Cerrado - Appendix

Collection 3

Version 1

General coordinator
Ane A. Alencar

Team
Camila Balzani Marques
Felipe E. B. Lenti
Isa Costa Alencar
Isabel de Castro Silva
João Ribeiro
Julia Zanin Shimbo
Ricardo Souza
Valderli J. Piontekowski
Vera L. S. Arruda
Vivian Ribeiro

1. Landsat image mosaics

The first step to classify the native vegetation (Forest, Savanna and Grassland) of Cerrado was to generate the mosaic of images that were used in the classification scheme. The mosaic of images represents a mosaic composition of the best pixels that are extracted from all the images available between a defined period within a year. Once the initial and final dates of this period were defined, the median of these pixels was calculated generating one value for each pixel. The aggregation of these composed pixels generated then the annual mosaic that was submitted to classification.

Several tests were done to define the optimum period of images to compose the annual mosaics. Due to the high impact of the seasonality on Cerrado vegetation spectral response it was evaluated compositions of images from the raining and dry seasons. The tests included classification using the end of the raining season when the Cerrado vegetation is still vigorous and there is higher probability of getting images with lower cloud observation if compared with the peak of the raining season. Tests were done also with a composition of images generated in the end of the dry season which includes the months between July and September. The tests demonstrated that if only images from the raining season were used, the result was a greener mosaic and the chances of increasing the commission errors in the classification were higher, since more areas that were not forest (*i.e.* dense savannas) can be classified as forest. On the other hand, if only the images acquired during the last three months of the raining season were selected, the mosaic results in a drier aspect, underestimate the forest cover mainly due to the lost ability for mapping deciduous forests (Figure 1).

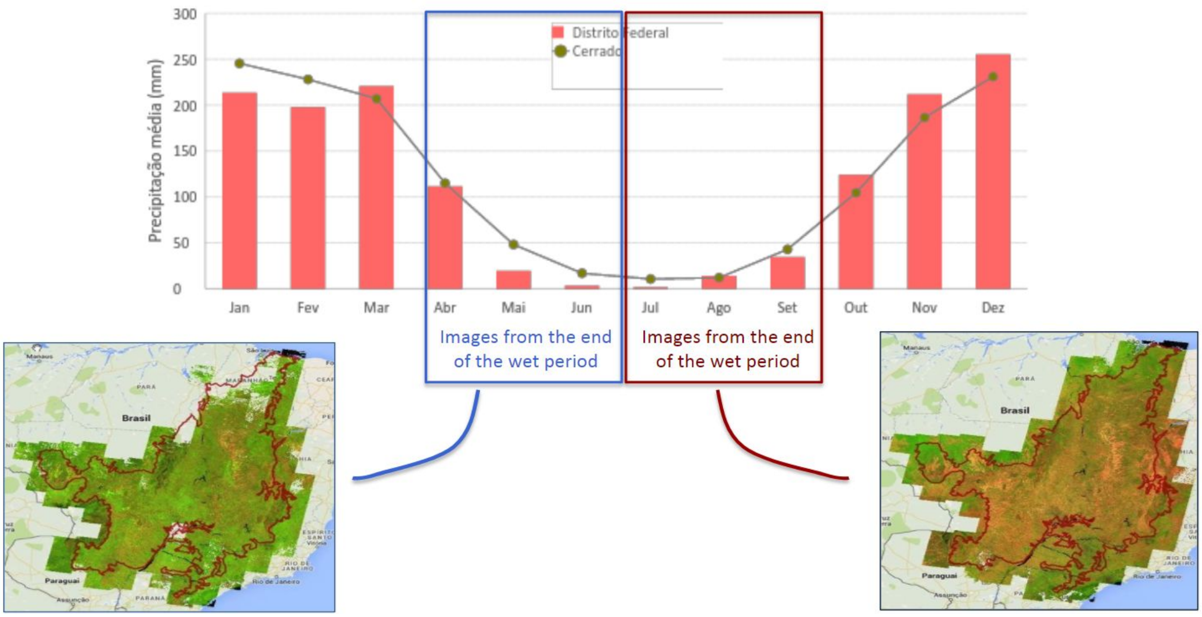


Figure 1. Pixel composite mosaics in the end of the raining season and the end of the dry season in the Cerrado biome.

Based on the tests described above, a large window was decided for the selection of the initial and final dates for the mosaics. These dates were individually selected for each of the 173 tiles and for each year. The criteria for the select these dates included the use of a maximum six months window between the months of April to September (Figure 2). The median value of the pixels selected during this wider period demonstrated to accommodate better the difficulties in mapping forest presented in the shorter window tests. In fact, this strategy averaged the commission and omission errors of the strategies presented above generating 33 mosaics (Figure 3).

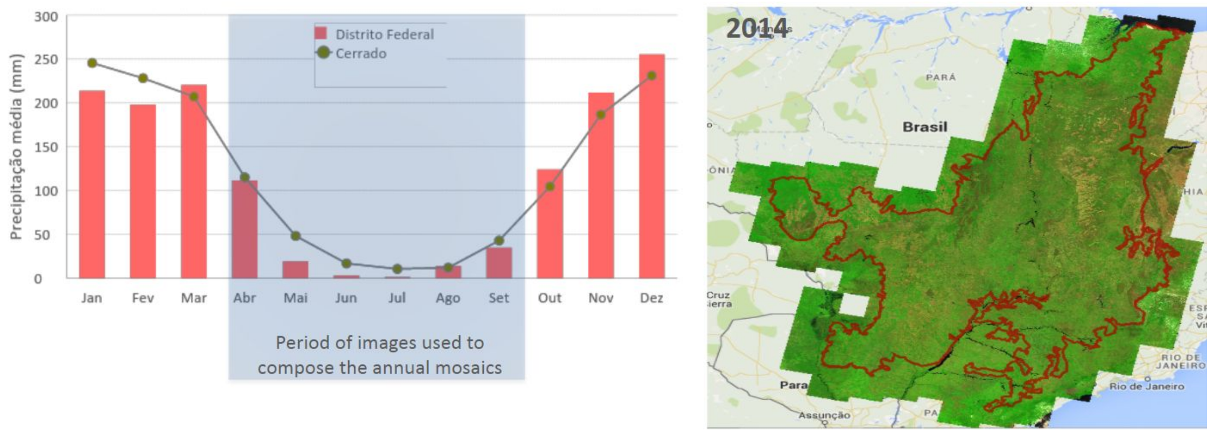


Figure 2. Window period used to define the final pixel composite annual mosaics used in the classification of MapBiomas Collection 3 in the Cerrado biome.

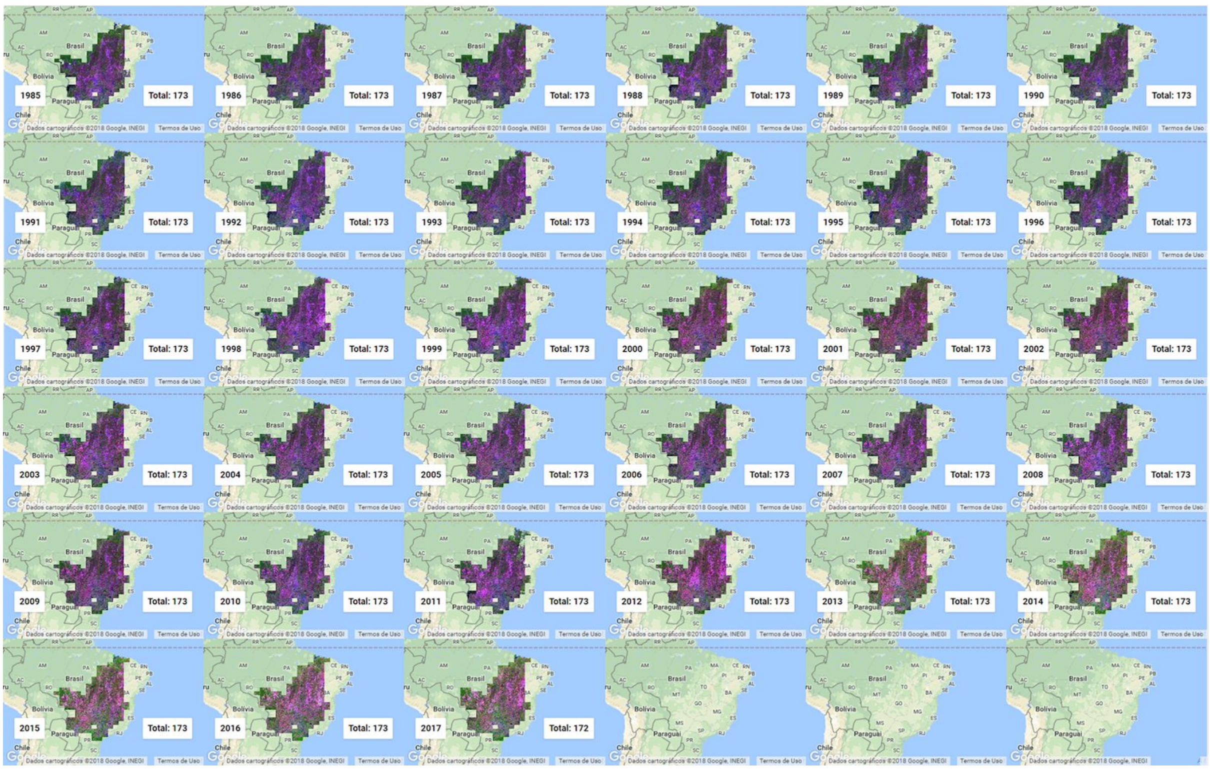


Figure 3. Annual Landsat image mosaics of the Cerrado biome from 1985 to 2017 in the MapBiomas Collection 3.

2. Classification

The Collection 3 was improved with the Random Forest approach by collecting training samples both manually (*i.e.* vectorization based on visual interpretation) and automatically (*i.e.* random points within the areas of stable classes of the Collection 2.3). The manual approach was necessary to ensure sufficient samples in classes that covered a small area in the stable classes map.

2.1. Classification scheme

The classification of the Landsat mosaics for the Cerrado biome considered seven land use and land cover (LULC) classes of the MapBiomas Collection 3 Legend (Table 1), which later was integrated with the cross-cutting themes classes in a further step.

Table 1. Land cover and land use categories considered for digital classification of Landsat mosaics for the Cerrado 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	
2.2. Grassland Formation	12	
3.3 Mosaic of Agriculture and Pasture	21	
4.4. Rocky Outcrop	29	
5. Water	26	
6. Non Observed	27	

The development of the Collection 3 with annual maps of Cerrado main natural vegetation formations from 1985 to 2017 followed the steps:

(1) definition of areas of stable classes considering the Collection 2.3 time series (2000 - 2016) and thresholds in terms of years of pixel persistence in a given class;

(2) definition of training samples by sorting 12,500 random points per tile inside the areas of stable classes (considering Collection 2.3) and extracting remote sensing variables from the mosaics generated for Collection 3;

(3) running the Random Forest classifier at the Google Earth Engine Platform to generate a preliminary series of LULC maps from 1985 to 2017;

(4) definition of areas of stable classes considering the preliminary time series (1985 - 2017) and thresholds in terms of years of pixel persistence in a given class;

(5) definition of training samples by sorting 12,500 random points per tile inside the areas of stable classes (the preliminary time series from 1985 and 2017) and extracting remote sensing metrics from the mosaics generated for Collection 3;

(6) manual collection of complementary training samples using the WebCollect platform to cover unstable areas in order to improve sample availability for infrequent classes and to enhance spatial consistency in regions problematic in this regard. Polygons were drawn and sample points were collected inside of them.

(7) running the Random Forest classifier at the Google Earth Engine Platform to generate the definitive series of LULC maps from 1985 to 2017.

2.2. Feature space

The feature space used in the Cerrado biome in the Collection 3 was defined empirically by running several pilot classifications in tiles for which there were comparable reference maps, such as the TerraClass Cerrado (2013), the Third National Inventory of Anthropic Emissions (2010) and the Brazilian Foundation for Sustainable Development - FBDS (2012). The pilot results were also evaluated in a preliminary accuracy assessment based on approximately 7,500 sample points in the Cerrado, collected by LAPIG team. The feature space with highest accuracy and spatial/temporal consistency in these pilot classifications was then generalized to all tiles.

Table 2. Feature space subset considered in the classification of the Pampa biome Landsat image mosaics in the MapBiomas Collection 3.

Variable	Description	Statistic	Temporal range	Script acronym
Blue	Landsat band	median	year	median_blue
Blue dry season	Landsat band	median	seasonal ; NDVI below first quartile	median_blue_dry
Blue wet season	Landsat band	median	seasonal ; NDVI above first quartile	median_blue_wet
Green	Landsat band	median	year	median_green
Greendr dry season	Landsat band	median	seasonal ; NDVI below first quartile	median_green_dry

Green wet season	Landsat band	median	seasonal ; NDVI above first quartile	median_green_wet
Red	Landsat band	median	year	median_red
Red dry season	Landsat band	median	seasonal ; NDVI below first quartile	median_red_dry
Red wet season	Landsat band	median	seasonal ; NDVI above first quartile	median_red_wet
Near Infrared (NIR)	Landsat band	median	year	median_nir
Near Infrared (NIR) dry season	Landsat band	median	seasonal ; NDVI below first quartile	median_nir_dry
Near Infrared (NIR) wet season	Landsat band	median	seasonal ; NDVI above first quartile	median_nir_wet
Shortwave Infrared 1 (SWIR 1)	Landsat band	median	year	median_swir1
Shortwave Infrared 1 (SWIR 1) dry season	Landsat band	median	seasonal ; NDVI below first quartile	median_swir1_dry
Shortwave Infrared 1 (SWIR 1) wet season	Landsat band	median	seasonal ; NDVI above first quartile	median_swir1_wet
Shortwave Infrared 2 (SWIR 2)	Landsat band	median	year	median_swir2
Shortwave Infrared 2 (SWIR 2) dry season	Landsat band	median	seasonal ; NDVI below first quartile	median_swir2_dry
Shortwave Infrared 2 (SWIR 2) wet season	Landsat band	median	seasonal ; NDVI above first quartile	median_swir2_wet

EVI2	Enhanced vegetation index 2	median	year	median_evi2
EVI2 dry season	Enhanced vegetation index 2	median	seasonal ; NDVI below first quartile	median_evi2_dry
EVI2 wet season	Enhanced vegetation index 2	median	seasonal ; NDVI above first quartile	median_evi2_wet
GV	Green vegetation fraction	median	year	median_gv
GVS	GV / (100 - shade)	median	year	median_gvs
GVS dry season	GV / (100 - shade)	median	seasonal ; NDVI below first quartile	median_gvs_dry
GVS wet season	GV / (100 - shade)	median	seasonal ; NDVI above first quartile	median_gvs_wet
NPV	Non-photossintetic vegetation fraction	median	year	median_npv
Soil	Soil fraction	median	year	median_soil
Shade	Shade fraction	median	year	median_shade
NDFI	Normalized Difference Fraction Index	median	year	median_ndfi
NDFI dry season	Normalized Difference Fraction Index	median	seasonal ; NDVI below first quartile	median_ndfi_dry
NDFI wet season	Normalized Difference Fraction Index	median	seasonal ; NDVI above first quartile	median_ndfi_wet
NDVI	Normalized Difference Vegetation Index	median	year	median_ndvi
NDVI dry season	Normalized Difference Vegetation Index	median	seasonal ; NDVI below first quartile	median_ndvi_dry
NDVI wet season	Normalized Difference Vegetation Index	median	seasonal ; NDVI above first quartile	median_ndvi_wet
NDWI	Normalized Difference	median	year	median_ndwi

	Water Index			
NDWI dry season	Normalized Difference Water Index	median	seasonal ; NDVI below first quartile	median_ndwi_dry
NDWI wet season	Normalized Difference Water Index	median	seasonal ; NDVI above first quartile	median_ndwi_wet
SAVI	Soil-adjusted vegetation index	median	year	median_savi
SAVI dry season	Soil-adjusted vegetation index	median	seasonal ; NDVI below first quartile	median_savi_dry
SAVI wet season	Soil-adjusted vegetation index	median	seasonal ; NDVI above first quartile	median_savi_wet
SEFI	Savanna ecosystem fraction index	median	year	median_sefi
SEFI dry season	Savanna ecosystem fraction index	median	seasonal ; NDVI below first quartile	median_sefi_dry
SEFI wet season	Savanna ecosystem fraction index	median	seasonal ; NDVI above first quartile	median_sefi_wet
delta SEFI*	Temporal difference in SEFI	identity	year	dif_sefi
SEFI entropy*	Spatial texture	identity	year	entr_sefi
Slope*	Terrain slope	identity	fixed	slope

2.3. Classification algorithm, training samples and parameters

For the final classification a Random Forest classifier was trained for each tile using approximately 18,000 sample points. Samples were randomly selected from the internal area of the tile and from a 50 km-wide buffering area around its border (*i.e.* from adjacent tiles). The usage of samples from neighboring tiles was adopted as it has improved the spatial consistency of the maps resulting from preliminary tests. The ratio between manually collected samples and automatically collected samples was about even, but the exact number of samples varied tile-wise as there could be tiles with no stable area of a given class, impeding the use of automatically collected samples for that class in that tile. The minimum amount of sample points for any given class was set to 200, while the maximum was set to 6,000; all tiles were classified using 120 decision trees.

3. Post-classification

3.1. Complementary samples (in steep terrains and water bodies)

After classification was finished we observed confusion between cast shadows and water bodies in Cerrado regions with steep terrain associated with mountains (nine tiles). To address this issue we collected complimentary samples of shade pixels and trained a Random Forest classifier using only water samples, shade samples and a reduced feature space, focused on variables that could help differentiate between this two targets. We then generated a layer containing only pixels that were originally mapped as water (water mask) and ran this supplementary classification to assign each pixel as "confirmed" water or shade pixel, in which case it was remapped to the "Not observed" class. Finally, pixels originally classified as water in each of the nine tiles were substituted by the layer generated in this post-classification step.

3.2. Temporal filter

We defined 94 temporal filter rules considering both a three years kernel and a five years kernel (Table 2). This rules were applied in order to eliminate the presence of misclassified pixels that generated impossible transitions throughout the time series. Our main focus was to avoid false detection of vegetation loss and regeneration. For example, a pixel mapped as natural vegetation could only change to agriculture if it persisted as the later class for three or more years; if this criterium was not satisfied, then the pixel was reassigned to the natural vegetation class. The same idea was applied regarding regenerating vegetation, so that a transition from agriculture to forest formation, for example, was considered only if the pixels persisted as forest for three or more years after the transition.

Table 2. General and specific rules of the temporal filter for the Cerrado biome in the MapBiomas Collection 3. RG = General Rule, RP = First Year Rule, RU = Last Year Rule, FF = Forest Formation, SF = Savanna Formation; GL = Grassland, AG = Mosaic of Agriculture and Pasture, AR = Rocky Outcrop, WB = Water, NO = Non Observed.

Rule	Type	Kernel	Notes	Tminus2	Tminus1	T	Tplus1	Tplus2	Result
R1	RG	3	FF-NO-FF	null	3	27	3	null	3
R2	RG	5	FF-FF-AG-FF-FF	3	3	21	3	3	3
R3	RG	5	FF-FF-GL-FF-FF	3	3	12	3	3	3

R4	RG	5	FF-FF-SF-FF-FF	3	3	4	3	3	3
R5	RG	5	FF-FF-AG-AG-FF	3	3	21	21	3	3
R6	RG	3	FF-GL-FF	null	3	12	3	null	3
R7	RG	3	SF-NO-SF	null	4	27	4	null	4
R8	RG	3	SF-NO-AG	null	4	27	21	null	4
R9	RG	3	SF-NO-FF	null	4	27	3	null	4
R10	RG	3	SF-NO-GL	null	4	27	12	null	4
R11	RG	3	SF-NO-NO	null	4	27	27	null	4
R12	RG	3	SF-NO-RO	null	4	27	29	null	4
R13	RG	3	SF-NO-WB	null	4	27	26	null	4
R14	RG	3	FF-NO-AG	null	3	27	21	null	3
R15	RU	3	SF-SF-NO	null	4	4	27	null	4
R16	RG	5	SF-SF-AG-SF-SF	4	4	21	4	4	4
R17	RG	5	SF-SF-FF-SF-SF	4	4	3	4	4	4
R18	RG	5	SF-SF-GL-SF-SF	4	4	12	4	4	4
R19	RG	5	SF-SF-RO-SF-SF	4	4	29	4	4	4
R20	RG	3	SF-FF-SF	null	4	3	4	null	4
R21	RG	3	SF-GL-SF	null	4	12	4	null	4
R23	RG	3	GL-NO-GL	null	12	27	12	null	12
R24	RG	3	GL-FF-GL	null	12	3	12	null	12
R25	RG	3	GL-SF-GL	null	12	4	12	null	12
R26	RG	3	GL-NO-AG	null	12	27	21	null	10
R27	RG	3	GL-NO-FF	null	12	27	3	null	10

R28	RG	3	FF-NO-GL	null	3	27	12	null	3
R29	RG	3	GL-NO-NO	null	12	27	27	null	10
R30	RG	3	GL-NO-RO	null	12	27	29	null	10
R31	RG	3	GL-NO-SF	null	12	27	4	null	10
R32	RG	3	GL-NO-WB	null	12	27	26	null	10
R33	RU	3	GL-GL-NO	null	12	12	27	null	10
R34	RG	5	GL-GL-AG-GL-GL	12	12	21	12	12	12
R35	RG	5	GL-GL-AG-AG-GL	12	12	21	21	12	12
R36	RG	5	GL-GL-FF-FF-GL	12	12	3	3	12	12
R37	RG	5	GL-GL-SF-SF-GL	12	12	4	4	12	12
R38	RG	5	GL-GL-RO-GL-GL	12	12	29	12	12	12
R39	RG	5	GL-GL-SF-GL-GL	12	12	4	12	12	12
R40	RG	5	GL-GL-FF-GL-GL	12	12	3	12	12	12
R41	RG	5	GL-GL-FF-AG-AG	12	12	3	21	21	21
R42	RG	3	FF-NO-NO	null	3	27	27	null	3
R43	RG	3	AG-NO-AG	null	21	27	21	null	21
R44	RG	3	AG-NO-FF	null	21	27	3	null	21
R45	RG	3	AG-NO-GL	null	21	27	12	null	21
R46	RG	3	AG-NO-NO	null	21	27	27	null	21
R47	RG	3	AG-NO-RO	null	21	27	29	null	21
R48	RG	3	AG-NO-SF	null	21	27	4	null	21
R49	RG	3	AG-NO-WB	null	21	27	26	null	21
R50	RU	3	AG-AG-NO	null	21	21	27	null	21

R51	RG	5	AG-AG-FF-GL-GL	21	21	3	12	12	21
R52	RG	3	FF-NO-RO	null	3	27	29	null	3
R53	RG	5	AG-AG-RO-AG-AG	21	21	29	21	21	21
R54	RG	5	AG-AG-SF-AG-AG	21	21	4	21	21	21
R55	RG	5	AG-AG-FF-AG-AG	21	21	3	21	21	21
R56	RG	3	RO-NO-RO	null	29	27	29	null	29
R57	RG	3	RO-NO-AG	null	29	27	21	null	29
R58	RG	3	RO-NO-FF	null	29	27	3	null	29
R59	RG	3	RO-NO-GL	null	29	27	12	null	29
R60	RG	3	RO-NO-NO	null	29	27	27	null	29
R61	RG	3	RO-NO-SF	null	29	27	4	null	29
R62	RG	3	RO-NO-WB	null	29	27	26	null	29
R63	RG	3	FF-SF-FF	null	3	4	3	null	3
R64	RG	3	FF-GL-FF	null	3	12	3	null	3
R65	RU	3	RO-RO-NO	null	29	29	27	null	29
R66	RG	5	RO-RO-SF-RO-RO	29	29	4	29	29	29
R67	RG	5	RO-RO-FF-RO-RO	29	29	3	29	29	29
R68	RG	5	RO-RO-FF-RO-RO	29	29	3	29	29	29
R69	RG	3	WB-NO-WB	null	26	27	26	null	26
R70	RG	3	WB-NO-AG	null	26	27	21	null	26
R71	RG	3	WB-NO-FF	null	26	27	3	null	26
R72	RG	3	WB-NO-GL	null	26	27	12	null	26
R73	RG	3	WB-NO-NO	null	26	27	27	null	26

R74	RG	3	WB-NO-RO	null	26	27	29	null	26
R75	RG	3	FF-NO-WB	null	3	27	26	null	3
R76	RG	3	WB-NO-SF	null	26	27	4	null	26
R77	RU	3	WB-WB-NO	null	26	26	27	null	26
R78	RP	3	NO-FF-FF	null	27	3	3	null	3
R79	RP	3	NO-SF-SF	null	27	4	4	null	4
R81	RP	3	NO-GL-GL	null	27	12	12	null	10
R82	RP	3	NO-AG-AG	null	27	21	21	null	21
R83	RP	3	NO-WB-WB	null	27	26	26	null	26
R84	RP	3	NO-RO-RO	null	27	29	29	null	29
R85	RG	5	AG-AG-FF-FF-AG	21	21	3	3	21	21
R86	RU	3	FF-FF-NO	null	3	3	27	null	3
R87	RG	5	AG-AG-GL-GL-AG	21	21	12	12	21	21
R88	RG	5	FF-FF-GL-GL-FF	3	3	12	12	3	3
R89	RG	5	FF-FF-SF-SF-FF	3	3	4	4	3	3
R90	RG	5	SF-SF-AG-AG-SF	4	4	21	21	4	4
R91	RG	5	SF-SF-GL-GL-SF	4	4	12	12	4	4
R92	RG	5	SF-SF-FF-FF-SF	4	4	3	3	4	4
R93	RG	5	AG-AG-SF-SF-AG	21	21	4	4	21	21
R94	RG	5	FF-FF-RO-FF-FF	3	3	29	3	3	3

3.3. Integration with cross-cutting themes

The final classification products after the temporal filter application were then integrated with the cross-cutting themes for each of the 33 years in the period 1985-2017.

Different from the other biomes, all cross-cutting themes had prevalence over Cerrado native vegetation classes, including Pasture class over the Forest, Savanna and Grassland Formations.

4. Validation strategies

Sample points collected by LAPIG were used to generate preliminary estimations of global accuracy, which were extensively used to compare pilot classification results and preliminary filtered maps (Figure 4). Several classification approaches (versions) were compared based on this dataset, differing in terms of variables considered in the feature space and training sampling scheme; the version with the highest estimated global accuracy was then chosen as the definitive one. Also, preliminary accuracy assessments were considered while consolidating integration rules, by changing the prevalence of Cerrado classes over cross-cutting themes (and vice-versa) and then comparing the results of different approaches in terms of estimated global accuracy.

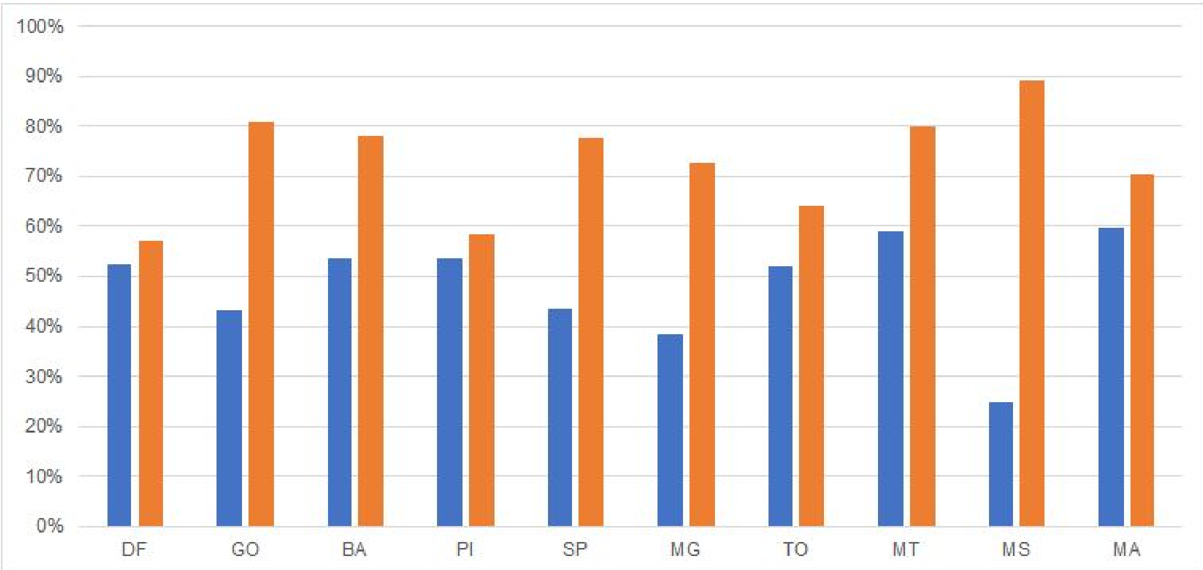


Figure 4. For each Brazilian state in the Cerrado, blue bars show the global accuracies of the classification results before spatial and temporal filters while orange bars shows global accuracies after filtering. Sample points collected by the LAPIG team were used as reference for these comparisons. Large differences in some states indicate that the Landsat mosaic was particularly noisy for the year illustrated (2015), leading to classification noise which was then corrected by the filtering process.

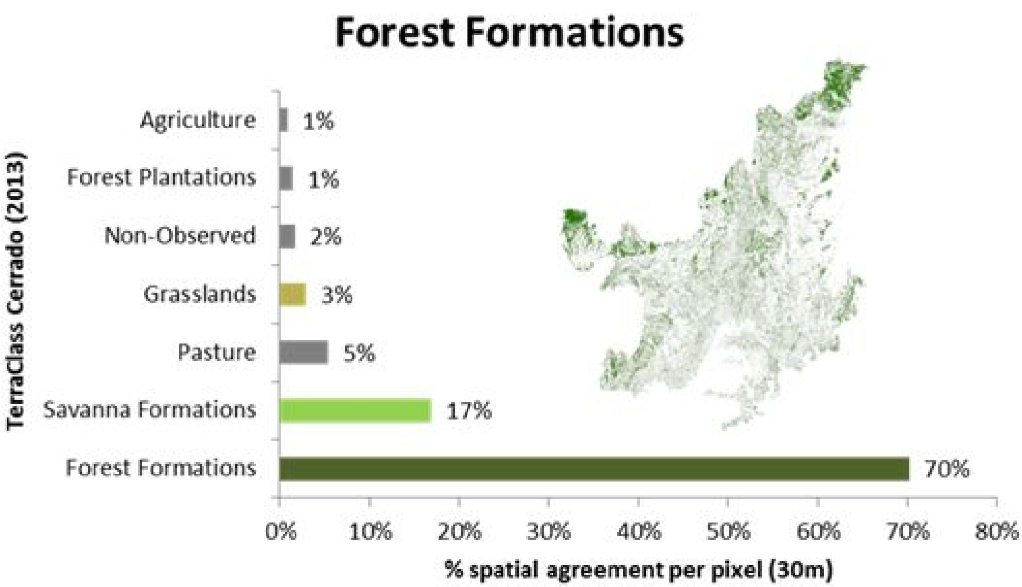
4.1. Reference maps

The Collection 3 LULC maps in the Cerrado were compared with available reference maps available for the entire biome, as TerraClass Cerrado (from 2013 by INPE), the Third National Inventory of Anthropic Emissions (from 2010 by MCTI) and some states from FBDS (from 2012 using Rapideye images).

The Collection 3 mapped more savanna than TerraClass Cerrado, whereas Forest and Grassland Formations were more mapped in TerraClass Cerrado (Table 3). The Forest Formation area mapped in Collection 3 corresponded to 70% of this class mapped in TerraClass Cerrado, Savanna Formation in 47% and Grasslands 58%, mostly due to the confusion between Savanna Formation with Forest and Grassland Formations (Figure 5).

Table 3. Comparison between area mapped in hectares as Forest, Savanna and Grassland Formations in the MapBiomas Collection 3 and TerraClass for the entire biome in 2013.

Land cover	**MapBiomas Collection 3**	**TerraClass Cerrado**
Forest formation	367.186	431.647
Savanna formation	519.975	414.568
Grassland formation	241.996	298.146
Total of native vegetation	1.129.157	1.144.361



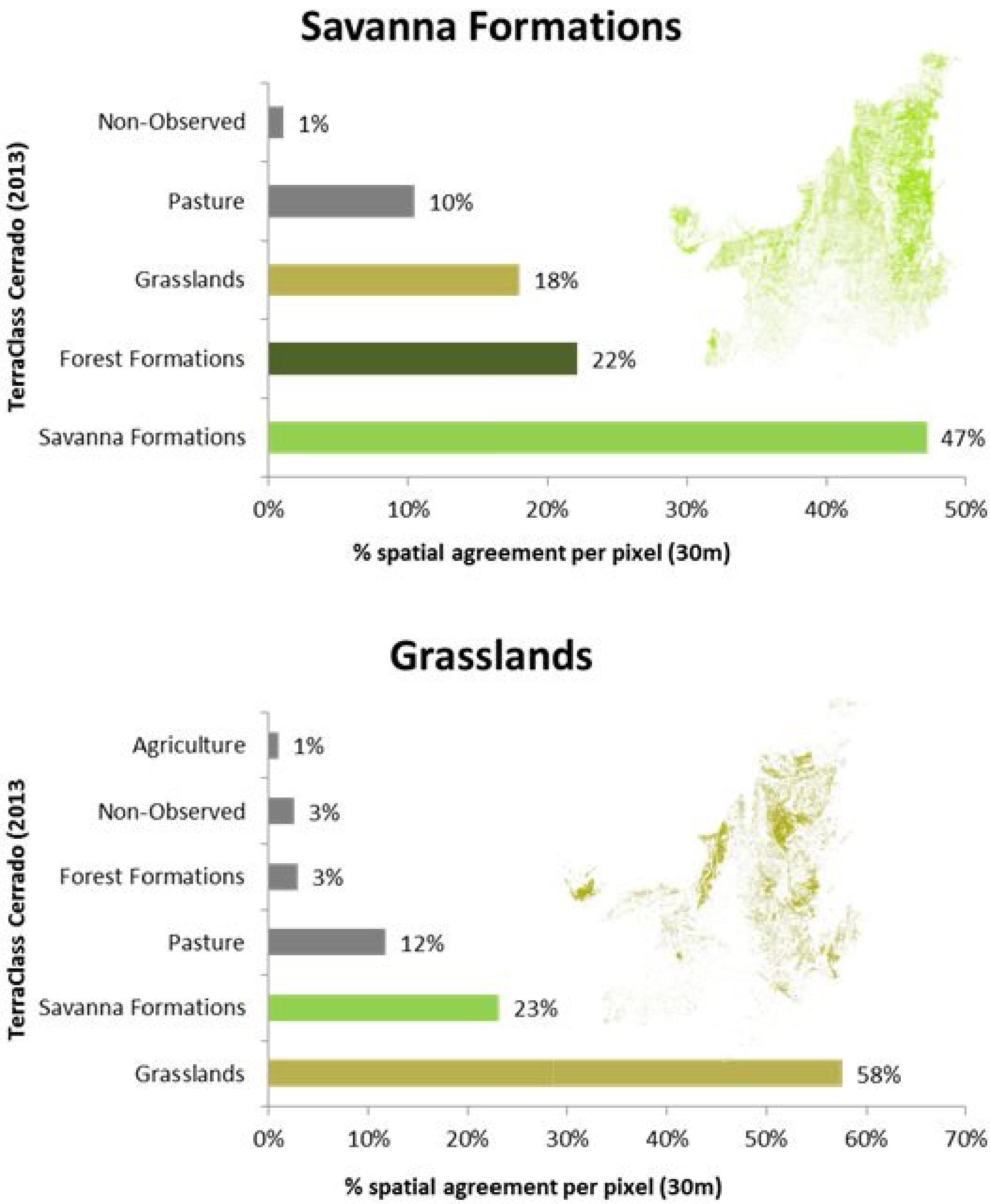


Figure 5. Spatial concordance between the classification of forest, savanna and grassland formations in the MapBiomas Collection 3 and TerraClass Cerrado of the year 2013.

The Third National Inventory of Anthropic Emissions mapped forest and savanna together as FMN and FM (Forest Formations) classes. The correspondent formations mapped in Collection 3 agrees in area to 60% of this class in the Third National Inventory and grasslands (GM, GMN and GSE in the Third National Inventory map) agrees to 58% (Figure 6).

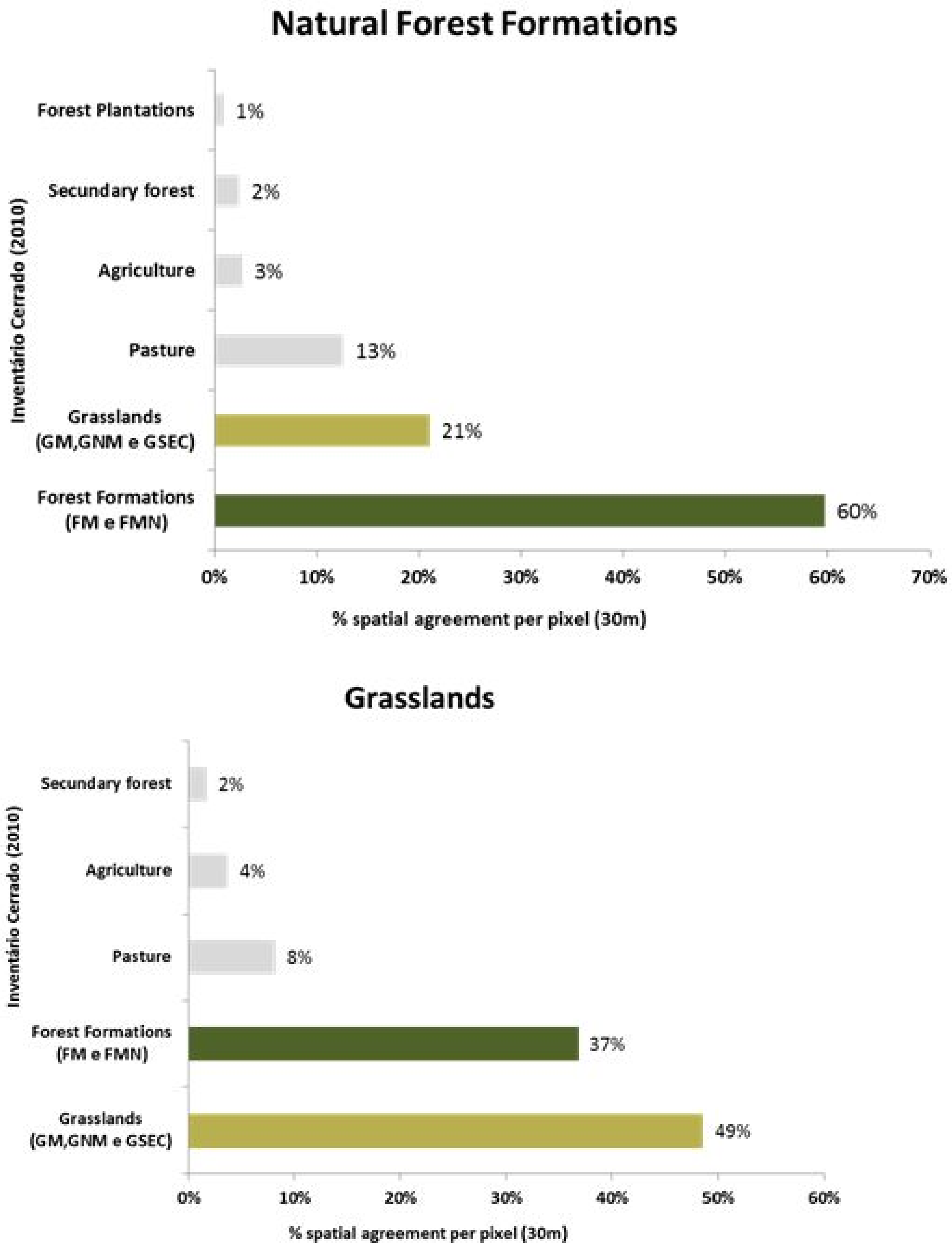


Figure 6. Spatial concordance between the classification of natural forest and grassland formations in the MapBiomas Collection 3 and The Third National Inventory of Anthropic Emissions of the year 2010.

The Forest Formation area mapped in Collection 3 agrees to 61% of this class mapped for FBDS and Non-Forest Formations (Savanna Formations and Grasslands in MapBiomas legend) agrees to 62% (Figure 7). This comparison was made with region mapped for FBDS in the Cerrado biome.

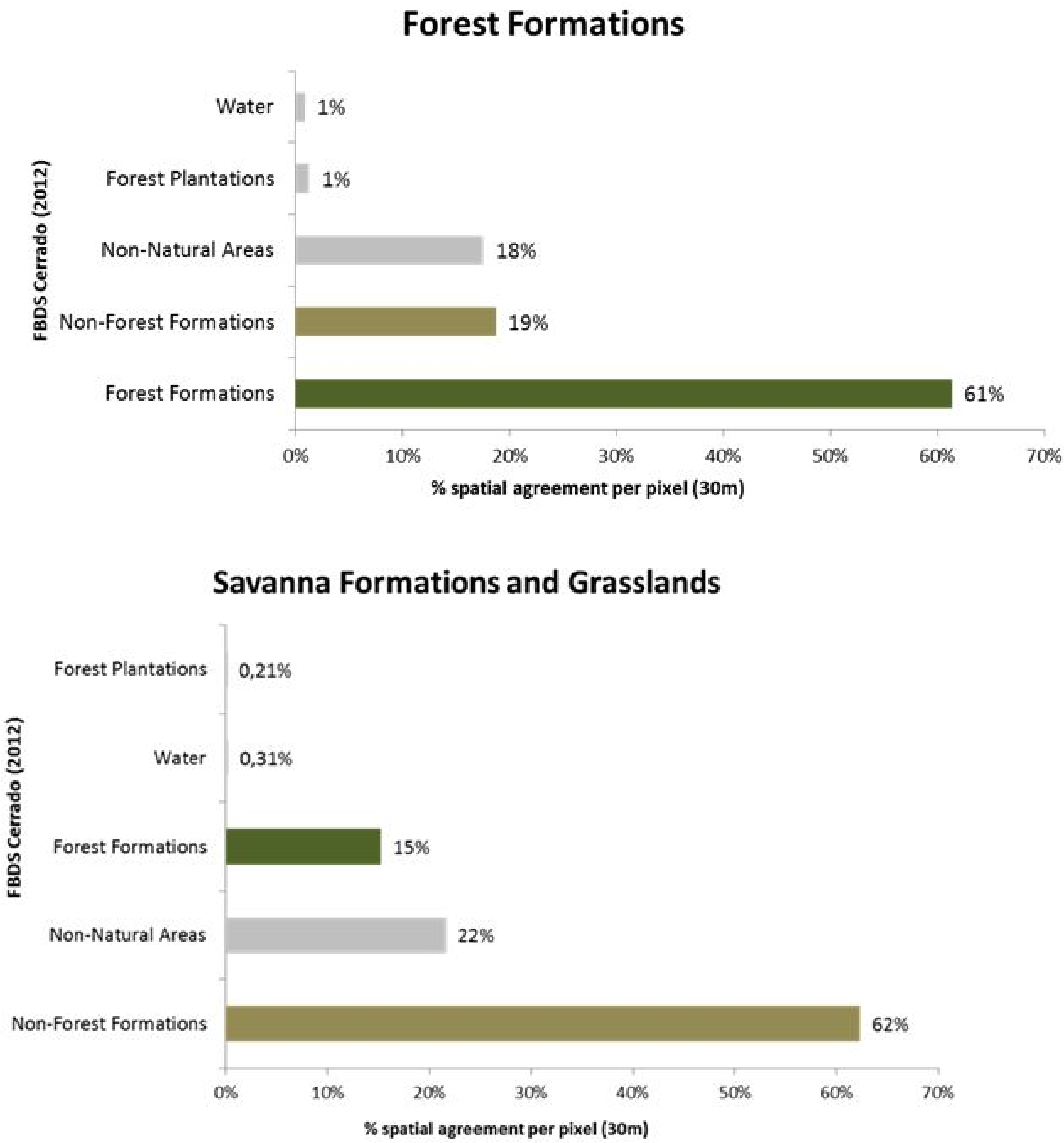


Figure 7. Spatial concordance between the classification of Forest, Savanna and Grassland Formations in the MapBiomas Collection 3 and FBDS data of the year 2012.

5. References

MCTI - MINISTRY OF SCIENCE, TECHNOLOGY AND INNOVATION. 2015. III Brazilian inventory of anthropogenic emissions and removals of greenhouse gases not controlled by the Montreal Protocol. Brasília, DF: MCTI.

MMA - MINISTRY OF THE ENVIRONMENT. 2015. TerraClass Cerrado - Mapping of the Use and Coverage of Cerrado Land in the year 2013. Brasília, DF: MMA.

FBDS - BRAZILIAN FOUNDATION FOR SUSTAINABLE DEVELOPMENT. 2015. Land use mapping for the Cerrado. Rio de Janeiro, RJ: FBDS.