



Pantanal - Appendix

Collection 4

Version 1

General coordinator

Marcos Reis Rosa

Team

Eduardo Reis Rosa

Viviane Cristina Mazin

1. Overview of classification method

The production of the Collection 4, with land cover and land use annual maps for the period of 1985-2018, followed a sequence of steps in the Pantanal biome, similar to those used in the previous Collection 3.1 (**Figure 1**). However, some improvements were added up, particularly in the geographical units for classification and in the post classification filters.

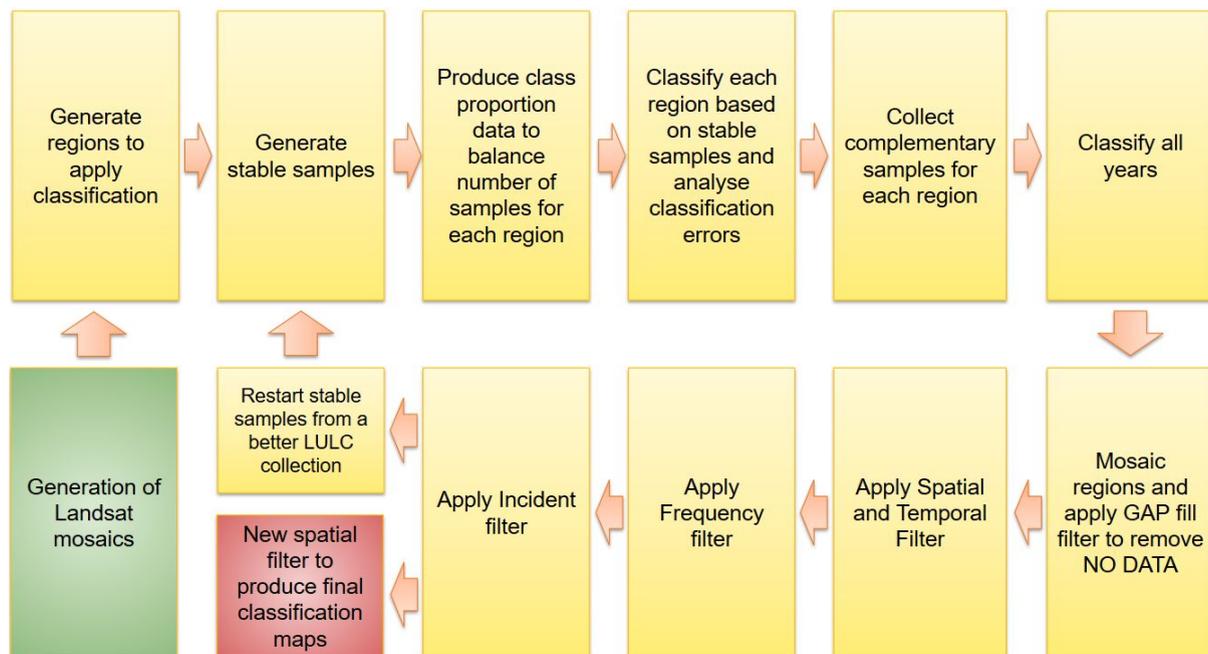


Figure 1. Classification process to produce Collection 4 in the Pantanal biome.

2. Landsat image mosaics

2.1. Definition of the temporal period

The image selection period for the Landsat image mosaics in the Pantanal biome was defined aiming the selection in the dry season (Figure 2) to reduce the wetlands. The use of images in the driest period in 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 fields and pastures and reduces possible confusions in the identification of the areas of Forested Savannas and Forests existing in the plain and that also is influenced by the periodic floods.

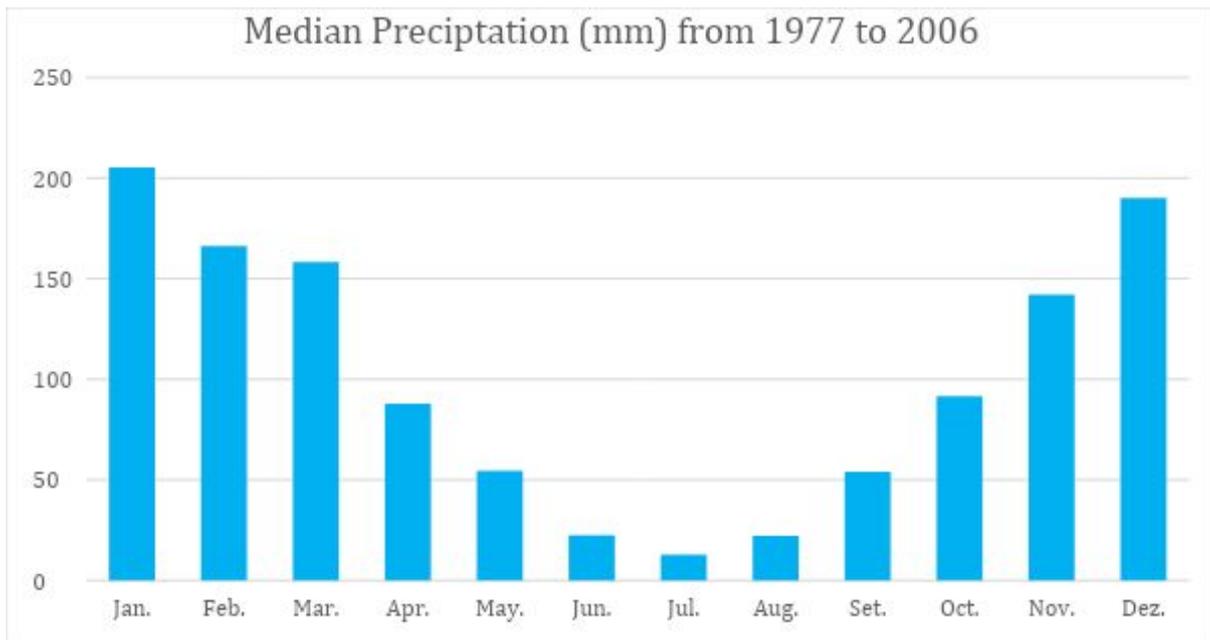


Figure 2. Median monthly precipitation values from 1977 to 2006 (MARCUIZZO et al., 2010).

2.2. Image selection

For the selection of Landsat scenes to build the mosaics of each chart for each year, within the driest 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 dry/winter season.

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

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

- 1985 to 1999 – Landsat 5
- 2000 to 2002 – Landsat 7
- 2003 to 2011 – Landsat 5
- 2012 – Landsat 7
- 2013 to 2017 – Landsat 8

2.3. Final quality

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

3. Definition of regions for classification

In Collection 3.1, the sheets of the World International Chart to the Millionth (1:250.000), hereafter called 'charts', were the spatial units adopted for data processing. A total of 22 charts were used to cover the biome. Each chart sets the geographical limits to build up the temporal and spatial Landsat mosaics, to collect training samples and to proceed with digital classification procedures. The final map of the Pantanal biome was generated merging these 22 units.

In Collection 4 the charts approach was applied only to manage the Landsat mosaics. For the years 1985-2017 the same mosaics were used with the addition of new ones for 2018. All the following steps were based on newer geographical units, corresponding to six homogeneous regions based on wet and dry areas and border with Amazon and Cerrado Biomes. (**Figure 2**).

This new approach reduce confusion of samples and classes and improve the balance of samples, and avoid abrupt transitions when merging the geographical units of classification.

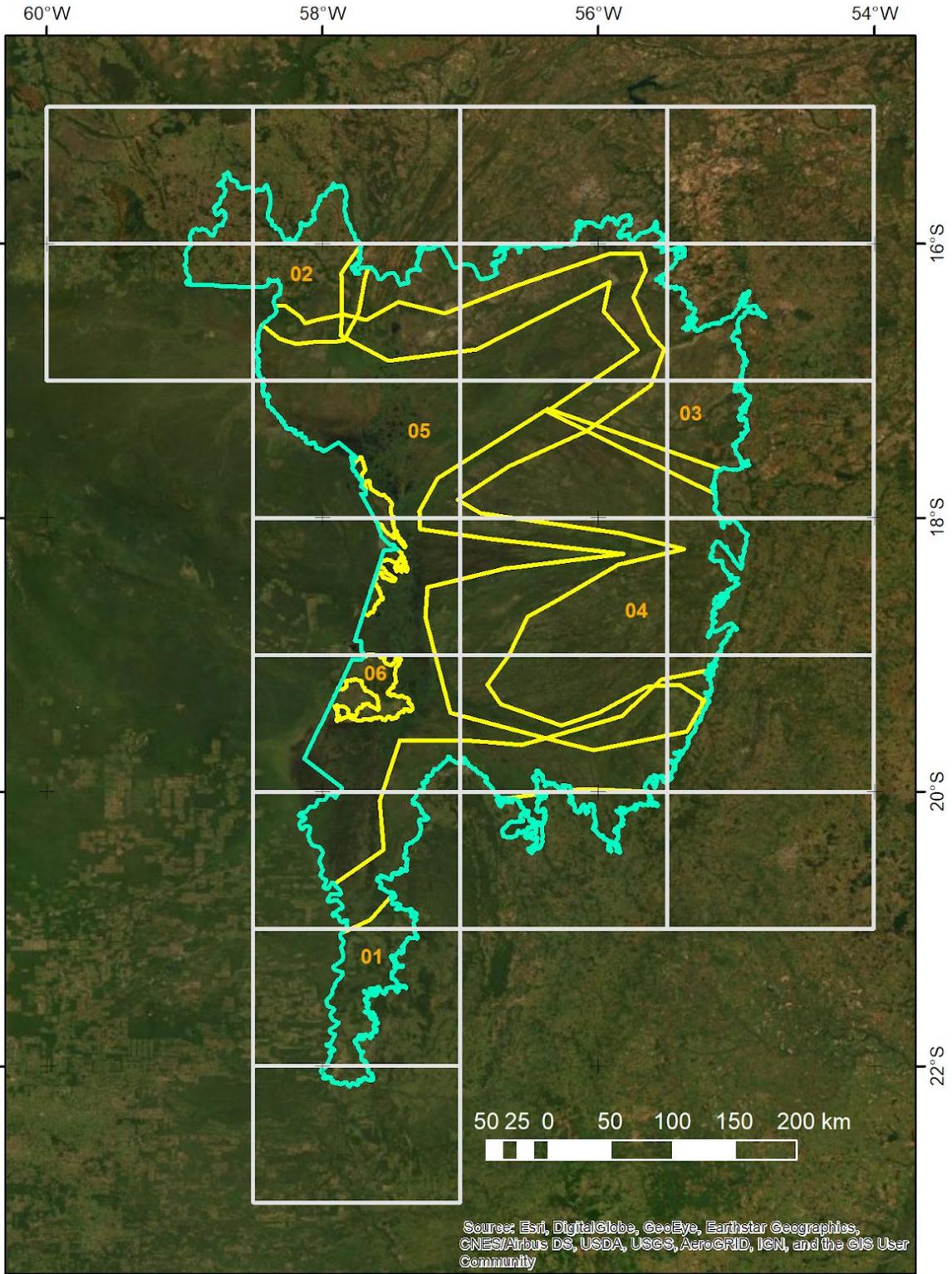


Figure 3. Regions used in the classification of the Pantanal biome.

4. Classification

4.1. Classification scheme

The digital classification of the Landsat mosaics for the Pantanal biome aimed to individualize a subset of eight land cover and land use classes from the complete legend of MapBiomias Collection 4 (Table 1), which were integrated with the cross-cutting themes in a further step.

Table 1. Land cover and land use categories considered for digital classification of Landsat mosaics for the Pantanal biome in the MapBiomias Collection 4.

Legend class of Collection 4	Numeric ID	Color
1.1.1. Forest Formation	3	
1.1.2, Savanna Formation	4	
2.1. Wetland	11	
2.2. Grassland	12	
3.3 Mosaic of Agriculture and Pasture	21	
5. Water	33	
6. Non Observed	27	

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 4. They include 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 the subset was made based on the expected usefulness 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 in the MapBiomas Collection 4 (1985-2018).

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

4.3. Classification algorithm, training samples and parameters

Digital classification was performed region by region, year by year, using a Random Forest algorithm (Breiman, 2001) available in *Google Earth Engine*, running *XX iterations (random forest trees)*. Training samples for each region were defined following a strategy of using pixels for which the land cover and land use remained the same along the 33 years of Collection 3.1, so named “stable samples”. An ensemble taken from three main sources was made: extracted from Collection 3.1; manually drawn polygons; and complementary samples.

4.3.1. Stable samples from collection 3.1

The extraction of stable samples from the previous Collection 3.1 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 33 years of Collection 3.1 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 used to rare classe that does 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 means of 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 34 years during the polygon drawing. Based in the knowledge of each regions samples from Forest, Savanna, Grassland or Wetlands were added. Samples from Forests that were not well represented in the stable map were also added where need.

4.3.3. Final classification

Final classification was performed for all regions and 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.

5. Post-classification

Due to the pixel-based classification method and the long temporal series, a list of post-classification spatial and temporal filters was applied. The post-classification process includes the application of gap-fill, temporal, spatial and frequency filters. The temporal filter rules were adapted for the land cover and land use classes used in the Pantanal biome and were complemented by specific rules to adjust for cases where a pixel appeared.

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 avoid 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 (neighbours) that share the same pixel value. Thus, only pixels that do not share connections to a predefined number of identical neighbours 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 search for any native vegetation class (12, 4, 3, 21) that is not this class in 85 and is equal in 86 and 87, and then corrects 85 value to avoid any regeneration in the first year of the time series.

In the second process the filter search for a pixel value in 2018 that is not 21 (Mosaic of Agriculture and Pasture) and is equal 21 in 2016 and 2017. The value in 2018 is 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 is 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 is a 4- and 5-years moving window that corrects all middle years.

5.4. Frequency filter

Frequency filters were applied only in pixels that were considered "stable native vegetation" (at least 33 years as [3, 11, 12, 4]). If a "stable native vegetation" pixel is at least 60% of years of the same class, all years are changed to this class. The result of these frequency filters is a classification with more stable classification between navel classes (e.g. Forest and Savanna). Another important result is the removal of noises in the first and last year in the classification.

5.5. Incident filter

An incident filter were applied to remove pixels that change too much times in the 34 years. All pixels that change more than eight times and are connected to less than 22 pixels that also change more than eight times weres replaced by the MODE value. This incident filter avoids changes in the border of the classes.

6. Validation strategies

A new set of 2,008 independent validation points provided by Lapig (*Laboratório de Processamento de Imagens e Geoprocessamento - UFG*) was used to perform accuracy analysis (**Figure 5**).

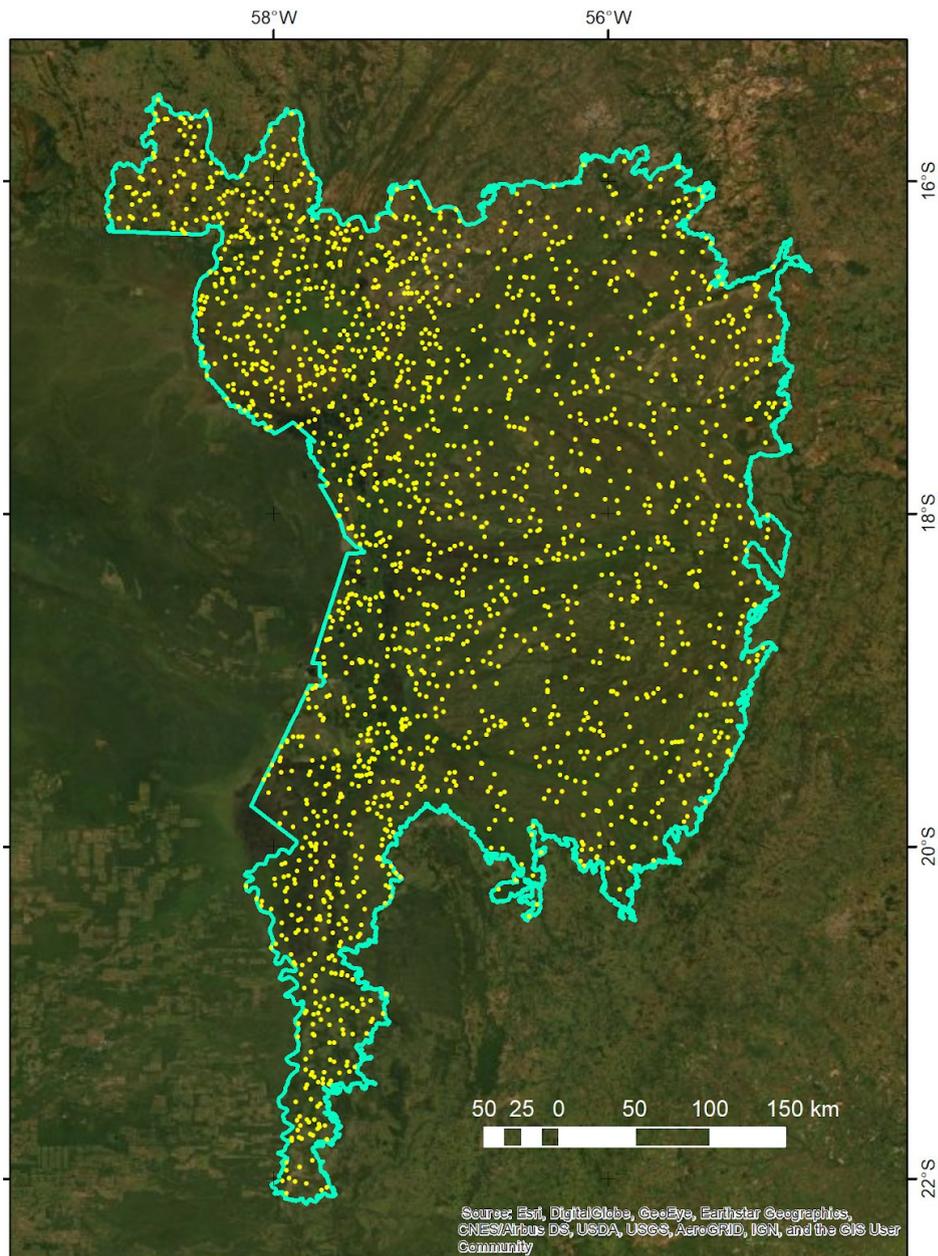


Figure 5. Accuracy points in Pantanal.

To evaluate the improvements of different classification filters of Collection 4 we performed a comparative analysis of accuracy, including the previous Collection 3.1. Those classes presented only in the integration were previously remapped to the corresponding classes in

the biome classification to allow direct comparison of accuracy results between classifications with and without the integration with cross-cutting themes.

Accuracy results indicate positive outcomes for the different filters and better results for Collection 4 when compared with Collection 3.1 (**Figure 6**).

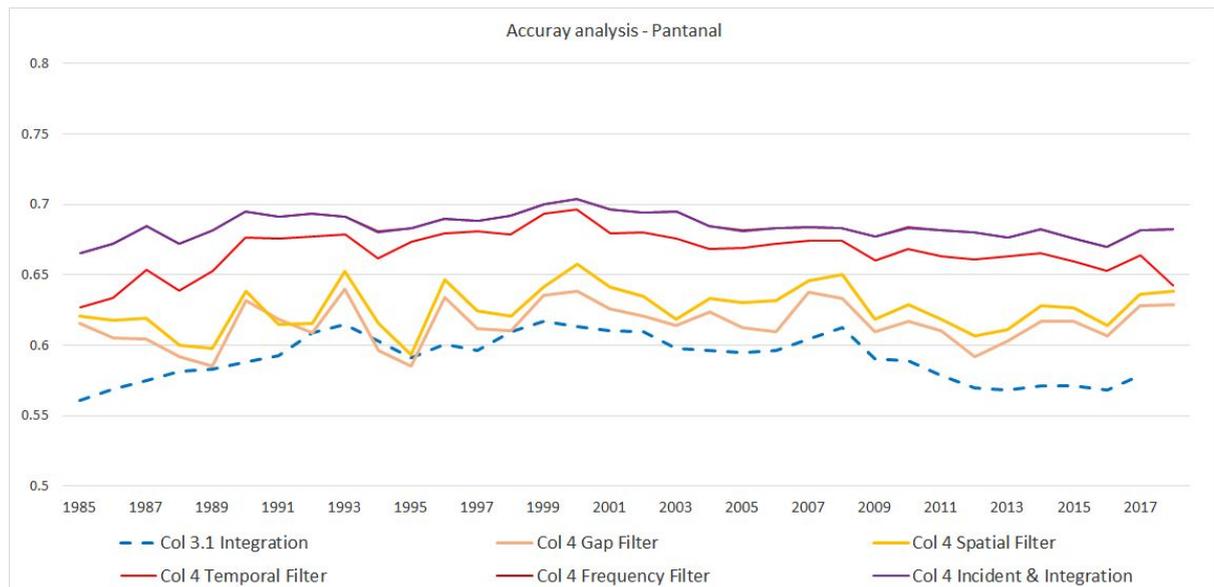


Figure 6. Accuracy of Pantanal classification, including different classification filters in the Collection 4 and Collection 3 comparison.

7. References

- Breiman, L. Random forests. Machine learning, v. 45, n. 1, p. 5-32, 2001.
- MARCUZZO, F. et al. Chuvas no Pantanal brasileiro: análise histórica e tendência futura. **Anais 3º Simpósio de Geotecnologias no Pantanal**, p. 5, 2010.