (IBGE 2019).

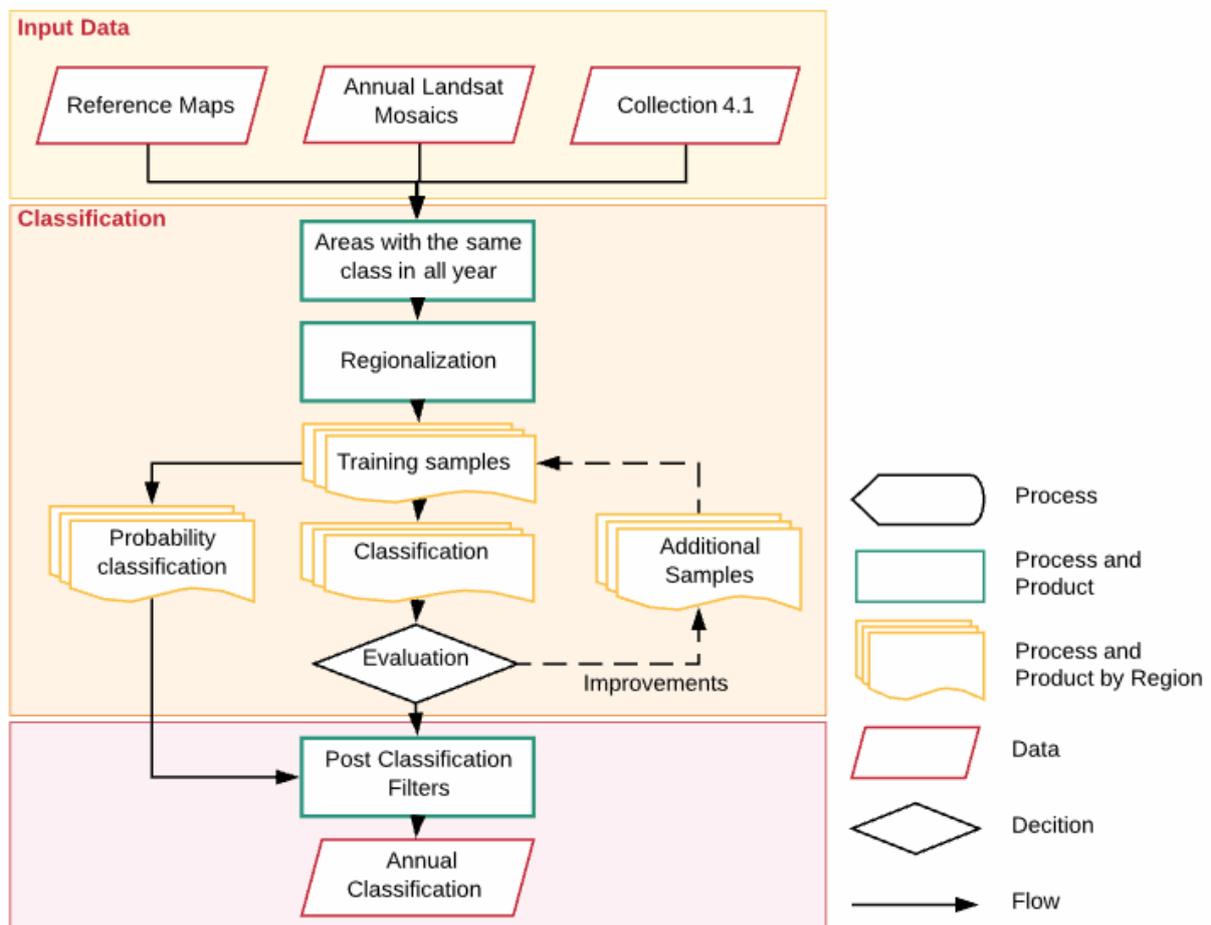


Figure 1. Classification process of Pantanal classification in the MapBiomias Collection 5.

2. Landsat image mosaics

2.1. Definition of the temporal period

The image selection period for the Pantanal biome was defined aiming the selection in the dry season (Figure 2). The use of images in the driest period of the Pantanal reduces the occurrence of wetlands that can reach areas of natural fields and pastures. It also helps to detect the variations in the natural grasslands and pastures, as

well as reduces possible confusions in the identification of the areas of Savanna and Forest Formations in the floodplain.

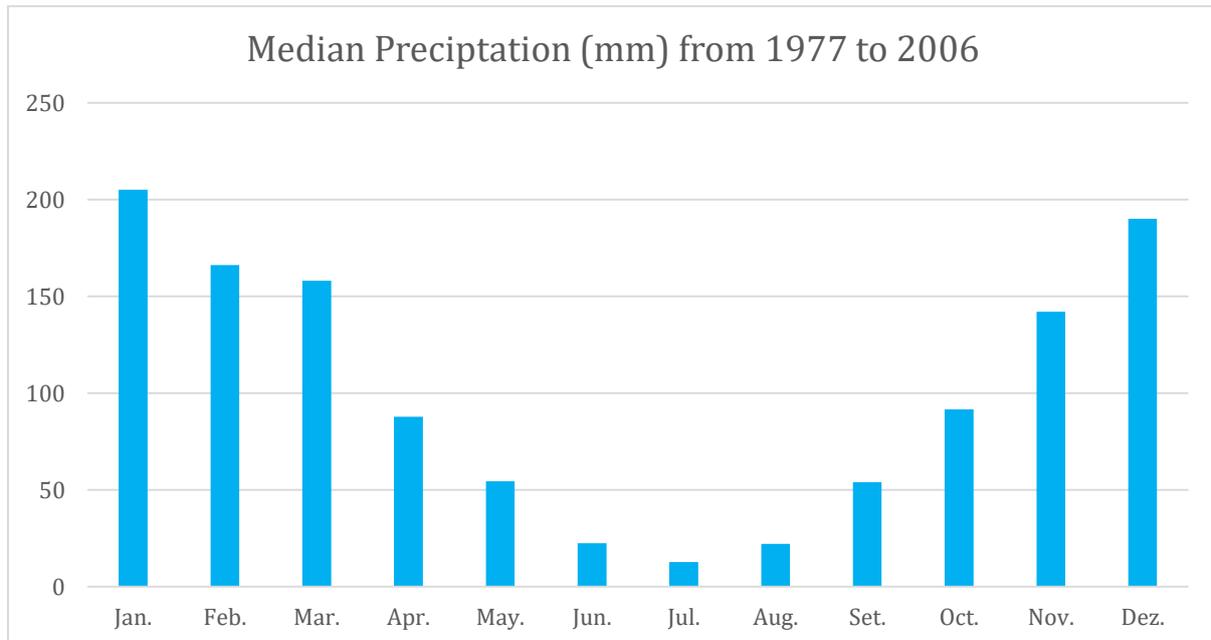


Figure 2. Median monthly precipitation values from 1977 to 2006 (MARCUIZZO et al., 2010).

2.2. Image selection

For the selection of Landsat scenes to build the mosaics of each chart for each year, within the acceptable period, a threshold of 50% of cloud cover was applied (i.e., any available scene with up to 50% 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. When needed, due to excessive cloud cover and/or lack of data, the acceptable period was extended to encompass a larger number of scenes in order to allow the generation of a mosaic without holes. Whenever possible, this was made by including months in the beginning of the period, in the winter season.

In most cases the dry period from May 1st to 30 to August was the best to get a good mosaic with none or few missing information caused by clouds and shades.

For each year we used the best Landsat images available from:

- 1985 to 1999 – Landsat 5
- 2000 to 2002 – Landsat 7
- 2003 to 2011 – Landsat 5
- 2012 – Landsat 7
- 2013 to 2019 – Landsat 8

2.3. Final quality

As a result of the selection criteria, most of the mosaics presented satisfactory quality. The first years of the collection still have some noise caused by haze.

2.4. New set of mosaics

We exported a new set of mosaics for the Collection 5 based on a visual selection of images that still had atmospheric noises after cloud and shade filters. These mosaics were used as a complement to the mosaics from Collection 3.

3. Definition of regions

Classifying more homogenous regions reduces confusion of samples and classes and allows a better balance of samples. Thus the Pantanal biome was divided into six regions based in dry and wet areas and latitude (Figure 3).



Figure 3. Regions of Pantanal biome defined for the classification of Collection 5.

4. Classification

4.1. Classification scheme

The digital classification of the Landsat mosaics in the Pantanal biome aimed to individualize a subset of eight classes from the MapBiomias Collection 5 legend (Table 1), which were integrated with the cross-cutting themes in a further step.

Table 1. Land cover and land use classes considered for digital classification of Landsat mosaics in the Pantanal biome in the MapBiomias Collection 5.

Land cover and land use classes	Numeric ID	Color
1.1.1. Forest Formation	3	
1.1.2. Savanna Formation	4	
2.1. Flooded Grassland and Swamped Area	11	
2.2. Grassland Formation	12	
3.3 Mosaic of Agriculture or Pasture	21	
4.5. Other non-vegetated area	25	
5. Water	33	
6. Non Observed	27	

The land cover class 11 (Flooded Grassland and Swamped Area) has a new concept in the Collection 5 which represents the flood in each year without any temporal filter to better represent the Pantanal flood dynamic.

Permanently flooded areas or flooded areas that flooded at least once a year, particularly due to the flood pulse.

The pulse of water flooding over the countryside is a natural factor. Areas with a water surface, but difficult to classify due to the amount of macrophytes, eutrophication or sediments, were also included in this category.

The woody element may be present on the field matrix forming a mosaic with shrub or tree plants

4.2. Feature space

The feature space for digital classification of the categories of interest for the Pantanal biome comprised a subset of 38 variables (Table 2), taken from the complete feature space of MapBiomias Collection 3. The feature space included the original Landsat reflectance bands, as well as vegetation indexes, spectral mixture modeling-derived variables, terrain morphometry (slope), and a spatial texture measure. The definition of this subset was based on the ability of each variable to discriminate the targets of concern, considering local knowledge about their spectral, spatial and temporal dynamics.

Table 2. Feature space subset considered in the classification of the Pantanal biome Landsat image mosaics, 1985-2019.

'slope',	'median_gcvi_wet',	"median_ndfi_dry",	"stdDev_ndfi",
'textG',	'median_gcvi',	"median_ndfi_wet",	"stdDev_fns",
"median_blue",	'median_gcvi_dry',	"median_ndvi",	"stdDev_soil",
"median_evi2",	'median_hallcover',	"median_ndvi_dry",	"stdDev_gvs",
"median_green",	"median_gv",	"median_ndvi_wet",	'min_nir',
"median_nir",	"median_gvs",	"median_ndwi",	'amp_soil',
"median_swir1",	"median_npv",	"median_ndwi_dry",	'amp_ndfi'
"median_swir1_wet",	"median_soil",	"median_ndwi_wet",	"stdDev_ndfi",
"median_swir1_dry",	"median_shade",	"median_savi",	
"median_swir2",	"median_ndfi",	"median_sefi",	

4.3. Classification algorithm, training samples and parameters

Digital classification was performed by region, year by year, using a Random Forests algorithm (Breiman, 2001) available in *Google Earth Engine*. Training samples for each chart were defined following a strategy of using pixels for which the land cover and land use remained the same along the 34 years of Collection 4.1, so named “stable samples”. An ensemble taken from three main sources for the training samples was made: extracted from Collection 3.1; manually drawn polygons; and complementary samples.

4.3.1 Stable samples from Collection 4.1

The extraction of stable samples from the previous Collection 4.1 followed several steps aiming to ensure their confidence for use as training areas. First, based on a visual analysis, a threshold was established for each class, specifying a minimum number of years in which a pixel should remain with that class to be eligible as a stable sample. A layer of pixels with a stable classification along the 34 years of Collection 4.1 was then generated by applying such thresholds. From the resulting layer of stable samples, a subset 2,000 samples were randomly generated and balanced for each class based on the class cover percentage. A minimum of 600 samples was used for rare classes that do not cover at least 10% of the region area.

4.3.2. Complementary samples

The need for complementary samples was evaluated by visual inspection. Complementary sample collection was done by drawing polygons using Google Earth Engine Code Editor. The same concept of stable samples was applied, checking the false-color composites of the Landsat mosaics for all the 35 years during the polygon drawing. Based on the knowledge of each region samples from forest, savanna, grassland or flooded grassland were added. Samples from forests that were not well represented in the stable map were also added when needed.

4.3.3. Final classification

Final classification was performed for all regions/years with stable and complementary samples. All years used the same subset of samples and it was trained in the same mosaic of the year that was classified.

4.3.4. Probability classification

Natural classes of Forest, Savanna, Grassland, Water and Flood Grassland (Figure 4) were classified as single classes with the result as a probability map using both mosaics from Collection 3 and new mosaics from Collection 5.

The probability classification was integrated to the final classification. We have applied a threshold for each class and used the pixels above this threshold to update the final classification map using the following order:

- > 90% of chance to be forest;
- > 95% of chance to be grassland;
- > 90% of chance to be savanna;
- > 50% of chance to be water.

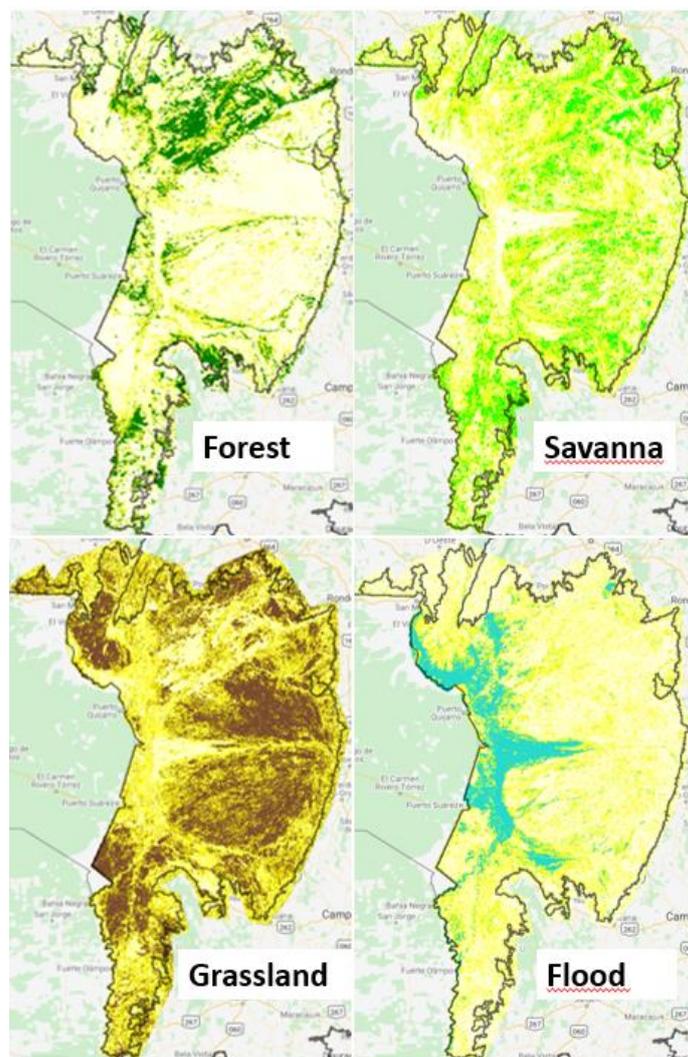


Figure 4. Probability maps for each natural vegetation class in the Pantanal biome.

5. Post-classification

5.1. Gap Fill filter

In this filter, no-data values (“gaps”) are theoretically not allowed and are replaced by the temporally nearest valid classification. In this procedure, if no “future” valid position is available, then the no-data value is 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.

5.2. Spatial filter

The spatial filter avoids unwanted modifications to the edges of the pixel groups (blobs), a spatial filter was built based on the "connectedPixelCount" function. Native to the GEE platform, this function locates connected components (neighbors) that share the same pixel value. Thus, only pixels that do not share connections to a predefined number of identical neighbors are considered isolated. In this filter, at least six connected pixels are 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 6pixels (~0,5 ha).

5.3. Temporal filter

The temporal filter uses the subsequent years to replace pixels that have invalid transitions.

In the first process the filter looked for any natural class (12, 4, 3, 21) that is not this class in 1985 and is equal to these values in 1986 and 1987 and then corrects 1985 values to avoid any regeneration in the first year.

In the second process the filter looked at the pixel value in 2018 that is not 21 and it is equal to 21 in 2016 and 2017. The value in 2018 was then converted to 21 to avoid any regeneration in the last year.

The third process looks in a 3-year moving window to correct any value that was changed in the middle year and return to the same class next year. This process was applied in this order: 12, 3, 4, 21, 33.

The last process is similar to the third process but it was a 4- and 5-years moving window that corrected all middle years.

5.4. Frequency filter

A frequency filter was applied only in pixels that are considered “stable natural vegetation” (at least 33 years as [3, 11, 12, 4]). If a “stable natural vegetation” pixel was at least 60% of the years in the same class, all years were changed to this class. The result of this frequency filter was a classification with more stable classification between natural classes (ex: forest and savanna). Another important result was the removal of noises in the first and last year in the classification.

5.5. Regeneration filter

A new filter was applied to prevent Forest or Savanna Formations converted to

anthropic to regenerate as Grassland Formation. Thus, it was eliminated part of the confusion between pasture areas and Grassland Formation.

5.6. Flooded areas

The probability map of flood was converted to an annual flood map using a threshold of 70%.

After all post classification filters, a mask with pixels that has flooded more than five times in 35 years was applied to convert forest and savanna to grassland.

The flood map was applied each year to replace the grassland (12) to flooded grassland (11), showing the flooded area of each year without any temporal filter.

6. Validation strategies

The validation of each process was produced using independent validation points provided by LAPIG/UFG. We used all points that both interpreters considered the same class, resulting in more than 1,600 validation points.

7. References

BREIMAN, L. Random forests. Machine learning, v. 45, n. 1, p. 5-32, 2001.