Pantanal - Appendix

Collection 3

Version 1

General coordinator
Marcos Reis Rosa

Team
Eduardo Reis Rosa
Jacqueline Freitas
Viviane Cristina Mazin

1 Landsat image mosaics

1.1 Definition of the temporal period

The image selection period for the Pantanal biome was defined aiming the selection in the dry season (Figure 1) to reduce the wetlands. 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 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.

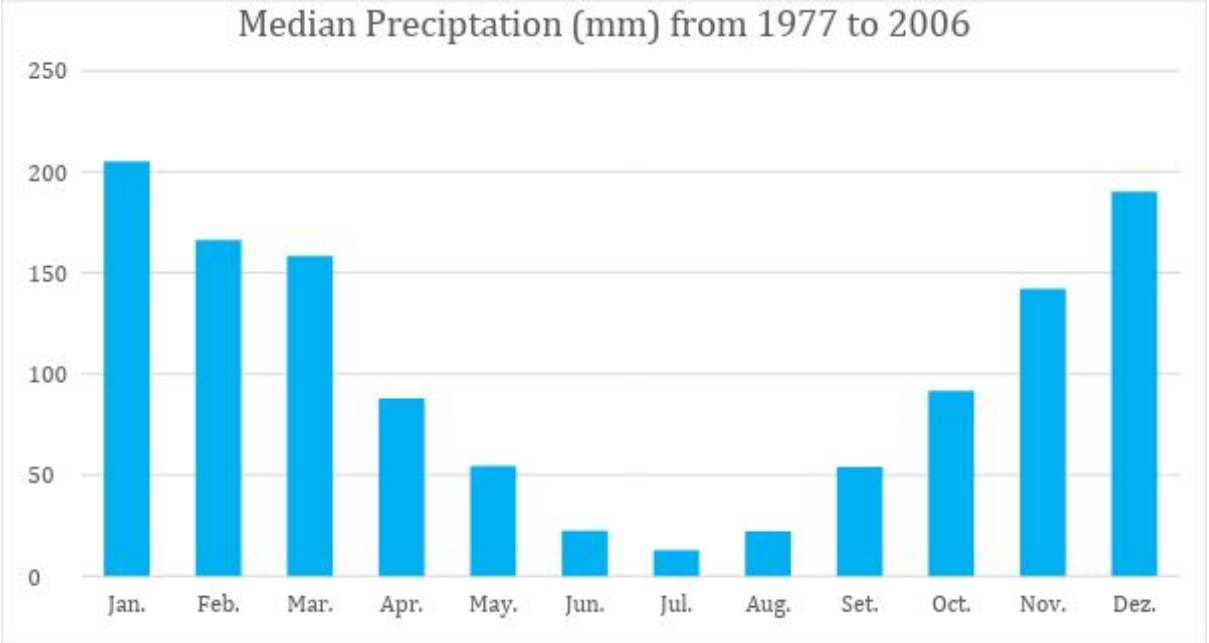

Figure 1. Median monthly precipitation values from 1977 to 2006 (MARCUZZO et al., 2010).

1.2 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 could 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 period from May 1st to August 30th was accepted as a good 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 Classification

2.1 Classification scheme

The digital classification of the Landsat mosaics for the Pantanal biome aimed to individualize a subset of eight land use and land cover (LULC) classes from the MapBiomas Collection 3 legend (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 MapBiomas Collection 3.

Legend class of Collection 3	Numeric ID	Color
1.1.1. Forest Formation	3	
1.1.2. Savanna Formation	4	
2.1. Wetland	11	
2.2. Grassland Formation	12	
3.3 Mosaic of Agriculture and Pasture	21	
4.5. Other non vegetated area	25	
5. Water	26	
6. Non Observed	27	

2.2 Feature space

The feature space for digital classification of the categories of interest for the Pantanal biome comprised a subset of 30 variables (Table 2), taken from the complete feature space of MapBiomas Collection 3. These variables include the original Landsat reflectance bands, as well as vegetation indexes, spectral mixture modeling-derived variables, terrain morphometry (slope), and a spatial texture measure. 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 3 (1985-2017).

ID	Script acronym	Variable	Description
1	'median_blue'	Blue	median Landsat band
2	'median_green'	Green	median Landsat band
3	'median_nir'	NIR	median Landsat band
4	'median_red'	Red	median Landsat band
5	'median_swir1'	SWIR 1	median Landsat band
6	'median_swir2'	SWIR 2	median Landsat band
7	'median_evi2'	Evi 2	Enhanced Vegetation Index 2
8	'median_ndvi'	Ndvi	median Normalized Difference Vegetation Index
9	'median_ndvi_dry'	Ndvi Dry	Normalized Difference Vegetation Index Dry Season
10	'median_ndvi_wet'	Ndvi Wet	Normalized Difference Vegetation Index Wet Season
11	'median_savi'	Savi	median Soil-adjusted Vegetation Index
12	'median_sefi'	Sefi	median Savanna Ecosystem Fraction Index
13	'stdDev_sefi'	Sefi	standard deviation Savanna Ecosystem Fraction Index
14	'median_wefi'	Sefi	median Wetlands Ecosystem Fraction Index
15	'median_ndwi'	Ndwi	median Normalized Difference Water Index
16	'median_ndwi_dry'	Ndwi Dry	Normalized Difference Water Index Dry Season
17	'median_ndwi_wet'	Ndwi Wet	Normalized Difference Water Index Wet Season
18	'median_gv'	Gv	median green vegetation fraction
19	'median_gvs'	Gvs	median GV / (100 - shade)
20	'median_ndfi'	Ndfi	Median Normalized Difference Fraction Index
21	'stdDev_ndfi'	Ndfi	standard deviation Normalized Difference Fraction Index
22	'median_ndfi_dry'	Ndfi Dry	Normalized Difference Fraction Index Dry Season
23	'median_ndfi_wet'	Ndfi Wet	Normalized Difference Fraction Index Wet Season
24	'median_npv'	Npv	Median npv fraction
25	'stdDev_npv'	Npv	standard deviation npv fraction
26	'median_shade'	Shade	median shade fraction
27	'median_soil'	Soil	median soil fraction
28	'stdDev_soil'	Soil	soil fraction
29	'textG'	Green spatial texture	Spatial texture
30	'slope'	Slope	Slope

2.3 Classification algorithm, training samples and parameters

Digital classification was performed chart by chart, year by year, using a Random Forest algorithm (Breiman, 2001) available in Google Earth Engine. Training samples for each chart were defined following a strategy of using pixels for which the LULC remained the same along the 33 years of Collection 3, so named "stable samples". An ensemble taken from three main sources was made: extracted from Collection 2.3; manually drawn polygons; and complementary samples.

2.3.1 Stable samples from Collection 2.3

The extraction of stable samples from the previous Collection 2.3 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 remained with that class to be eligible as a stable sample. A layer of pixels with a stable classification along the 17 years of Collection 2.3 was then generated by applying such thresholds. Following, a set of polygons delineating zones with errors in some class (*e.g.*, omission or commission) were drawn and used as a mask to delete misclassified pixels. A second cleaning was performed by comparing the stable pixels with a reference map from Instituto SOS Pantanal excluding all pixels whose class disagreed with it. From the resulting layer of stable samples, a subset of 10,000 pixels was randomly selected and used as training areas to classify all charts for each of the 33 years with the Random Forest algorithm, by running 50 iterations.

After classification, a temporal filter (see item 3.1) was applied to each chart in order to improve the classification consistency of each pixel along the period 1985-2017. The output of the temporal filter was then submitted to the same procedures described above: definition and application of a threshold for the selection of stable pixels along the 33 years, followed by the exclusion of misclassified pixels by drawing mask polygons, and by comparison with a reference. The product is the 1st version of the map of stable classes from 1985 to 2017.

2.3.2 Manually drawn polygons

Manually drawn polygons were used to add samples for classes with little occurrence, as well as to help to enrich class representation in zones which presented classification problems in 1st version of the map of stable classes from 1985 to 2017. Drawing was performed in Google Earth Engine Code Editor using Landsat mosaics as backdrop. Again, the concept of stable samples was applied: each of the polygons should delineate areas in which LULC remained unchanged, checking the mosaics for all the 33 years.

2.3.3 Preliminary classification

From both the sets of stable samples, a subset of 3,000 pixels was randomly selected and used as training areas to classify all charts for each of the 33 years with the Random Forest algorithm, now running 100 iterations. A number of 2,000 training pixels was selected from the 1st version of the map of stable classes from 1985 to 2017, while 1,000 were selected from the manually drawn polygons. The output of

the temporal filter was then submitted to the same procedures described above: definition and application of a threshold for the selection of stable pixels along the 33 years, followed by the exclusion of misclassified pixels by drawing mask polygons, and by comparison with a reference. The product is the 2st and final version of the map of stable classes from 1985 to 2017.

2.3.4 Complementary samples

The need for complementary samples was evaluated by visual inspection and by comparing the output of the preliminary classification for the year 2011 with a reference map of the same year. When adjacent charts presented class discontinuities in some year, or some classes and/or some portion of a chart presented a certain degree of disagreement with the reference of 2009, new samples were added aiming to improve classification for such specific cases. Complementary sample collection was also done by means of drawing polygons but 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 33 years during the polygon drawing. Based in the knowledge of each regions samples from savanna, grassland or Rocky outcrop were added. Samples from forests that were not well represented in the stable map were also added where need.

2.3.5 Samples to balance training dataset

Samples from classes that had lower distribution were added to balance the training dataset. Samples from Forest were added in some regions to balance when Mosaic of Agriculture and Pasture was predominant. Samples of Water and Non Vegetated Area were added to all chars/years.

2.3.6 Final classification

Final classification was performed for all charts/years and each chart used 4.000 samples from the map of stable classes, 2,000 samples from each class of complementary samples when available and 2,000 of samples collected to balance the training dataset. All years used the same subset of samples and it was trained in the same mosaic of the year that was classified.

3 Pos-classification

3.1 Temporal filter

The temporal filter rules were adapted for the LULC classes in the Pantanal biome and were complemented by specific rules to adjust for cases where a pixel appeared two subsequent years in the class "Non Observed". A number of 81 rules, distributed in three groups, were used: a) rules for cases not observed in the first year (RP); (b) rules for cases not observed in the final year (RU); (c) rules for cases of implausible transitions or not observed for intermediate years (Table 3).

Table 3. General and specific rules of the temporal filter for the Pantanal biome in the MapBiomas Collection 3. RG = General Rule, RP = First Year Rule, RU = Last Year Rule, FF = Forest Formation, SV = Savanna Formation, AU = Wetland, FC = Grassland, AG = Mosaic of Agriculture and Pasture, AR = Rocky Outcrop, NV = Other Non Vegetated Area ,CD = Water Bodies, NO = Non Observed.

rule	type	kernel	biome	notes	tminus2	tminus1	t	tplus1	tplus2	result
RG01	RG	3	PANTANAL	Correct FF		3	12	3		3
RG02	RG	3	PANTANAL	Correct FF		3	21	3		3
RG03	RG	3	PANTANAL	Correct FF		3	27	3		3
RG04	RG	3	PANTANAL	Correct FF		3	4	3		3
RG05	RG	3	PANTANAL	Correct FF		3	11	3		3
RG06	RG	3	PANTANAL	Correct FF		3	26	3		3
RG07	RG	3	PANTANAL	Correct SV		4	12	4		4
RG08	RG	3	PANTANAL	Correct SV		4	21	4		4
RG09	RG	3	PANTANAL	Correct SV		4	27	4		4
RG10	RG	3	PANTANAL	Correct SV		4	3	4		4
RG11	RG	3	PANTANAL	Correct SV		4	11	4		4
RG12	RG	3	PANTANAL	Correct SV		4	26	4		4
RG13	RG	3	PANTANAL	Correct FC		12	3	12		12
RG14	RG	3	PANTANAL	Correct FC		12	4	12		12
RG15	RG	3	PANTANAL	Correct FC		12	21	12		12
RG16	RG	3	PANTANAL	Correct FC		12	27	12		12
RG17	RG	3	PANTANAL	Correct FC		12	11	12		12
RG18	RG	3	PANTANAL	Correct FC		12	26	12		12
RG19	RG	3	PANTANAL	Correct AG		21	3	21		21
RG20	RG	3	PANTANAL	Correct AG		21	4	21		21
RG21	RG	3	PANTANAL	Correct AG		21	12	21		21
RG22	RG	3	PANTANAL	Correct AG		21	27	21		21

RG23	RG	3	PANTANAL	Correct AG		21	11	21		21
RG24	RG	3	PANTANAL	Correct AG		21	26	21		21
RG25	RG	3	PANTANAL	Correct AU		11	4	11		11
RG26	RG	3	PANTANAL	Correct AU		11	12	11		11
RG27	RG	3	PANTANAL	Correct AU		11	3	11		11
RG28	RG	3	PANTANAL	Correct AU		11	21	11		11
RG29	RG	3	PANTANAL	Correct AU		11	26	11		11
RG30	RG	3	PANTANAL	Correct AU		11	27	11		11
RG31	RG	3	PANTANAL	Correct CD		26	4	26		26
RG32	RG	3	PANTANAL	Correct CD		26	12	26		26
RG33	RG	3	PANTANAL	Correct CD		26	11	26		26
RG34	RG	3	PANTANAL	Correct CD		26	21	26		26
RG35	RG	3	PANTANAL	Correct CD		26	3	26		26
RG36	RG	3	PANTANAL	Correct CD		26	27	26		26

rule	type	kernel	biome	notes	tminus2	tminus1	t	tplus1	tplus2	result
RG37	RG	5	PANTANAL	Correct Deforestation	3	3	12	12	12	21
RG38	RG	5	PANTANAL	Correct Deforestation	3	3	12	12	21	21
RG39	RG	5	PANTANAL	Correct Deforestation	3	3	12	21	21	21
RG40	RG	5	PANTANAL	Correct Deforestation	4	4	12	12	12	21
RG41	RG	5	PANTANAL	Correct Deforestation	4	4	12	12	21	21
RG42	RG	5	PANTANAL	Correct Deforestation	4	4	12	21	21	21
RG43	RG	5	PANTANAL	Correct false Regen	21	21	3	3	21	21
RG44	RG	5	PANTANAL	Correct false Regen	21	21	4	4	21	21
RG45	RG	5	PANTANAL	Correct false Regen	21	21	12	12	21	21
RG46	RG	5	PANTANAL	Correct false Regen	3	3	4	4	3	3
RG47	RG	5	PANTANAL	Correct false Regen	4	4	3	3	4	4
RG48	RG	5	PANTANAL	Correct false Regen	11	11	3	3	11	12
RG49	RG	5	PANTANAL	Correct false Regen	11	11	3	3	12	12
RP01	RP	3	PANTANAL	Fill NO		27	3	3		3
RP02	RP	3	PANTANAL	Fill NO		27	4	4		4
RP03	RP	3	PANTANAL	Fill NO		27	12	13		13
RP04	RP	3	PANTANAL	Fill NO		27	11	11		11
RP05	RP	3	PANTANAL	Fill NO		27	21	21		21
RP06	RP	3	PANTANAL	Fill NO		27	26	26		26
RP07	RP	3	PANTANAL	Fill NO		27	3	4		3
RP08	RP	3	PANTANAL	Fill NO		27	4	3		4
RU01	RU	3	PANTANAL	Fill NO		3	3	27		3
RU02	RU	3	PANTANAL	Fill NO		4	4	27		4
RU03	RU	3	PANTANAL	Fill NO		12	12	27		12
RU04	RU	3	PANTANAL	Fill NO		11	11	27		11

RU05	RU	3	PANTANAL	Fill NO		21	21	27		21
RU06	RU	3	PANTANAL	Fill NO		26	26	27		26

3.2 Integration with cross-cutting themes

The products of digital classification after the application of the temporal filter, for each of the 33 years in the period 1985-2017, were then integrated with the cross-cutting themes using one specific rule: in the final product the class “Mosaic of Agriculture and Pasture” was reclassified as “Pasture”.

4 Validation strategies

4.1 Use of reference maps

The LULC map used as reference was the “Monitoramento da Bacia do Alto Paraguai” produced by “Instituto SOS Pantanal”. They produced LULC maps using visual interpretation of Landsat images in 2002, 2008, 2010, 2012, 2014, 2016 and 2017.The comparison of the maps will be included in the next version of ATBD.

5 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.