



## **Pantanal - Appendix**

**Collection 8**

**Version 1**

**General coordinator**

Eduardo Reis Rosa

**Team**

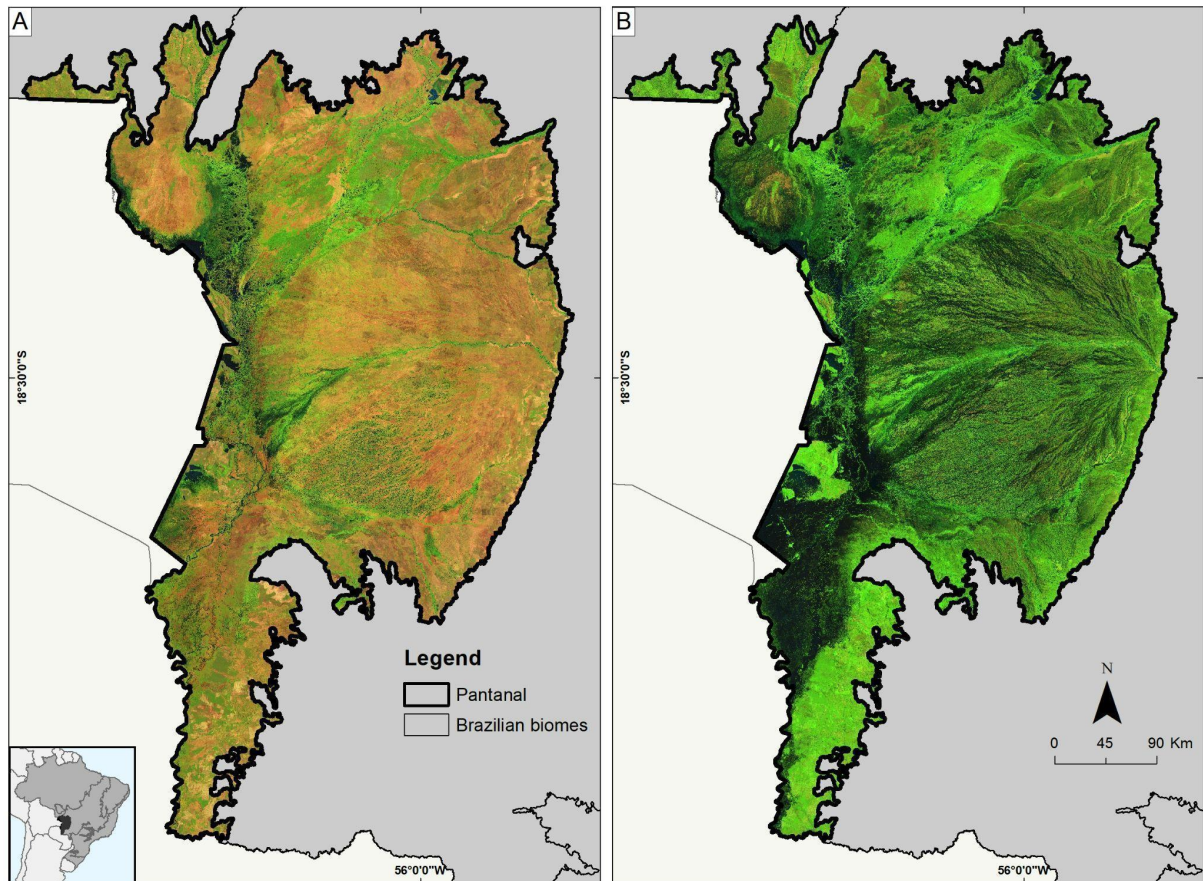
Marcos Reis Rosa

Mariana Dias

## 1. Introduction

The Pantanal biome is located within Brazilian territory between the Mato Grosso and Mato Grosso do Sul states. It also extends into the territories of the neighboring countries of Brazil, Bolivia, and Paraguay. It covers an area of 150,000 km<sup>2</sup> and is surrounded by the Amazon, Cerrado, and Atlantic Forest biomes. Despite being the smallest Brazilian biome in area, it boasts the highest proportion of remaining native vegetation.

The Pantanal is renowned as one of the largest continuous wetlands in the world, characterized by a vast alluvial plain subject to seasonal flooding due to a flood pulse, with pronounced flood and dry periods. (Figure 1). This flood pulse is influenced by global and regional climatic conditions and the hydro-sedimentological dynamics of the rivers in the Upper Paraguay Basin (HAMILTON; SOUZA; COUTINHO, 1998; PADOVANI, 2010).



**Figure 1:** Location of the Pantanal biome showing RGB color composites of TM bands 5, 4 and 3 for Landsat mosaics of the year 1988 for. A) dry season Landsat mosaic; B) wet season Landsat mosaic.

Understanding the dynamics of flooding in the biome is crucial to comprehending changes in land use and land cover over time. As part of the MapBiomas Collection 8, we conducted monthly mapping of water and wetland classes and made other improvements compared to previous collections.

This document describes the methodological flow used to map the land use and land cover in the Pantanal biome from 1985 to 2022. Also, all codes used are available on our public GitHub (<https://github.com/mapbiomas-brazil/pantanal>).

## **2. Overview of classification method**

The initial classification of the Pantanal biome within the MapBiomas project consisted of applying decision trees to generate annual maps of the predominant native vegetation (NV) types, distinguished in: Forest, Savanna, and Grassland. The method used to generate these annual maps evolved, with significant improvements from the first MapBiomas Collection to the present.

Collection 1.0 covered the period of 2008 to 2015 and was published in 2016. Collections 2.0 and 2.3 covered the period of 2000 to 2016 and were published in 2018.

The classification using Random Forest was implemented in Collection 2.3, and from this point onward, the empirical decision tree was used generate stable samples, which were classified as the same NV type over the considered period (2000-2016). These stable samples were used to train the Random Forest models to classify the entire time series. Collections 3 and 3.1 expanded the period covered to 1985–2017.

Collections 4 and 5 used training samples based on the stable samples from the previous collection and reference maps. Collection 6.0 used stable samples from collection 5.0. Collections 7 and 7.1 (published in 2022) used stable samples from collection 6.0. This Collection 8.0 used stable samples from Collection 7.1 and covered the period from 1985 to 2022 (Table 1).

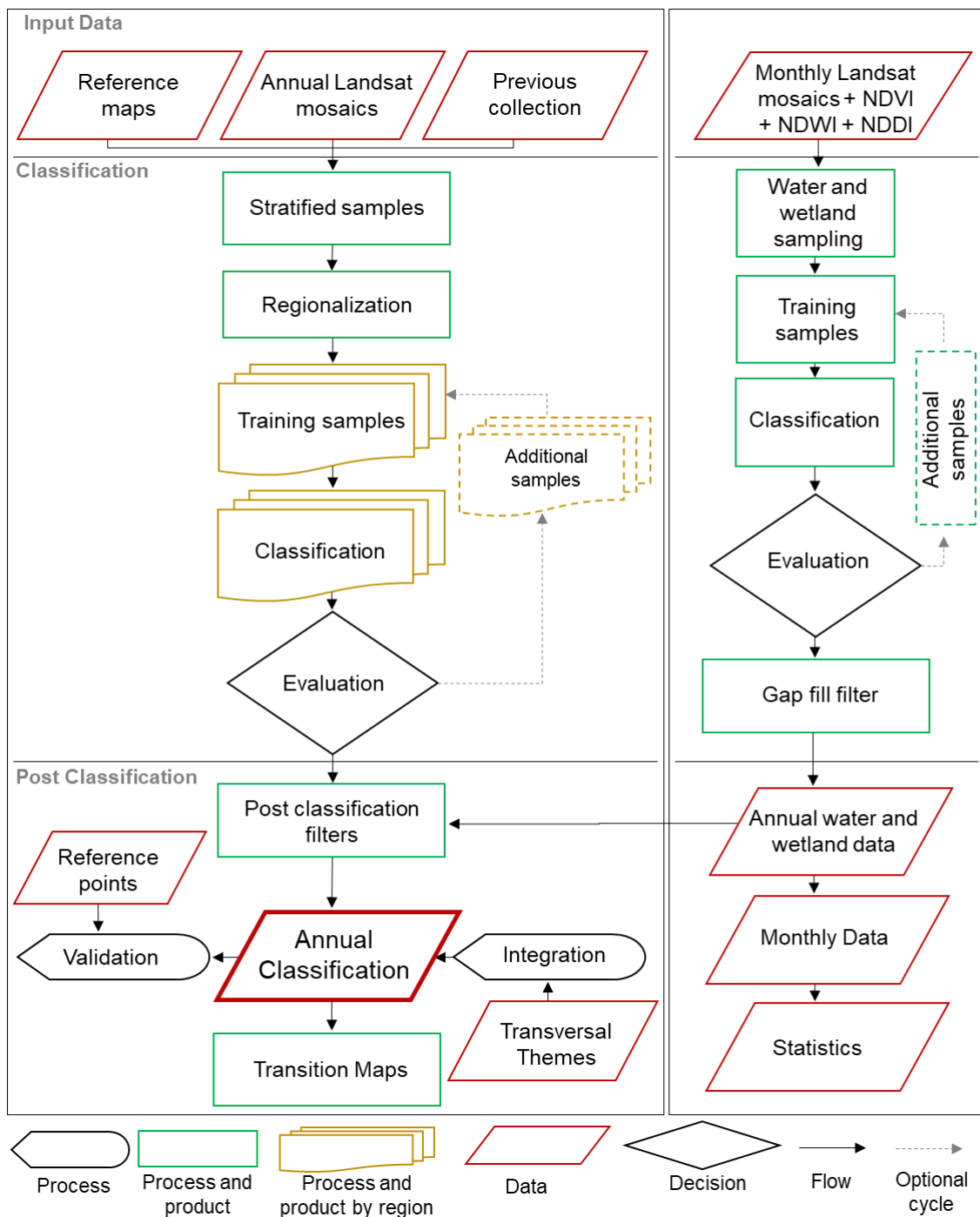
**Table 1.** The evolution of the Pantanal mapping collections in the MapBiomass Project, its periods, hierarchical level and number of classes, brief methodological description, and overall accuracy in level 1 and level 2.

Collection	Period	Levels /N. Classes	Method	Overall Accuracy
Beta & 1	8 years 2008-2015	1 / 7	Empirical Decision Tree	
2.0 & 2.3	16 years 2000-2016	3 / 13	Empirical Decision Tree & Random Forest (2.3)	
3.0 & 3.1	33 years 1985-2017	3 / 19	Random Forest	Level 1: 73.2% Level 3: 62.1% *
4.0 & 4.1	34 years 1985-2018	3 / 19	Random Forest	Level 1: 80.7% Level 3: 71.0% *
5.0	35 years 1985-2019	4 / 21	Random Forest	Level 1: 85.1% Level 3: 78.3% *
6.0	36 years 1985-2020	4 / 24	Random Forest	Level 1: 81.6% Level 2: 73.5%
7.0	37 years 1985-2021	4 / 27	Random Forest	Level 1: 85.9% Level 2: 79.5%
7.1	37 years 1985-2021	4 / 27	Random Forest	Level 1: 85.6% Level 2: 77.3%
8.0	38 years 1985-2022	4 / 29	Random Forest	Level 1: 84.0% Level 2: 77.6%

\* Due to hierarchy changes in the forest classes, level 2 of Collections 6 - 8 are being compared to level 3 of previous collections.

The production of the Collection 8.0 land use and land cover maps in the Pantanal biome followed steps similar to those used in the previous Collections (4 - 7.1). However, some improvements were added, particularly in the mosaics, a new regionalization approach, new balance of samples, the annual wetland and water classification based on monthly data, and improvement in the post-classification filters.

The classification workflow is summarized in Figure, 2 and each step is described in the following sessions.

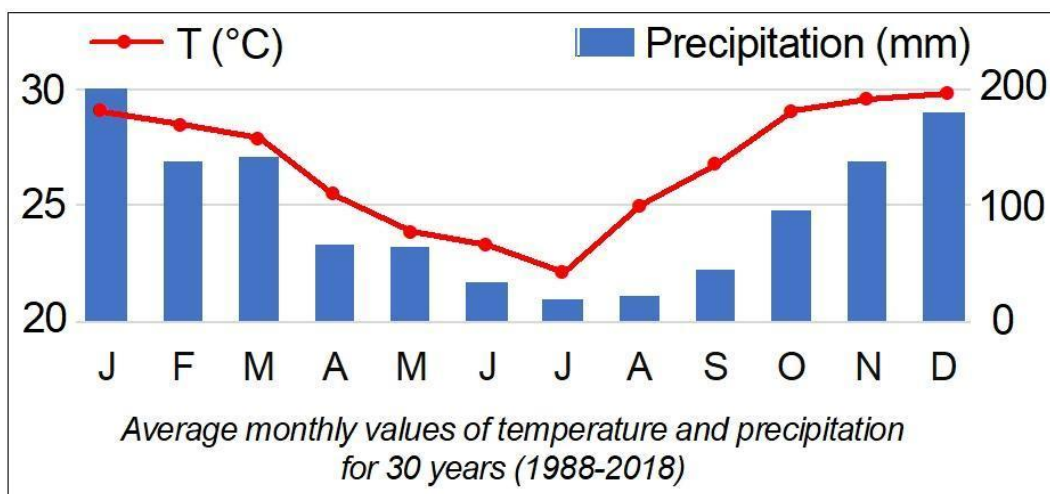


**Figure 2:** Workflow to generate Collection 8.0 land use and land cover maps for Pantanal.

### 3. Landsat image mosaics

#### 3.1 Temporal period and image selection

The image selection period for the annual mosaics in the Pantanal biome was defined considering that the dry season (Figure 3), would reduce the wetlands. Using images in the driest period reduces the identification of wetlands that can reach areas of grasslands and pastures, helping to detect the variations in both classes. It also helps to reduce possible confusion in the identification of Forested Savannas and Forests areas existing in the plain because they are also influenced by periodic floods.



**Figure 3:** Climograph from 1988 to 2018 with CHIRPS (Climate Hazards Group InfraRed Precipitation with Station) precipitation data and MODIS (Moderate-Resolution Imaging Spectroradiometer) temperature data. (FUNK; PETERSON; LANDSFELD, 2015).

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

Until Collection 5.0 the classification was performed by using top of atmosphere (TOA) data from Landsat 5 (TM), 7 (ETM+) and 8 (OLI) sensors. In collection 6.0, we discontinued the TOA data and adopted the use of surface reflectance (SR) data. Since collection 7.0 we adopted the ‘USGS Landsat Level 2, Collection 2, Tier 1’ dataset that is available on the Google Earth Engine platform, maintaining the use of SR data.

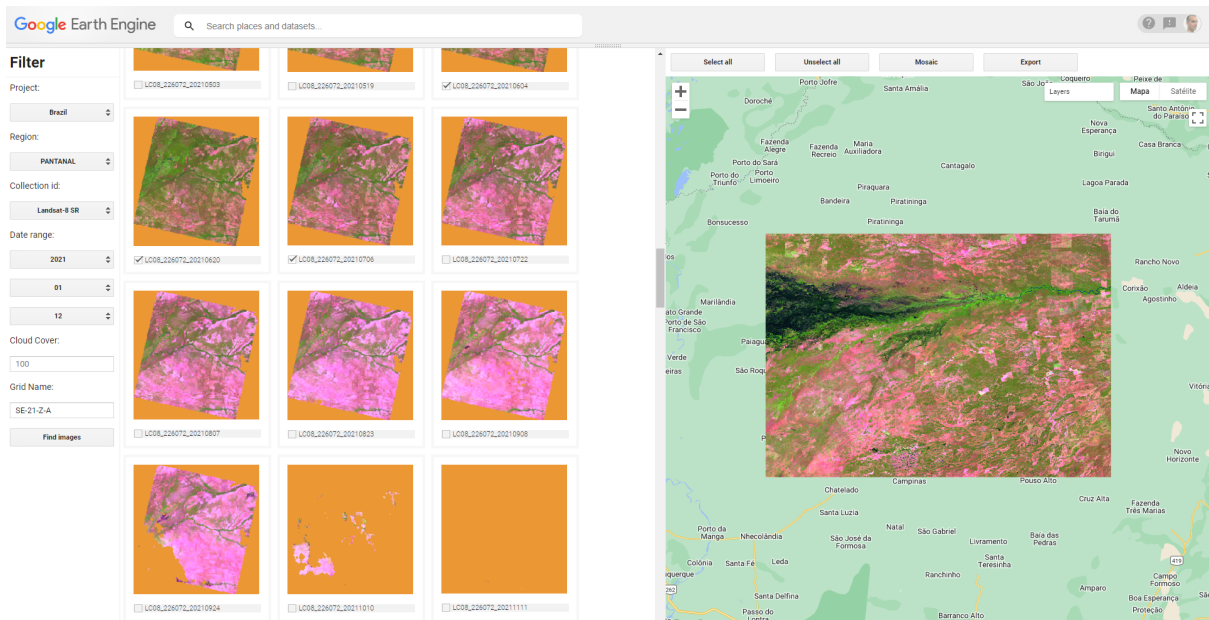
For each year we used images from the best Landsat sensor available:

**Table 2:** Landsat satellite for each period

Satellite	Years
Landsat 5	1985-1999; 2003-2011
Landsat 7	2000-2002; 2012
Landsat 8	2013-2022

The annual mosaics were built computing the median of the pixel values of all the available images in a period defined by an initial and a final date. A threshold of 50% of cloud cover was applied for the selection of Landsat scenes to be used in the mosaics of each year (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 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, that process was made by including months in the beginning of the period, in the winter season.

We made a visual analysis on the preliminary mosaics to identify and remove images with noises (clouds, shadow, or sensor defect) for each year (Figure 4).



**Figure 4:** Google Earth Engine tool used to identify and remove scenes with noise

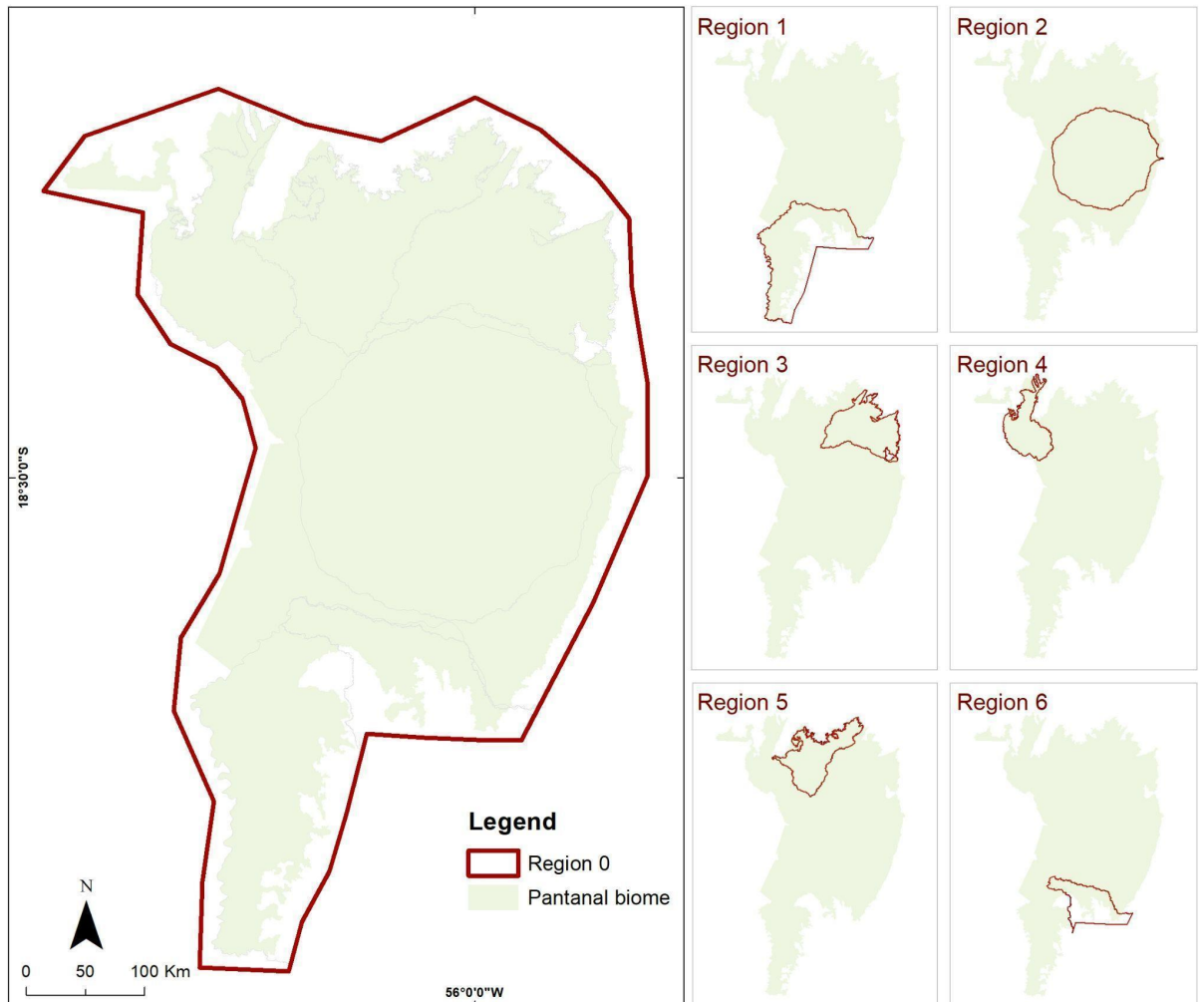
This process was conducted for each year (1985 - 2022), resulting in the annual Landsat mosaics, which were submitted to classification.

Most mosaics presented satisfactory quality as a result of the selection criteria. The first years of the collection (Landsat 5) still have some noise caused by haze.



#### 4. Regionalization

The Pantanal biome was divided into seven regions (Figure 5) based on dry and wet patterns, sub watersheds and native vegetation distribution presented on different regionalization approaches (SILVA & ABDON, 1998; ASSINE, 2015). The aim of this process was to reduce the confusion and classification noise improving the sample's balance with more homogeneous regions.







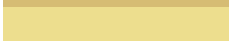



**Figure 5:** Regionalization used in the classification of the Collection 8 in the Pantanal biome.



## 5. Classification

The digital classification of the Landsat mosaics for the Pantanal biome aimed to individualize a subset of 8 classes from the complete legend of MapBiomias Collection 8 (Table 3), which were subsequently integrated with the cross-cutting themes in a further step (see section 7).

**Table 3:** Classes mapped by the Pantanal biome in the MapBiomias Collection 8.

Legend class of Collection 8	Numeric ID	Color
1.1.1. Forest Formation	3	
1.1.2. Savanna Formation	4	
2.1. Wetland	11	
2.2. Grassland Formation	12	
3.1. Pasture	15	
3.2.1.5. Other Temporary Crops	41	
4.5. Other non-vegetated area	25	
5. Water	33	

The classes known as Forest, Savanna, Grassland, Pasture, Other Temporary Crops and Other non-vegetated area were mapped based on annual Landsat mosaics. On the other hand, the water and wetland classes represent a minimum frequency of two monthly Landsat mosaics classified as these humid classes.

### 5.1 Stable Samples

The extraction of stable samples from the previous Collection 7.1 followed several steps aiming to ensure their confidence for use as training areas. The areas that did not change class from 1985 to 2021 in Collection 7.1 were used to generate 7.000 random training points in each region. During the classification the balance was done reducing the number of stable samples adjusted for each class.

The samples from forest, savanna, grassland, pasture and water were filtered using data from Global Forest Canopy Height (GFCH), 2019 (POTAPOV et al., 2019) based on GEDI data using the following rules:

- Forest: stable samples and  $GFCH > 7$
- Savanna: stable samples and  $GFCH > 3$  and  $< 6$
- Grassland and Pasture: stable samples and  $GFCH < 2$
- Water: stable samples and  $GFCH < 1$

### 5.2 Complementary samples

The need for complementary samples was evaluated by visual inspection. Complementary sample collection was made in Google Earth Engine Code Editor with the drawing polygons tool. The same concept of stable samples was applied, checking the false-color composites of the Landsat mosaics for all the 38 years during the polygon drawing. Based on the knowledge of each region, samples of forest, savanna, grassland or pasture were added.

### 5.3 Feature Space

A subset of 38 variables (Table 4) was defined as feature space for the Pantanal classification. These variables were composed of the original Landsat surface reflectance bands, spectral indexes and auxiliary data. The subset was defined based on a feature importance analysis produced with Random Forest classification applied in band and spectral indexes with 500 interactions.

**Table 4:** Feature space subset considered in the classification of the Pantanal biome Landsat image mosaics, 1985-2022.

Feature Space Subset			
Type	Band Name	Type	Band Name
Spectral Bands	green_median	Spectral Indexes	gcvi_median
	green_median_wet		gcvi_median_dry
	green_min		gcvi_median_wet
	red_median		gcvi_stdDev
	red_median_dry		nddi_dry
	red_median_wet		nddi_wet
	red_min		ndvi_median_dry
	nir_median		ndvi_median_wet
	nir_median_wet		ndwi_median
	swir1_median		ndwi_median_dry
	swir1_median_dry		ndwi_median_wet
	swir1_median_wet		ndwi_stdDev
	swir2_median		savi_median
	swir2_median_dry		savi_median_dry
swir2_median_wet	savi_median_wet		
Spectral Indexes	cai_median		wefi_median_wet
	evi2_median	Auxiliar Data	GEDI
	evi2_median_dry	Spatial Reference	latitude
	evi2_median_wet		longitude

## 5.4 Classification algorithm

A digital classification was performed year by year using a Random Forest algorithm (BREIMAN, 2001) available in Google Earth Engine. Random Forest was trained with stable and complementary samples, according to which feature space subset. Each subregion had a specific sample balance and classification. After the classification processing, the subregions were merged to form the Pantanal biome territory.

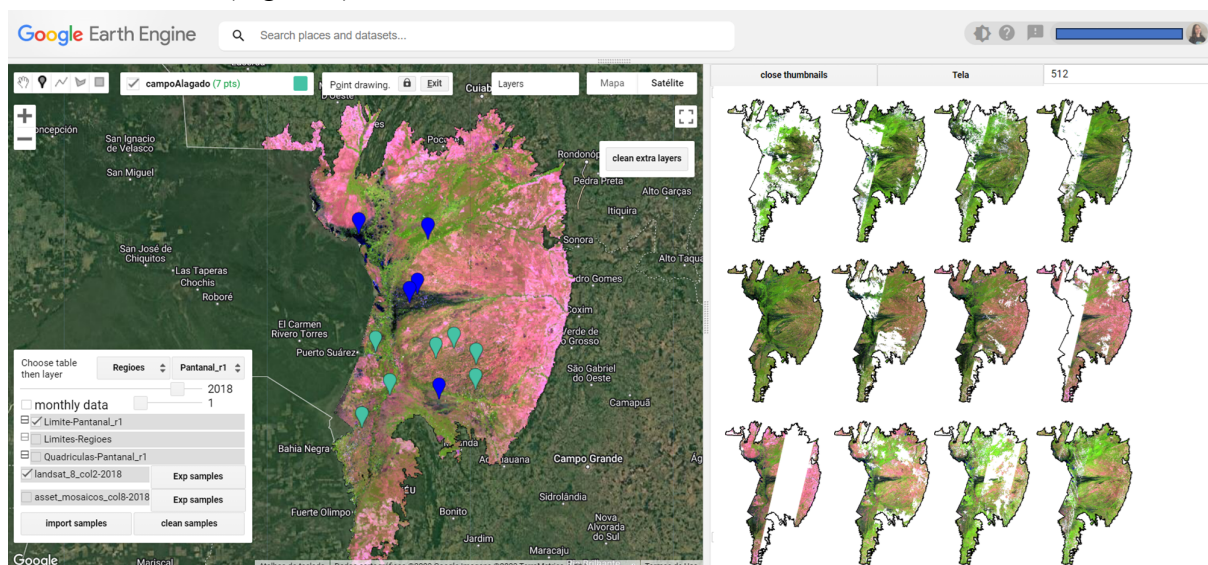
## 5.5 Water and Wetland Monthly and Annual Classification

For the first time we performed a monthly classification approach in Google Earth Engine platform. To guarantee the monthly classification of the maximum available images, we used monthly mosaics of each satellite and also mosaics composed of data from different satellites if they were available for the same period (Table 5).

**Table 5:** Satellites for each period

Satellite	Years
Landsat 5	1985-1999; 2003
Landsat 5 + 7	2000-2002; 2004-2011
Landsat 7	2012
Landsat 8	2013-2021
Landsat 8 + 9	2022

A sampling toolkit was developed to collect water and wetland points on the monthly Landsat mosaics (Figure 6).



**Figure 6:** Monthly sampling toolkit.

For classes considered 'non-wet', we raffled points in pixels that had a stable classification over the 37 years of the Collection 7.1. Each sample point was trained by Random Forest and all Landsat bands plus Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI) and Normalized Difference Drought Index (NDDI).

Then, a supervised classification was carried out using the Random Forest (RF) algorithm (BREIMAN, 2001) to classify each monthly mosaic into three classes: water, wetland and 'non-wet' areas.

The RF classification result was used to mask the 'non-wet' monthly mosaic areas and then a water-wetland separation threshold was defined for each satellite based on the NDDI value and applied on the 'wet' RF masked area.

Some monthly Landsat mosaics had no data pixels because of a lack of images or the cloud mask filter action. To overcome this challenge we developed a 'gap fill' filter on the classified data. The filter considered the local precipitation pattern (Figure 3) to fill each no data pixel with the classification of the closest and driest month available. We consider this a conservative strategy to not overestimate the mapping of the Pantanal flooded areas in periods without images, since the wettest months are also those with the highest levels of rainfall and, therefore, the cloudiest.

With the available monthly water and wetland classification for each year since 1985 it was possible to generate frequency data. To integrate the annual maps from MapBiomas Collection 8.0 we considered all pixels classified at least twice in a year as water or wetland classes, the water class was predominant over wetland class in pixels classified as both within a year. The monthly data detailing can be accessed at:

<https://gee-arcplan.users.earthengine.app/view/pantanalmensal>.

## **6. Post Classification**

### **6.1 Masks**

Some natural grassland areas in the western portion of the Pantanal can be easily confused and classified as exotic pastures by RF because of the overgrazing or the soils exposition during a dry period. To avoid this misclassification, a mask of 'non-pasture' was manually drawn based on historical reference maps (CI et al., 2009) and high resolution images, and then applied for excluding pixels misclassified as pasture. For the same purpose, a second 'non-pasture' mask was applied considering flooded pixels (classified monthly as water or wetland) in more than 75% of the entire series.

Also, in order to improve the classification for the most recent years, a mask of 'deforested areas' was build from the MapBiomas Alertas validated polygons. Following, all native vegetation pixels intercepting the mask were reclassified to anthropic class in the

deforestation event year and in the following years. It is worth remembering that these deforestation polygons are available from 2019.

## **6.2 Water and Wetland data integration**

Annual water and wetland data were added to the map of 'non-wet' classes only in areas classified as grasslands, considering that the methodology for the latter is not adapted to identify flooded forest and savannah. This is also a strategy to avoid false positives from humid areas in the shade of the relief, urban areas or roads.

## **6.3 Filters**

Some temporal and spatial filters were applied to the annual maps to correct classification errors or invalid class transitions with the help of long-term information from the temporal series.

### **6.3.1 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.

### **6.3.2. Transitions Forest Formation x Savanna Formation**

A frequency filter was applied only in pixels that are considered forest or savanna formation (classes 3 and 4 in Table 3). If a “stable natural vegetation” pixel is at least 60% of years of the same class and the first and last year are of this class, all years are changed to this class. The result of this frequency filter is a more stable classification of these natural classes (forest and savanna). Another important result is the removal of noises in the first and last year in classification.

### **6.3.3. Moving window temporal filter**

The temporal filter uses the subsequent years to replace pixels that have invalid transitions, and is applied in three steps. The first step 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 is applied in this sequential order [41, 3, 4, 15, 12] (classes listed in Table 3) . The second step is similar to the first process, but it is a 4-years moving window that corrects all middle years.

In the third step the filter looks at any native vegetation class [12, 3, 4] that is not this class in 1985 and is equal in 86 and 87 and then corrects 85 values to avoid any changes in the first year. In the last process the filter looks for the pixel value in 2022 that is not [12,15]

and is equal to 15 in 2020 and 2021. The value in 2022 is then converted to [12,15] to avoid any false regeneration in the last year.

#### 6.3.4. Regeneration filter

A filter was applied to prevent Forest or Savanna Formations converted to anthropic to regenerate as Natural Grassland Formation. Thus, it eliminated part of the confusion of pasture areas and Grassland Formation. This filter is used to ensure that any area converted to pasture stays as pasture for, at least, seven consecutive years to avoid some confusion in the classification.

Additionally, a forest stabilization filter was implemented in the time series from 1985 to 2022. By using the Stable Forest data, we corrected classification confusion between savanna and wetland areas, overestimating the Forest classification, especially in the early years.

Still to avoid false regeneration in agriculture areas, we use the data of agriculture in last year and increment year by year the classification since 1985. This filter considers the field knowledge in the Pantanal (in the southern part of subregion 6).

#### 6.3.5. Trajectory filter

To avoid false transitions between savanna formation and grassland in short time intervals, a filter was applied that considered the trajectory of the pixel according to Pontius (2022). Considering the trajectory of absence - alternation - absence, the number of changes that the class was involved in and also the pixel mode along the time series, the filter stabilizes these two classes.














#### 6.3.6. Spatial filter

A spatial filter was also applied to avoid unwanted modifications to the edges of the pixel groups, using the "connectedPixelCount" function. Native to the GEE platform, this function locates connected pixels (neighbors) with the same 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 6 Landsat pixels (~0,5 ha). This is the final step before integration with cross-cutting themes.

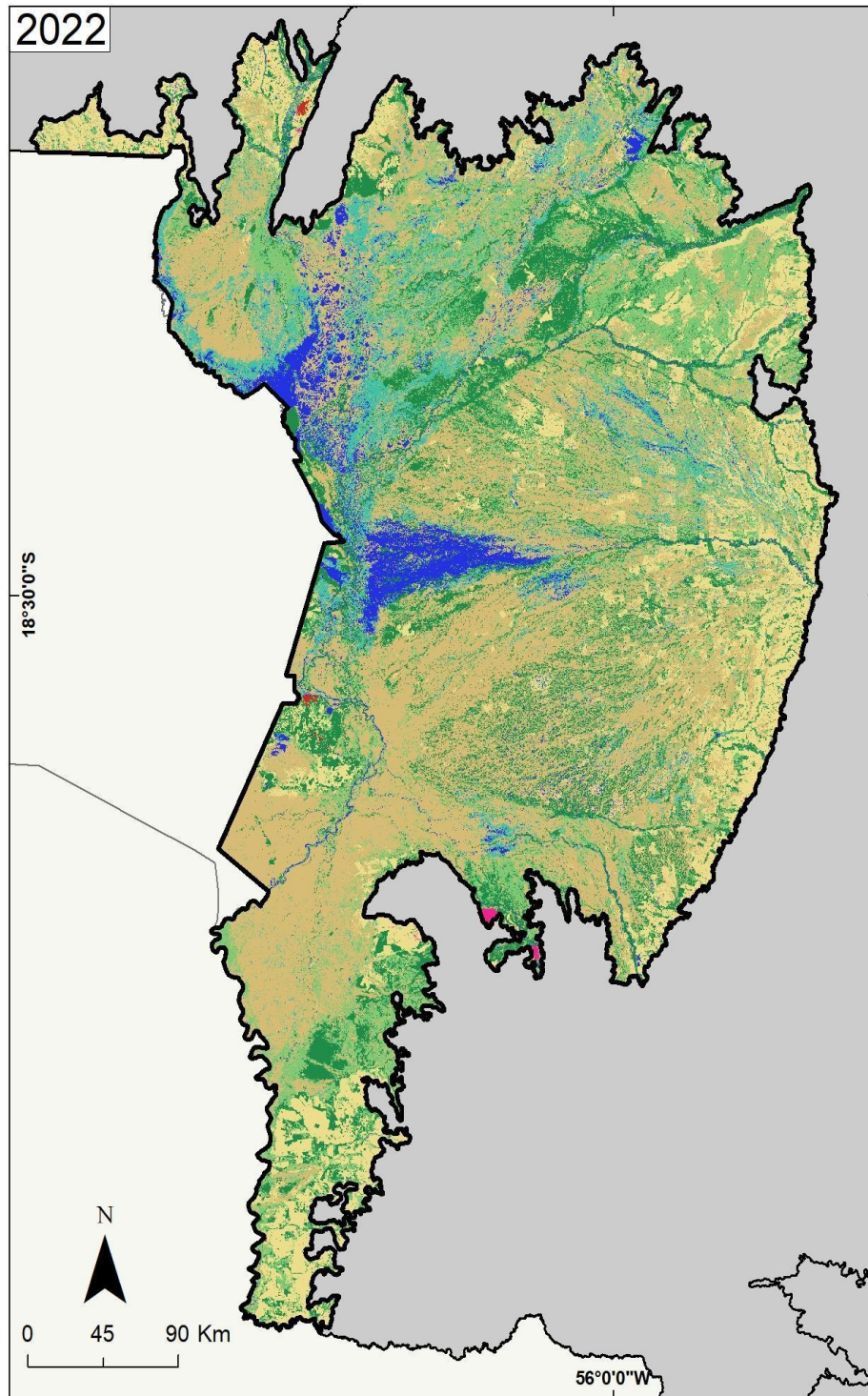
## 7. Integration with cross-cutting themes

The final map generated by the Pantanal biome team was integrated with maps from some cross-cutting themes, which represent rare classes or classes that demand a specific mapping strategy. These external themes comprise urban areas (24), mining (30), agriculture (39, 20, 41) and forest plantation (9). They were superimposed on the Pantanal data (Figure 7), resulting in final maps of the biome containing 13 classes (Table 6).

**Table 6:** Classes observed in the Pantanal biome.

Legend class of Collection 8	Numeric ID	Color
<b>1.1.1.</b> Forest Formation	3	
<b>1.1.2.</b> Savanna Formation	4	
<b>2.1.</b> Wetland	11	
<b>2.2.</b> Grassland Formation	12	
<b>3.1.</b> Pasture	15	
<b>3.2.1.1.</b> Soybean	39	
<b>3.2.1.2.</b> Sugar cane	20	
<b>3.2.1.5.</b> Other Temporary Crops	41	
<b>3.3.</b> Forest Plantation	9	
<b>4.2.</b> Urban area	24	
<b>4.3.</b> Mining	30	
<b>4.4.</b> Other non vegetated area	25	
<b>5.</b> Water	33	



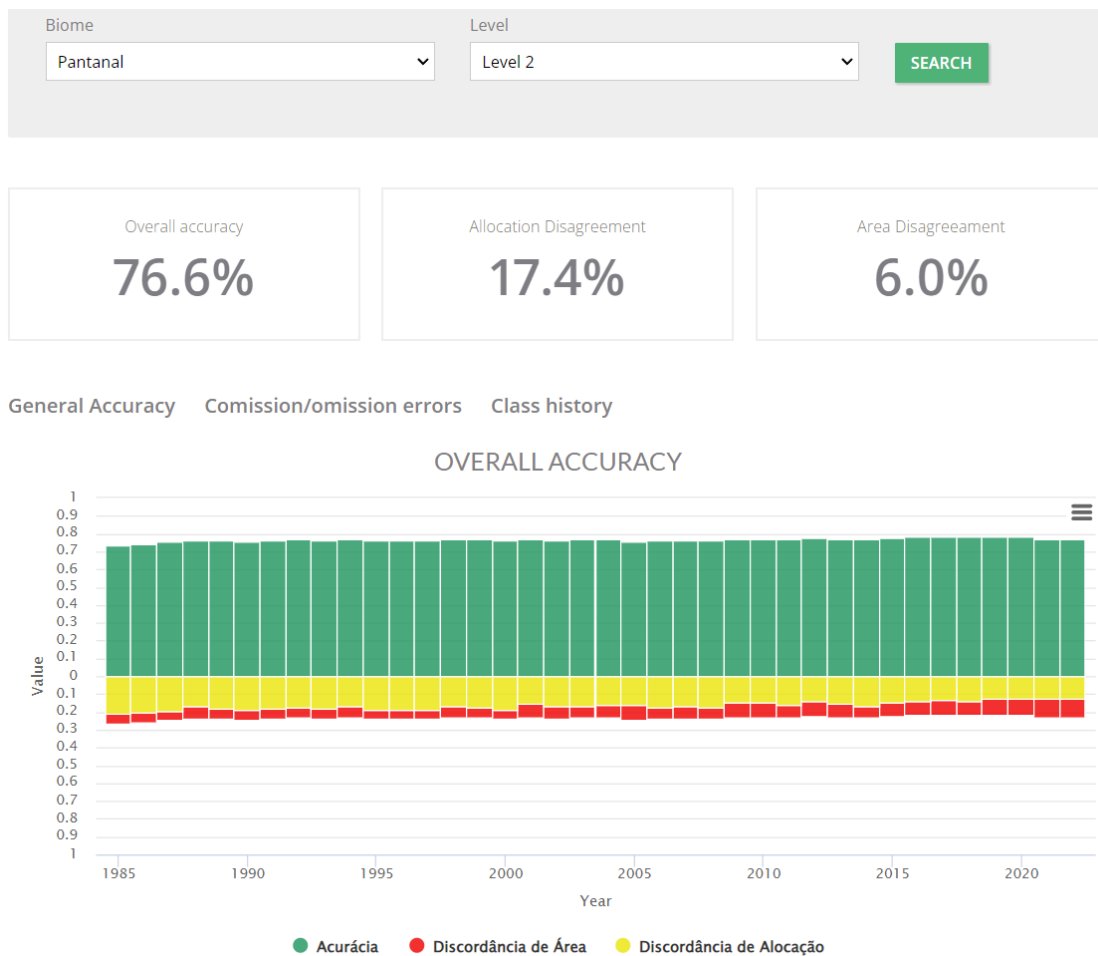


**Figure 7:** Final land use and land cover map of the Pantanal for the year 2022.

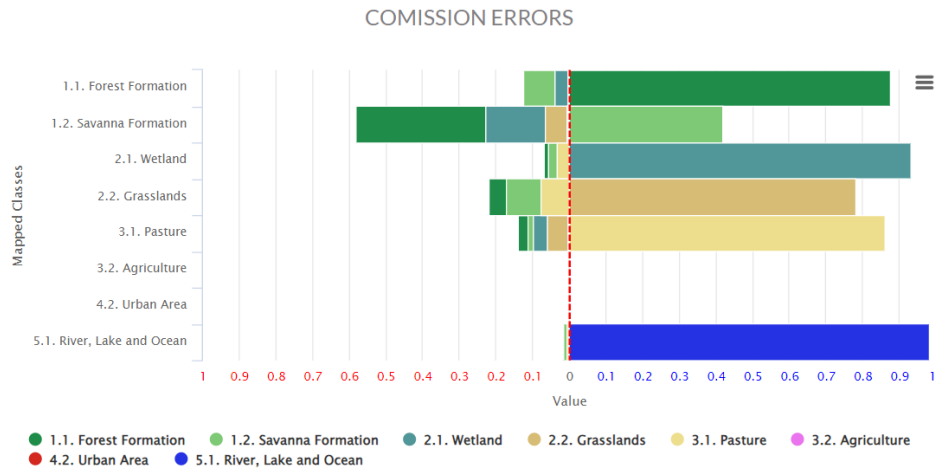
## 8. Validation

Mean overall accuracy (considering all years) in Collection 7 was 85.9% and 79.5% in levels 1 and 2, respectively, and in collection 7.1 it was 85.6%, in level 1 and 77.3% in level 2.

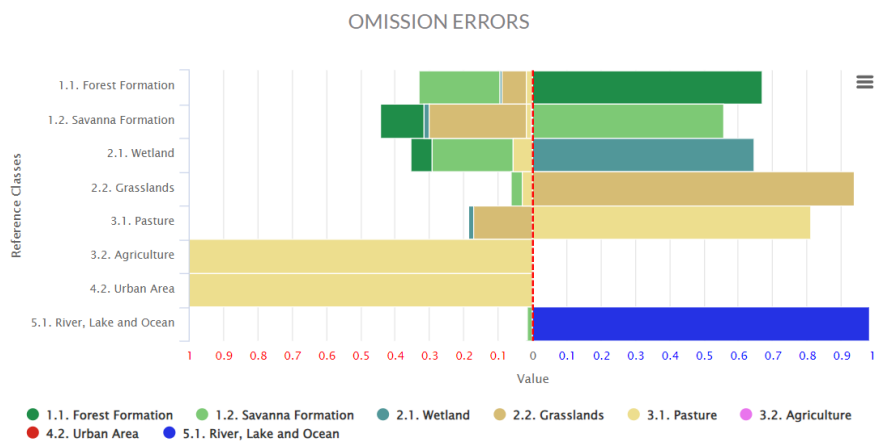
In Collection 8 the overall accuracy reached 84% and 77.6% in levels 1 and 2, respectively (Figure 8). Wetland and Water classes were added to the map by the maximum area in each year, but this criterion is different from the validation methodology. To solve that difference, the validation points of grassland or wetland that overlapped on the classified map as water were considered correct. The detailed information about inclusion and omission error are presented in Figures 9 and 10.



**Figure 8.** Accuracy of level 2 of MapBiomias Collection 8 in the Pantanal biome (1985-2022).

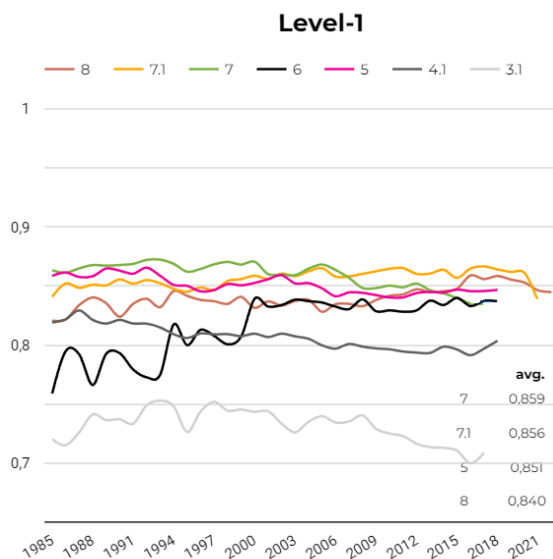


**Figure 9.** Inclusion error in 2022 in Pantanal for each class at level 2.



**Figure 10.** Omission error in 2022 in Pantanal for each class at level 2.

Considering the Global Accuracy for all collections in legend level 1 (Figure 11), the classification of Collection 8 has a steady growth throughout the historical mapping series, mainly in the last years.



**Figure 11.** Global accuracy for all collections in Pantanal over the classification period.

## 9. References

Assine, M. L. et al. The Quaternary alluvial systems tract of the Pantanal Basin, Brazil. **Brazilian Journal Of Geology**, [S. L.], v. 45, n. 3, p. 475-489, set. 2015. DOI: 10.1590/2317-4889201520150014.

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

Funk, C.; Peterson, P.; Landsfeld, M. The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. **Scientific Data 2**, [S. l.], v. 150066, 2015. DOI: 10.1038/sdata.2015.66.

Hamilton, S. K.; Souza, O. C. de; Coutinho, M. E. Dynamics of floodplain inundation in the alluvial fan of the Taquari River (Pantanal, Brazil). **SIL Proceedings**, [S. l.], v. 26, n. 3, p. 1922-2010, 1998. DOI: 10.1080/03680770.1995.11900852.

Padovani, C. R. **Dinâmica das Inundações do Pantanal**. 2010. 174 p. Tese (Doutorado em Ecologia Aplicada) - Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 2010. DOI: 10.11606/T.91.2010.tde-14022011-170515.

Pontius, R. G. **Metrics That Make a Difference**; Springer: Berlin/Heidelberg, Germany, 2022; pp. 1–130.

Potapov, P.; Li, X.; Hernandez-Serna, A.; Tyukavina, A.; Hansen, M. C.; Kommareddy, A.; Pickens, A.; Turubanova, S.; Tang, H.; Silva, C. E.; Armston, J.; Dubayah, R.; Blair, J. B.; Hofton, M. (2020) Mapping and monitoring global forest canopy height through integration of GEDI and Landsat data. **Remote Sensing of Environment**, 112165. <https://doi.org/10.1016/j.rse.2020.112165>.

CI – Conservação Internacional, ECOA - Ecologia e Ação, Fundación AVINA, Instituto SOS Pantanal, WWF-Brasil. **Monitoramento das alterações da cobertura vegetal e uso do solo na Bacia do Alto Paraguai – Porção Brasileira – Período de Análise: 2002 a 2008**. Brasília, 2009. 58 p.; Il.; 23 cm. ISBN 978-85-86440-25-0.

Silva, João & Abdon, Myrian. (1998). Delimitação do Pantanal Brasileiro e suas sub-regiões. **Pesq. Agropec. Bras.** 33.