Atlantic Forest - Appendix

Collection 3

Version 1

General coordinator
Marcos Reis Rosa

Team
Fernando Frizeira Paternost
Jacqueline Freitas
Viviane Cristina Mazin
Eduardo Reis Rosa

1 Landsat image mosaics

1.1 Definition of the temporal period

The image selection period for the Atlantic Forest biome was defined aiming to maximize the coverage by useful Landsat images after cloud removing/masking. Despite of the diversity of ecosystems and the great extent of the biome, both latitudinal amplitude and distance of the coast the Atlantic Forest have a well-defined dry period between the months of April to September (Figures 1 and 2).

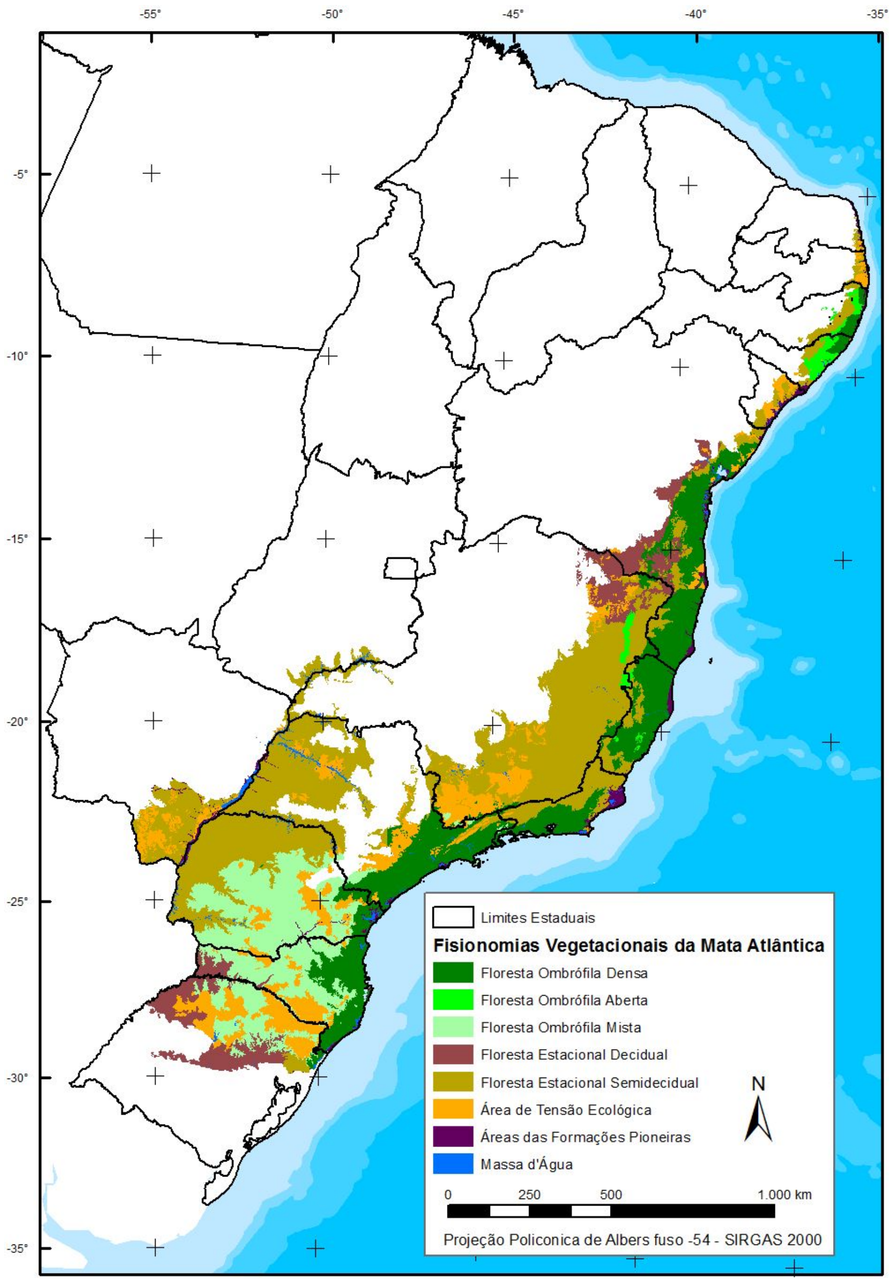

Figure 1. Atlantic Forest biome with forest formation (IBGE, 2017).

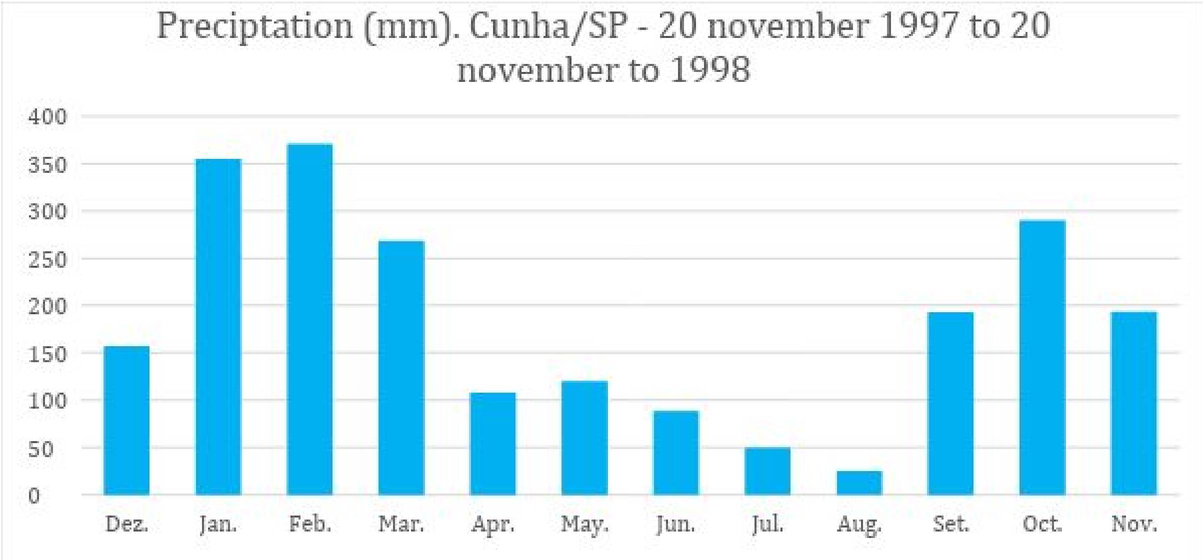

Figure 2. Monthly precipitation values from November 20, 1997 to November 20, 1998 in Cunha (SP) in the Atlantic Forest biome (Arcova et al., 2003).

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 April 1st to August 30th was accepted as a good mosaic with none or few missing information caused by clouds and shades. In some specific cases it was extended significantly the temporal period to include images from September and October. In the Northeast states the period was February 1st to October 30th to maximize the visible areas and avoid missing areas caused by clouds.

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 Atlantic Forest 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 Atlantic Forest 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	
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 Atlantic Forest biome comprised a subset of 29 variables (Table 2) from the 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 Atlantic Forest biome Landsat image mosaics in the MapBiomas Collection 3 (1985-2017).

ID	Variable	Description	Statistics	Temporal range	Script acronym	Group
1	Blue	Landsat band	median	mosaic months	'median_blue'	Landsat band
2	Green	Landsat band	median	mosaic months	'median_green'	Landsat band
3	Near Infrared (NIR)	Landsat band	median	mosaic months	'median_nir'	Landsat band
4	Red	Landsat band	median	mosaic months	'median_red'	Landsat band
5	Shortwave Infrared (SWIR) 1	Landsat band	median	mosaic months	'median_swir1'	Landsat band
6	Shortwave Infrared (SWIR) 2	Landsat band	median	mosaic months	'median_swir2'	Landsat band
7	Evi 2	Enhanced Vegetation Index 2	median	mosaic months	'median_evi2'	Vegetation Index
8	Ndvi	Normalized Difference Vegetation Index	median	mosaic months	'median_ndvi'	Vegetation Index
9	Ndvi Dry	Normalized Difference Vegetation Index	median	year -below first quartile	'median_ndvi_dry'	Vegetation Index
10	Ndvi Wet	Normalized Difference Vegetation Index	median	year - above third quartile	'median_ndvi_wet'	Vegetation Index
11	Savi	Soil-adjusted Vegetation Index	median	mosaic months	'median_savi'	Vegetation Index
12	Sefi	Savanna Ecosystem Fraction Index	median	mosaic months	'median_sefi'	Vegetation Index
13	Sefi	Savanna Ecosystem Fraction Index	standard deviation	complete year	'stdDev_sefi'	Vegetation Index
14	Ndwi	Normalized Difference Water Index	median	mosaic months	'median_ndwi'	Water Index
15	Ndwi Dry	Normalized Difference Water Index	median	year -below first quartile	'median_ndwi_dry'	Water Index
16	Ndwi Wet	Normalized Difference Water Index	median	year - above third quartile	'median_ndwi_wet'	Water Index
17	Gv	green vegetation fraction	median	mosaic months	'median_gv'	Spectral Mixture Modeling
18	Gvs	GV / (100 - shade)	median	mosaic months	'median_gvs'	Spectral Mixture Modeling
19	Ndfi	Normalized Difference Fraction Index	median	mosaic months	'median_ndfi'	Spectral Mixture Modeling
20	Ndfi	Normalized Difference Fraction Index	standard deviation	complete year	'stdDev_ndfi'	Spectral Mixture Modeling
21	Ndfi Dry	Normalized Difference Fraction Index	median	year -below first quartile	'median_ndfi_dry'	Spectral Mixture Modeling

22	Ndfi Wet	Normalized Difference Fraction Index	median	year - above third quartile	'median_ndfi_wet'	Spectral Mixture Modeling
23	Npv	npv fraction	median	mosaic months	'median_npv'	Spectral Mixture Modeling
24	Npv	npv fraction	standard deviation	complete year	'stdDev_npv'	Spectral Mixture Modeling
25	Shade	shade fraction	median	mosaic months	'median_shade'	Spectral Mixture Modeling
26	Soil	soil fraction	median	mosaic months	'median_soil'	Spectral Mixture Modeling
27	Soil	soil fraction	standard deviation	complete year	'stdDev_soil'	Spectral Mixture Modeling
28	Green spatial texture	Spatial texture	median	mosaic months	'textG'	Spatial Texture
29	Slope	Slope	-	permanent	'slope'	Geomorphometric

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 LULC 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. Later, a set of polygons delineating zones with errors in some classes (*e.g.*, omission or commission) was 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 SOS Mata Atlântica, 2014 and FBDS, 2013 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 LULC 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 from SOS Mata Atlântica, 2014 and FBDS, 2013. 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, 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 where 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 LULC 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 Post-classification

3.1 Temporal filter

The temporal filter rules were adapted for the LULC classes used in the Atlantic Forest 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 Atlantic Forest 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	Typ e	Kerne l	Notes	Tminus 2	Tminus 1	T	Tplus 1	Tplus 2	Resul t
R01	RG	3	Fill NO		3	27	4		3
R02	RG	3	Fill NO		3	27	21		3
R03	RG	3	Fill NO		3	27	26		3
R04	RG	3	Fill NO		3	27	12		3
R05	RG	3	Fill NO		3	27	27		3
R06	RG	3	Fill NO		4	27	3		4
R07	RG	3	Fill NO		4	27	21		4
R08	RG	3	Fill NO		4	27	26		4
R09	RG	3	Fill NO		4	27	12		4
R10	RG	3	Fill NO		4	27	27		4
R11	RG	3	Fill NO		21	27	3		21
R12	RG	3	Fill NO		21	27	4		21
R13	RG	3	Fill NO		21	27	12		21
R14	RG	3	Fill NO		21	27	26		21
R15	RG	3	Fill NO		26	27	3		26
R16	RG	3	Fill NO		26	27	21		26
R17	RG	3	Fill NO		26	27	4		26
R18	RG	3	Fill NO		26	27	12		26
R19	RG	3	Fill NO		26	27	27		26
R20	RG	3	Fill NO		12	27	3		12
R21	RG	3	Fill NO		12	27	21		12
R22	RG	3	Fill NO		12	27	4		12
R23	RG	3	Fill NO		12	27	26		12
R24	RG	3	Fill NO		12	27	27		12
R25	RG	3	Correct NV		3	21	3		3
R26	RG	3	Fill NO		3	27	3		3
R27	RG	3	Correct FF		3	26	3		3
R28	RG	3	Correct FF		3	12	3		3
R29	RG	3	Correct FF		3	4	3		3
R30	RG	3	Fill NO		4	27	4		4
R31	RG	3	Correct SV		4	21	4		4
R32	RG	3	Correct SV		4	3	4		4
R33	RG	3	Correct SV		4	12	4		4
R34	RG	3	Correct SV		4	26	4		4
R35	RG	3	Correct AG		21	3	21		21
R36	RG	3	Correct AG		21	4	21		21
R37	RG	3	Fill NO		21	27	21		21
R38	RG	3	Correct AG		21	12	21		21
R39	RG	3	Correct AG		21	26	21		21
R40	RG	3	Correct CD		26	21	26		26
R41	RG	3	Fill NO		26	27	26		26
R42	RG	3	Correct CD		26	3	26		26
R43	RG	3	Correct CD		26	12	26		26
R44	RG	3	Correct CD		26	4	26		26
R45	RG	3	Correct CD		12	21	12		12
R46	RG	3	Fill NO		12	27	12		12
R47	RG	3	Correct FC		12	26	12		12
R48	RG	3	Correct FC		12	3	12		12
R49	RG	3	Correct FC		12	4	12		12
R50	RG	3	Fill NO		21	27	27		21

R51	RG	3	Fill NO		25	27	27		25
R52	RG	3	Correct FC		12	21	12		12
R53	RG	5	Correct false Regen	21	21	3	3	21	21
R54	RG	5	Correct false Regen	21	21	12	12	21	21
R55	RG	5	Correct false Regen	21	21	4	4	21	21
R56	RG	5	Correct false deforestation	3	3	21	21	3	3
R57	RG	5	Correct false deforestation	4	4	21	21	4	4
R58	RG	5	Correct false deforestation	12	12	21	21	12	12
R59	RG	5	Correct false use change	25	25	21	21	25	25
RP01	RP	3	Fill NO		27	27	26		26
RP02	RP	3	Fill NO		27	27	4		4
RP03	RP	3	Fill NO		27	27	3		3
RP04	RP	3	Fill NO		27	21	21		21
RP05	RP	3	Fill NO		27	26	26		26
RP06	RP	3	Fill NO		27	12	12		12
RP07	RP	3	Fill NO		27	3	3		3
RP08	RP	3	Fill NO		27	4	4		4
RP09	RP	3	Fill NO		27	3	4		3
RP10	RP	3	Fill NO		27	27	21		21
RU01	RU	3	Fill NO		21	27	27		21
RU02	RU	3	Fill NO		3	27	27		3
RU03	RU	3	Fill NO		12	12	27		12
RU04	RU	3	Fill NO		4	27	27		4
RU05	RU	3	Fill NO		3	3	27		3
RU06	RU	3	Fill NO		4	4	27		4
RU07	RU	3	Fill NO		21	21	27		21
RU08	RU	3	Fill NO		25	25	27		25
RU09	RU	3	Correct NV		25	25	21		25
RU10	RU	3	Fill NO		26	26	27		26
RU11	RU	3	Fill NO		12	12	27		12
RU12	RU	3			21	21	25		21

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 the standard rules, with no specific rule. As output of this step, a final LULC map for each chart of the Atlantic Forest biome for each year was obtained.

4 Validation strategies

4.1 Use of reference maps

Overall, there have been few previous initiatives on LULC mapping in the Atlantic Forest biome with spatial and thematic detail compatible with the MapBiomas Project. The best reference is the “Atlas dos remanescentes Florestais da Mata Atlântica” produced by “Fundação SOS Mata Atlântica” and “Instituto Nacional de Pesquisas Espaciais-INPE”.

5 References

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

ARCOVA, F. C. S.; CICCO, V. DE; ROCHA, P. A. B. Precipitação efetiva e interceptação das chuvas por floresta de Mata Atlântica em uma microbacia experimental em Cunha - São Paulo. Revista Árvore, v. 27, n. 2, p. 257–262, 2003.

FUNDAÇÃO BRASILEIRA PARA O DESENVOLVIMENTO SUSTENTÁVEL (FBDS). Repositório público de mapas e shapefiles para download. Disponível em: <http://www.fbds.org.br/article.php3?id_article=594>. Acesso em: 30 maio. 2018.

FUNDAÇÃO SOS MATA ATLÂNTICA; INPE. Atlas dos Remanescentes Florestais da Mata Atlântica. Disponível em: <http://mapas.sosma.org.br/>. Acesso em: 30 maio. 2018.

IBGE. Vegetação RADAM. Disponível em: <ftp://geoftp.ibge.gov.br/informacoes_ambientais/acervo_radambrasil/vetores/>. Acesso em: 30 maio. 2018.