Simplified method of comparing covers from different years of MapBiomas 7 (Brazilian Land Cover Classification), using standard QGIS tools - Resources for Talk on Foss4g 2023

Leandro Meneguelli Biondo; National Institute of the Atlantic Forest, Ministry of Science, Technology and Innovation, Brazil;
Bernardo Araújo de Moraes Trovao, National Institute of Colonization and Agrarian Reform, Ministry of Agrarian Development, Brazil;
Clayton Borges da Silva, Energy Research Office - EPE, Brazil;

Simplified method of comparing covers from different years of MapBiomas 7, using standard QGIS tools (Raster Calculator, Pixels to points and Sample Raster Values) to create images that represent land use change and a database of pixels for combinatorial analysis time of classes in the same places over the years, for the production of maps and representation of changes in Sankey Diagram-type graphs. Source code, sample data, references and history at:

https://github.com/leandromet/geo_postgis/tree/master/2023_carbon_biodiversity

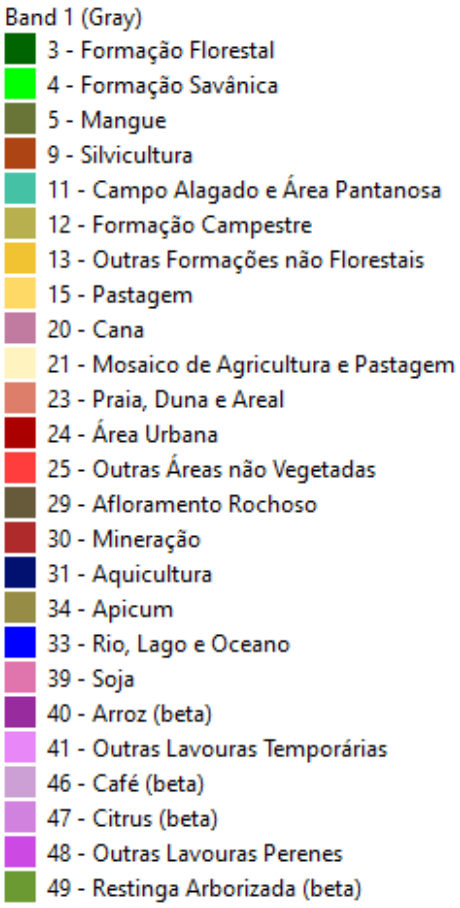
For all processing, database hosting and data analysis it was used a notebook computer with i7-10750H processor (6 cores, 12 threads) with 16GB RAM, 1TB SSD running Ubuntu 22.10 with QGIS 3.22, PostgeSQL 14.7, PostGIS 3.2.3 and PGAdmin 4 v6.21. There was no tuning, variable change or performance adjust from the default installation of all the tools out of Ubuntu repositories via apt-get.

Part 1 - processing, grouping and presentation of georeferenced images with the general panorama of changes in land use.

This methodology was organized with the objective of supporting the characterization of areas in the municipalities of São Sebastião do Anta, Inhapim, Alvarenga, Tarumirim, Conselheiro Pena, Santa Rita do Itueto and Resplendor in the state of Minas Gerais. The region is of interest to INMA due to the concentration of rocky fields that were the target of field expeditions and collection of ecological and biodiversity data, within the scope of the PCI/INMA program.

Mapbiomas collection 7 - 1985 to 2020
The subtitle file suggested by the platform for QGIS has 25 distinct classes of land use, which were separated into 12 natural and 13 anthropic or non-natural for the application of filters of probable natural vegetation or remaining native. (translation of classes on the next page)

Natural	
3	Formação Florestal
4	Formação Savânica
5	Mangue
11	Campo Alagado e Área Pantanosa
12	Formação Campestre
13	Outras Formações não Florestais
23	Praia, Duna e Areal
25	Outras Áreas não Vegetadas
29	Afloramento Rochoso
34	Apicum
33	Rio, Lago e Oceano
49	Restinga Arborizada (beta)
Artificial	
9	Silvicultura
15	Pastagem
20	Cana
21	Mosaico de Agricultura e Pastagem
24	Área Urbana
30	Mineração
31	Aquicultura
39	Soja
40	Arroz (beta)
41	Outras Lavouras Temporárias
46	Café (beta)
47	Citrus (beta)
48	Outras Lavouras Perenes

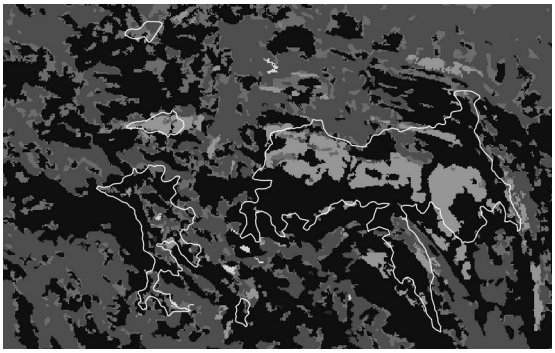


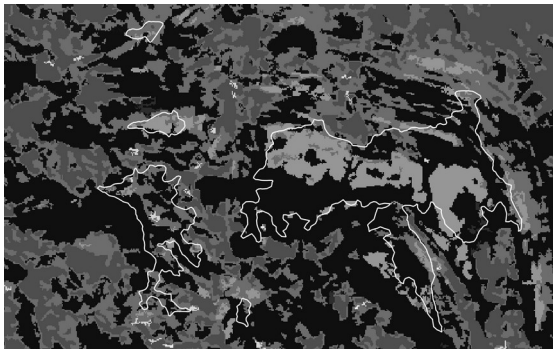
class_id	digital number	original name	color	translated
1	3	Formação Florestal	#006400	Forest
2	4	Formação Savânica	#00ff00	Savana
3	5	Mangue	#687537	Mangrove
4	9	Silvicultura	#ad4413	Silviculture
5	11	Campo Alagado e Área Pantanosa	#45c2a5	Wetland
6	12	Formação Campestre	#b8af4f	Coutryside
7	13	Outras Formações não Florestais	#f1c232	Other non Forest
8	15	Pastagem	#ffd966	Pasture
9	20	Cana	#c27ba0	Cane
10	21	Mosaico de Agricultura e Pastagem	#fff3bf	Agriculture and Pasture
11	23	Praia, Duna e Areal	#dd7e6b	Sand
12	24	Área Urbana	#aa0000	Urban
13	25	Outras Áreas não Vegetadas	#ff3d3d	Non Vegetated
14	29	Afloramento Rochoso	#665a3a	Rock
15	30	Mineração	#af2a2a	Minning
16	31	Aquicultura	#02106f	Aquiculture
17	34	Apicum	#968c46	Peak
18	33	Rio, Lago e Oceano	#0000ff	Waterstream
19	39	Soja	#e075ad	Soy
20	40	Arroz (beta)	#982c9e	Rice
21	41	Outras Lavouras Temporárias	#e787f8	Other Crops
22	46	Café (beta)	#cca0d4	Coffe
23	47	Citrus (beta)	#d082de	Citrus
24	48	Outras Lavouras Perenes	#cd49e4	Other Perennial Crops
25	49	Restinga Arborizada (beta)	#6b9932	Restinga with Tree

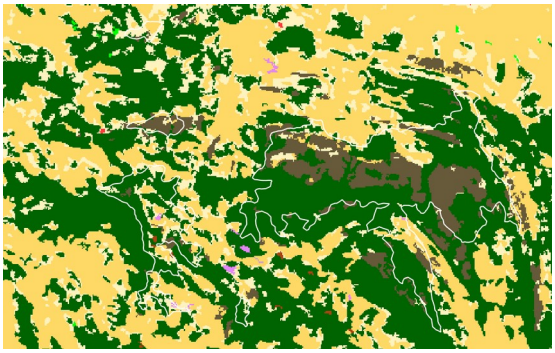
Example area near the Sete Salões State Park (Middle Rio Doce)

1. The processing of mapbiomas raster data was done in four steps: Crop in all years from 1985 to 2020 to speed up processing, a region limited by the coordinates -44,554,-20,507 : -39,349,-17,209 (EPSG:4674) was considered in files of 19,312x12,240 pixels equal to the originals. The 960MB Brazil TIFF file results in a 220MB clipping for this area with dimensions of 580x370Km. The region of interest with 26 boundaries of rocky fields is in the center of the chosen polygon and is approximately 75x45Km. Example for clipping:

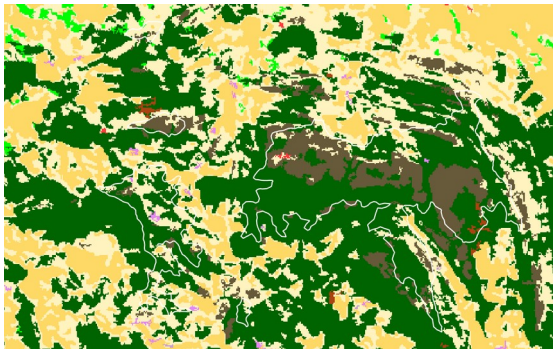
   gdal_translate -projwin -44.553 -17.209 -39.349 -20.507 -of GTiff source.tif extent.tif


1985 original


2020 original



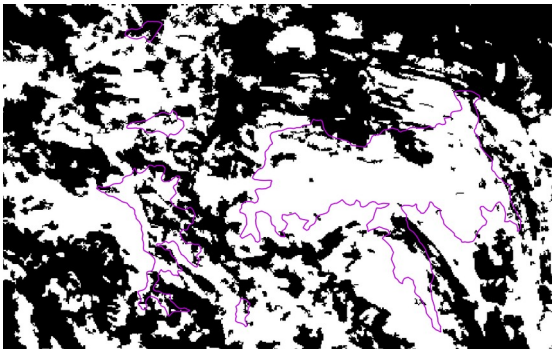
1985 colored



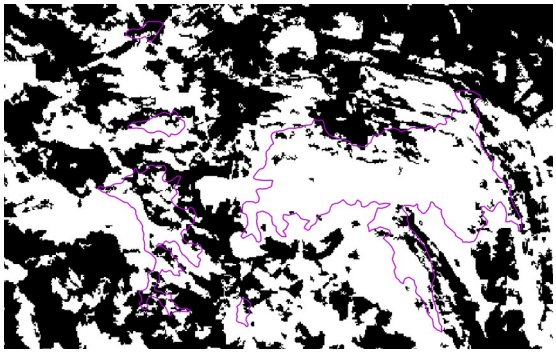
2020 colored

2. Grouping of natural and artificial classes for each year. Used the QGIS Raster Calculator, in a TIFF file with values 1 for natural and 0 for the rest. For an extent layer with the mapbiomas classes it looks like this:

   ("extent@1" = 3 or "extent@1" = 4 or "extent@1" = 5 or "extent@1" = 11 or "extent@1" = 12 or "extent@1" = 13 or "extent@1" = 23 or "extent@1" = 25 or "extent@1" = 29 or "extent@1" = 34 or "extent@1" = 33 or "extent@1" = 49)



Natural Land Use 1985



Natural Land Use 2020

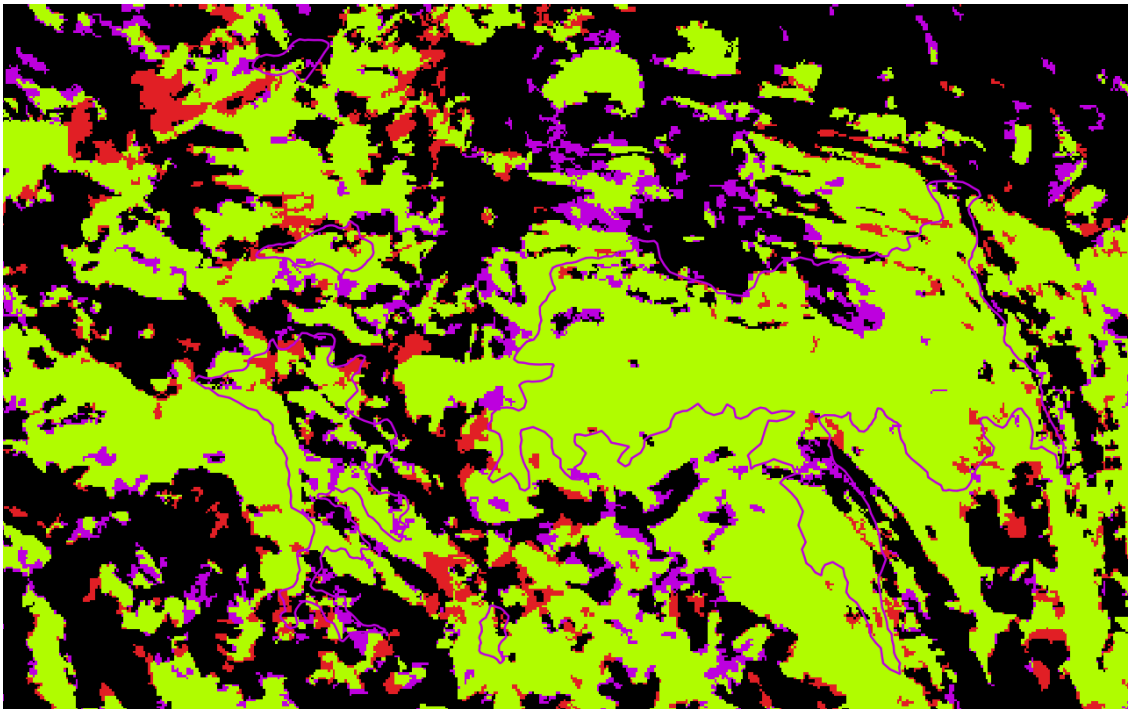
3. In cases where you want to accumulate information on natural areas covering more than one year, use the previous expression in the raster calculator for all the years you want to group together, with the OR operator between them. Thus, the result is also a file with 1 for all pixels that in any of the years were identified as natural class.

   ("extent_year1985@1" = 3 or "extent_year1985@1" = 4 or "extent_year1985@1" = 5 or "extent_year1985@1" = 11 or "extent_year1985@1" = 12 or "extent_year1985@1" = 13 or "extent_year1985@1" = 23 or "extent_year1985@1" = 25 or "extent_year1985@1" = 29 or "extent_year1985@1" = 34 or "extent_year1985@1" = 33 or "extent_year1985@1" = 49)
   OR
   ("extent_year1986@1" = 3 or "extent_year1986@1" = 4 or "extent_year1986@1" = 5 or "extent_year1986@1" = 11 or "extent_year1986@1" = 12 or "extent_year1986@1" = 13 or "extent_year1986@1" = 23 or "extent_year1986@1" = 25 or "extent_year1986@1" = 29 or "extent_year1986@1" = 34 or "extent_year1986@1" = 33 or "extent_year1986@1" = 49)

4. Generate a pixel-by-pixel change raster using two years or two groupings of years to be compared. A simple and direct method was applied, using the masks generated in step 2 or 3, the sum of two files with different weights is made so that later the compositions of natural areas

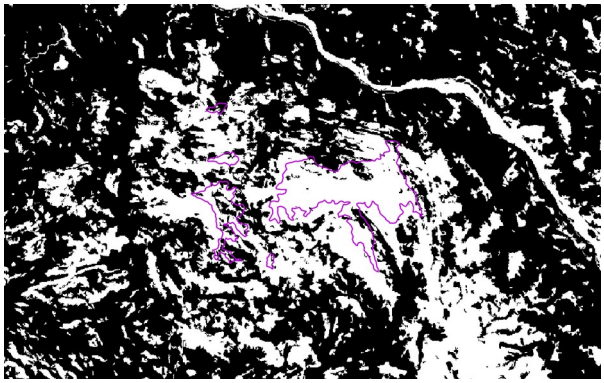
can be evaluated. Weights of 2 were adopted for the first file and 3 for the second, so as not to be confused with the value 1 for the natural area of original masks, and we have the possibility of pixels assuming values 0, 2, 3 or 5 in the result. Value 0 means no natural area in both files, value 2 only in the first, value 3 only in the second and value 5 indicates natural class in both.

In the following example, the land use masks with natural class from the years 1985 to 2000, and from 2011 to 2020 were grouped. Afterwards, the weighted sum was made as explained above, and we have the result of the changes identified between any natural pixel in the oldest 16-year period with the most recent 10 years. The 10-year jump was made because it was a sufficient window to indicate that areas that changed from natural to artificial were lost and the reverse recovered.

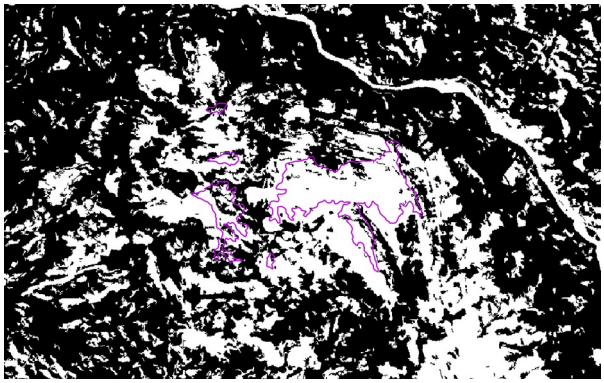


Result of the weighted sum of 1985 and 2020, in black value 0, green value 5 (persistent natural), red value 2 (loss of natural coverage) and purple value 3 (gain of natural coverage) that represent the identifiable change in the pixel with temporal distance of 35 years.

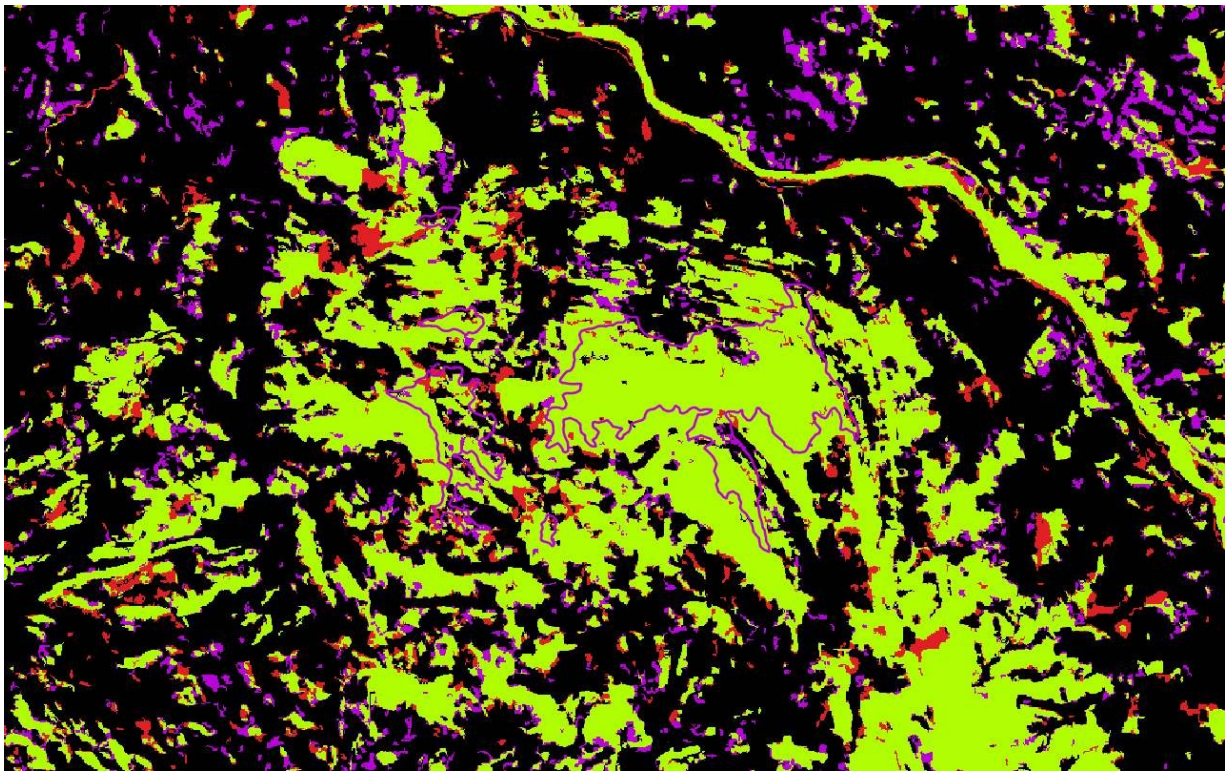
5. To compare multiple files with this simplified method, other increasing prime numbers can be used, which are not the result of the sum of those previously used, this way it is possible to identify the origin of the combinations only with the pixel value (for example if used 2, 3 and 7 when adding three files, we know that the result will have these values if there was a natural area in only one of them, value 0 if it did not occur in any, 5 if persistent in files one and two, 10 if in files two and three or 9 if in one and three, finally value 12 recurring case in all three files - final raster has pixels with possible values 0-2-3-5-7-9-10-12)



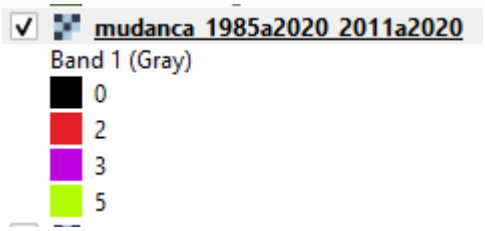
Natural pixel grouping 1985 to 2000



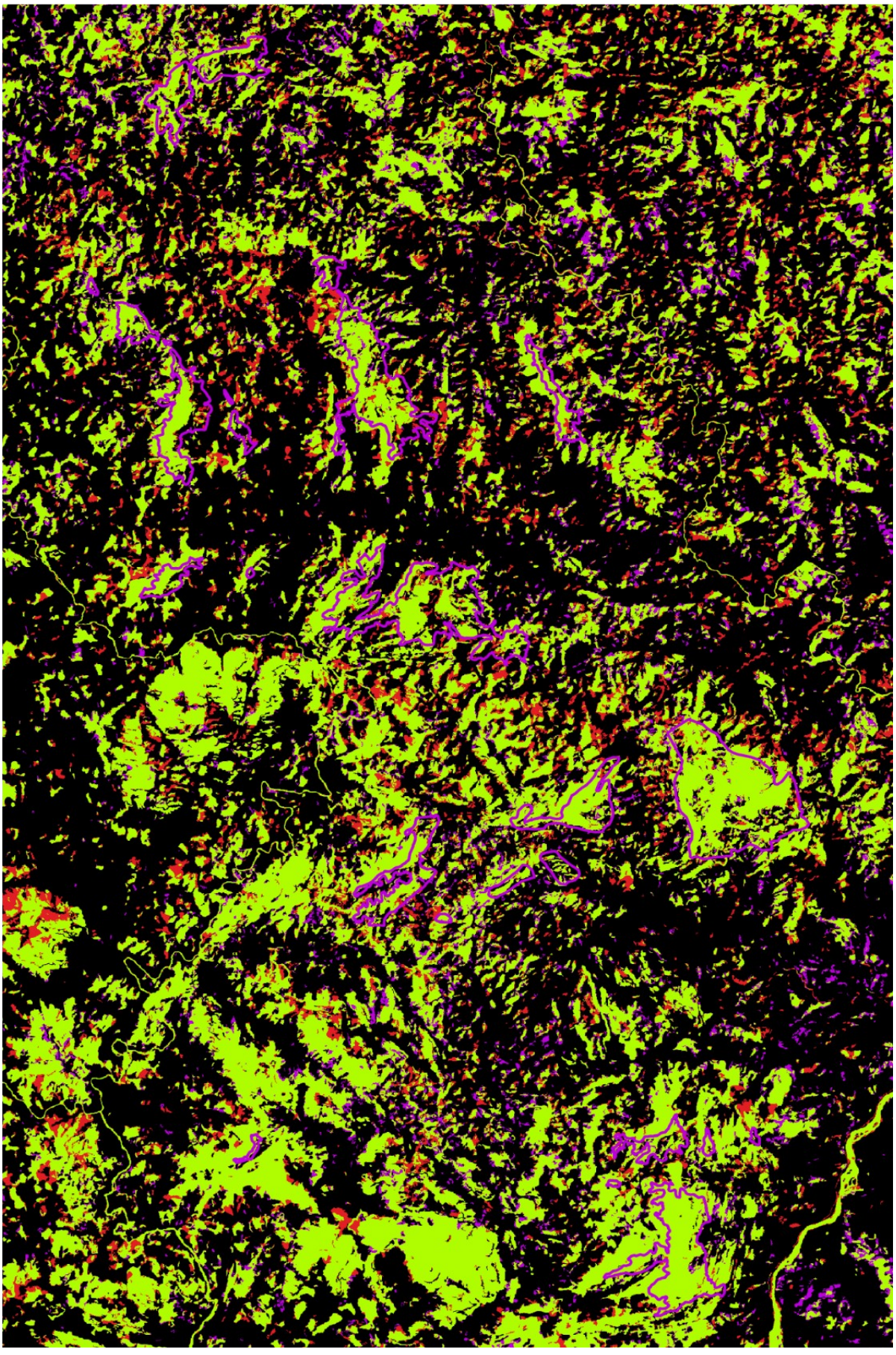
Natural pixel grouping 2011 to 2020



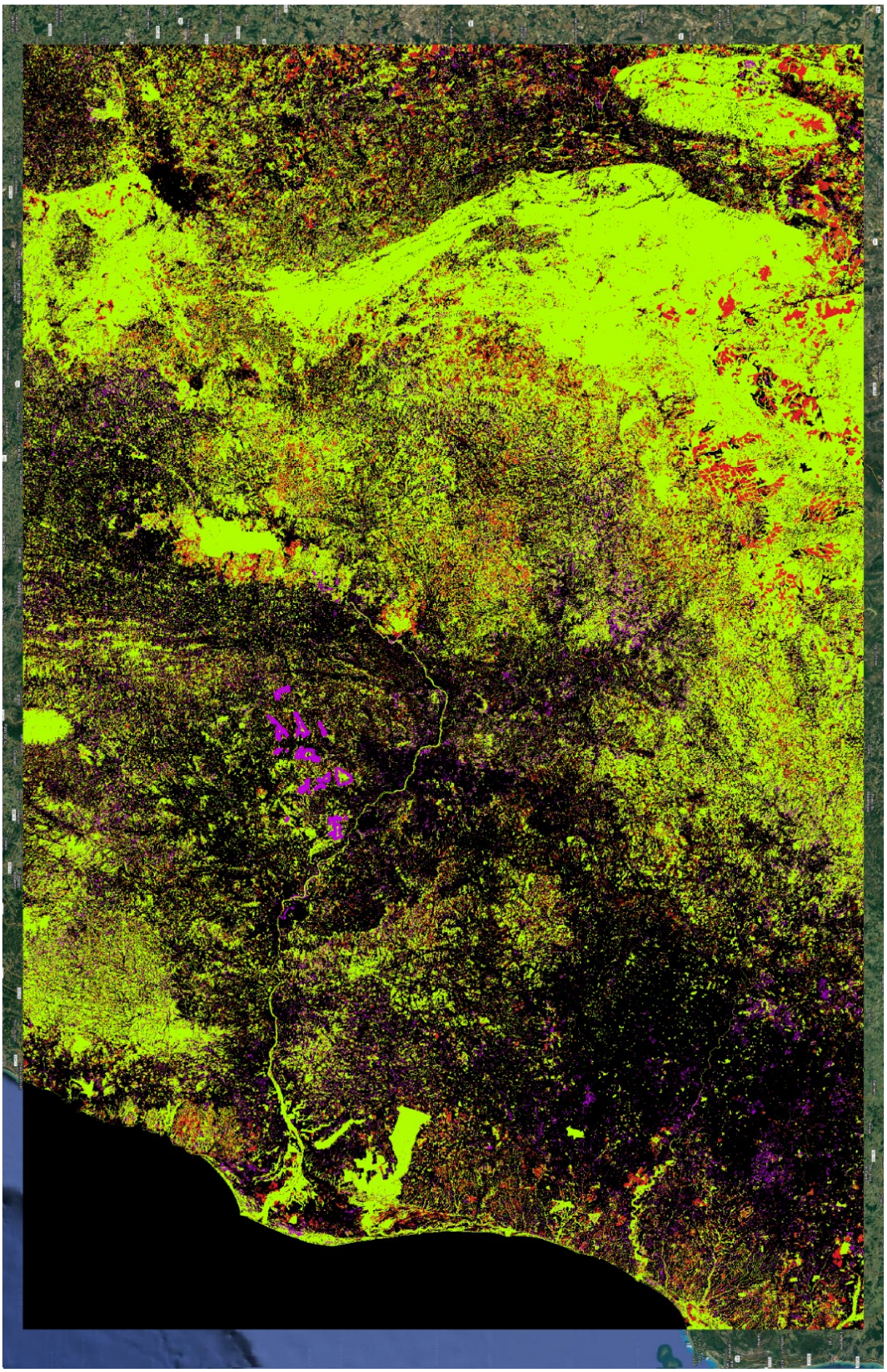
Sum of both images weighted 2 to 85-00 and weighted 3 to 11-20.



Result (rotated 90 degrees clockwise) for Campos rupestre region



Result (rotated 90 degrees clockwise) for the full original raster file



Part 2 - Analysis of land use changes pixel by pixel using a spatial database.

As shown in the previous part, it is relatively simple to evaluate the differences between two superimposed images of any region in Brazil using the “mapbiomas” land use classification data. There are several ways to do the same process, which is faster and more efficient in smaller areas and allows a visual or statistical analysis of the images with a pixel count per class, display of the differences in colors, only the changes, by classes or grouping of classes.

However, it is necessary to repeat the image processing for all the analyzes that one intends to do that have some difference in the area of application, classes that are grouped, sampling period or combinations between classes of different years of classification. As we are using raster data from the same location and which has overlapping information for different sampling years, we are looking at different pixel tables of the same size and with data of the same nature. It is possible to store these tabular data files in different ways, initially we are using TIF files which, as offered by Mapbiomas, have about 1GB of data per year.

We can use the raster data storage format in a PostgreSQL database, for example, which only changes the location where the raster will be stored and more or less maintains its size, which facilitates organization and access by different users and systems. Still in a “mosaic” format that increases the total file size, as the data is stacked in several layers with different resolutions, with acceleration of the display of images on a desktop or online geospatial environment. In these cases, to reprocess and compare different regions and periods, we have to once again access the class tables contained in the image files and run relation algorithms between them.

One way that has become popular to accelerate the access and processing of vector spatial data is to use a spatial database. In the case of PostgreSQL there is the PostGIS extension that works efficiently and quickly to store, process and provide spatial data to applications and users. It is possible to polygonize the TIFF images and place them in a database of this type, and start carrying out vector operations with the data to take advantage of the spatial indices that speed up geolocation operations such as intersection, touch and superimposition.

By polygonizing the data, the areas that are composed of the same class are homogenized, which instead of having thousands of pixels becomes a contour geometry that delimits the original area occupied by similar neighbors. Both this polygonization modifies the original data and if done in a very high resolution (equivalent to 30m of the raster file) the data can become heavy and the processing more time consuming and problematic than using image files.

For this section of the Atlantic Forest worked on in the previous part, an unconventional method of representing the rasters in a database was applied, representing all the pixels of the images as records of a PostgreSQL table. Using the “Raster pixels to points” tool with a TIF clipping of 580x370Km with output in a local database table, a table with 19,312x12,240 = 236,378,880 lines was assembled, each with a unique ID and a spatial point with the coordinates of the center of the original pixel.

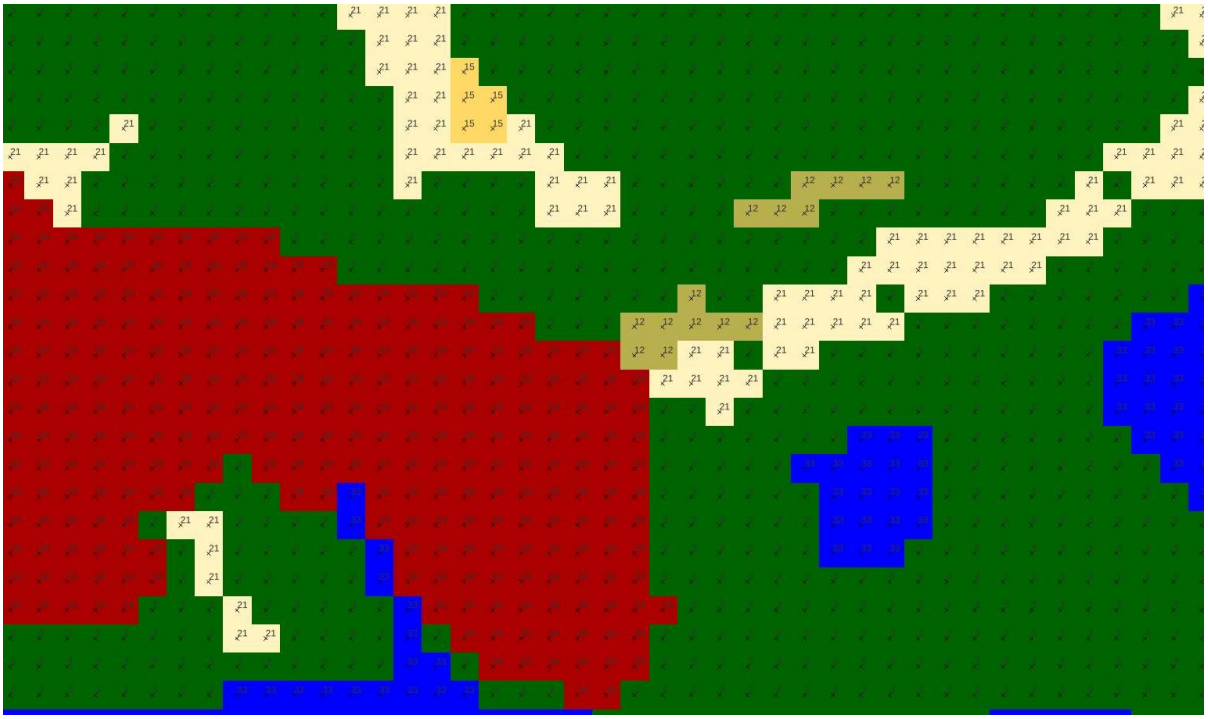
The only data cleaning done was to remove the database records with a classification value of 0 in 1985 and 2020, which were unsampled points, most of them over the ocean. The table of this clipping was then with 180 million rows.

Then a “stacked” raster of the 36 years of classification of the mapbiomas was made, the pre-processing was done with the “Clip Raster by Extent” tool that was clipped every year (explained at the beginning of part 1) and then the “Clip Raster by Extent” tool. Merge…”, checking the option “place each input file into a separate band” with the selection of the 36 files placed in ascending order by sampling year. The resulting raster has 36 bands or layers with the original resolution, which is useful for the next step, putting this data in the same database table.

With the “Sample Raster Values” tool, the 36-layer raster was placed as a raster layer, the database table as an input layer and a new table name in the same database as “Sampled”. This tool samples the raster value for each feature of the input vector and stores it in a column of the output vector. In the case of a raster with more than one layer, the tool creates a column for each layer, and as our sampling vector is a point in the center of each pixel, the result is a table with the same 180 million lines and 38 columns that are the unique ID, the spatial point with coordinate at the center of the pixel and 36 years of mapbioma data for each pixel.

The complete table with the data for this test region had a total of 50GB, against the original 8GB of TIF rasters with “DEFLATE” compression. As the type of data stored in the database was “byte” for all the class information of each attribute of each record, the physical storage space is practically the same as a TIF file without compression in 8bits (after all, it is also a table that each item is 1 byte).

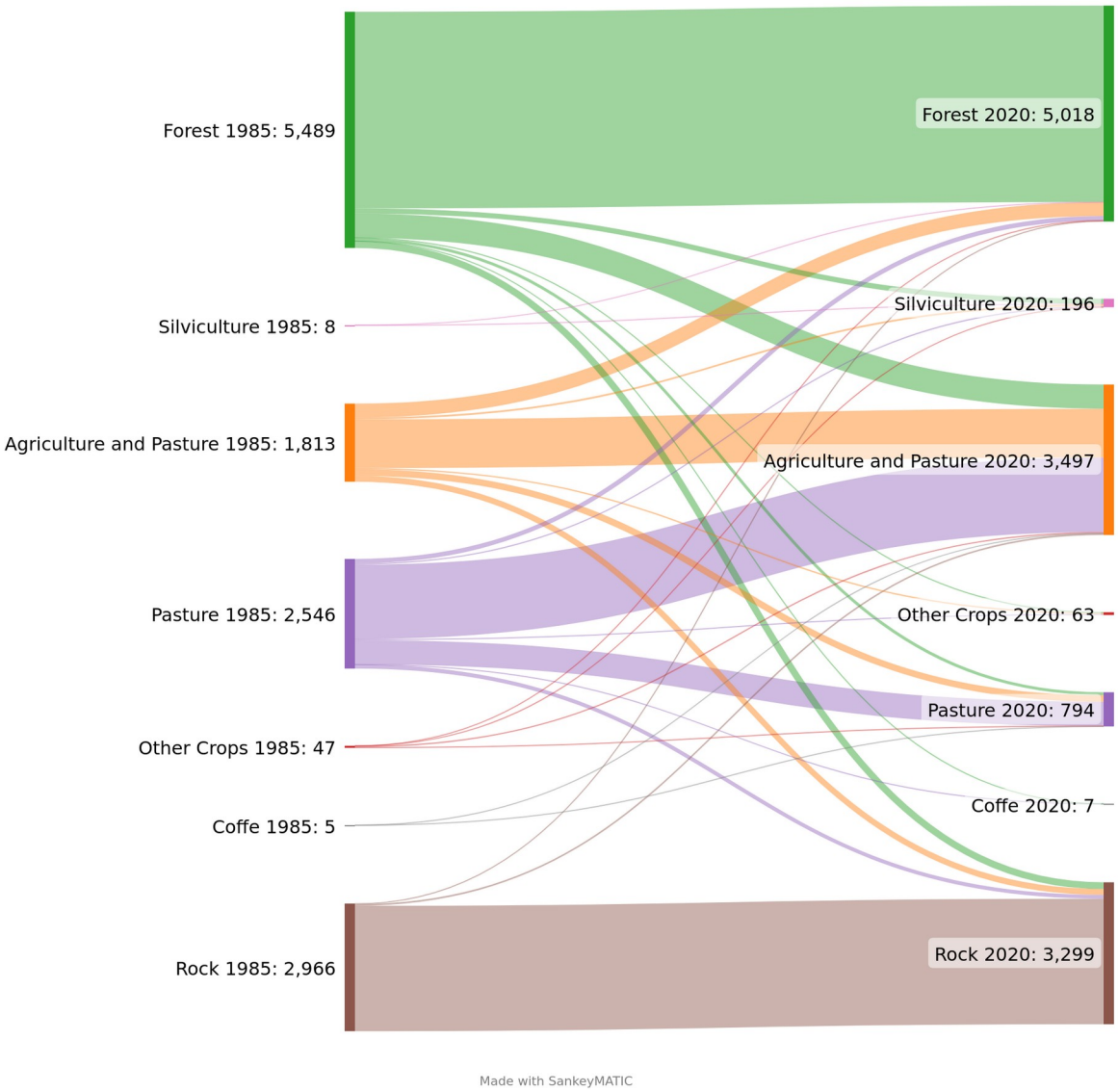
The first gain verified in this strange format of storing all rasters was to cross spatial data to filter the processing, which takes advantage of the spatial indexes of the reference layers and the spatial index of the points of the mapbiomas table. The second and most important is the possibility of making different queries at the level of the pixels of the images, being able to use groupings and mathematical operations and queries from a structured database. While all previous processing took about 20 hours of machine time, querying all records that have not changed in 36 years and grouping pixel counts by land use class across the entire region took 5 minutes. Other test operations proved to be quick and easy to change and reprocess, since years are treated as fields in a table and images as lines.



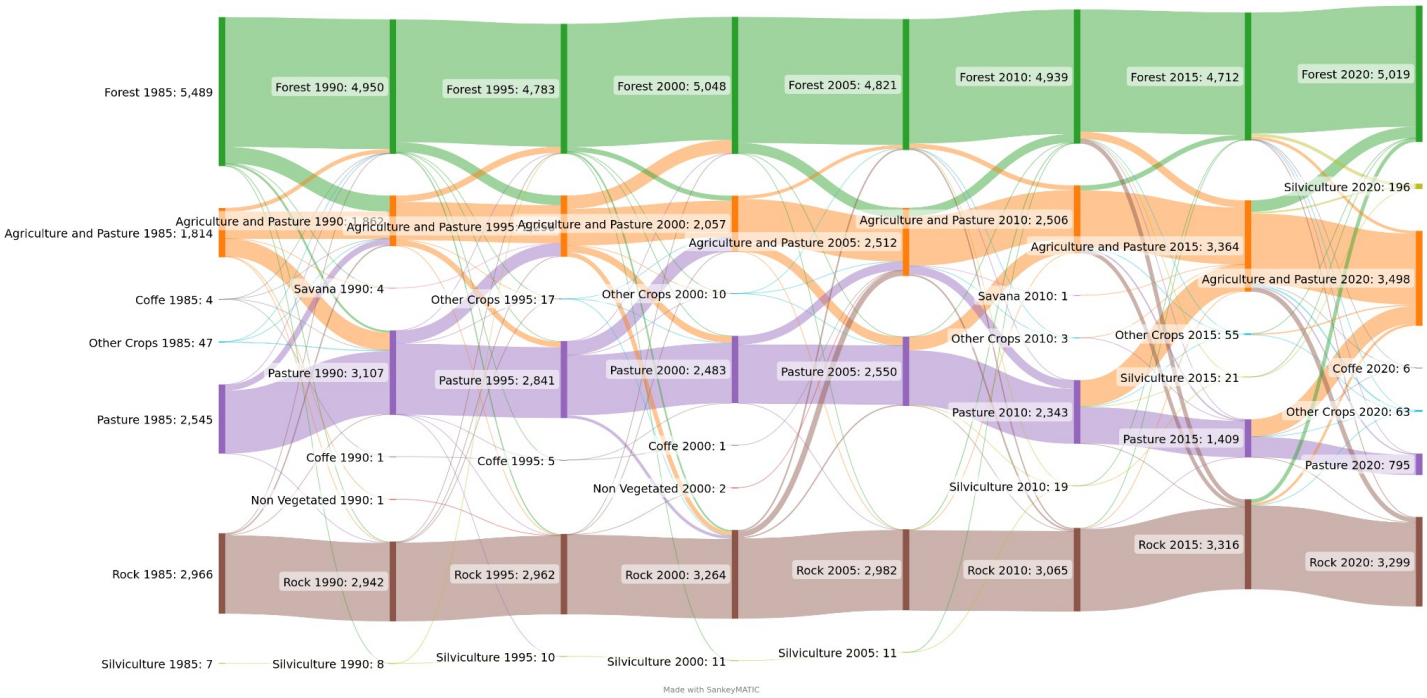
Raster with original classes from mapbiomas with digital numbers extracted and stored in a spatial database, shown here are the 2020 coverage with the 2020 column.

Labels of points from the database superimposed on the google maps image, the original resolution of mapbiomas is 30 meters while the background image has a resolution below 1 meter.

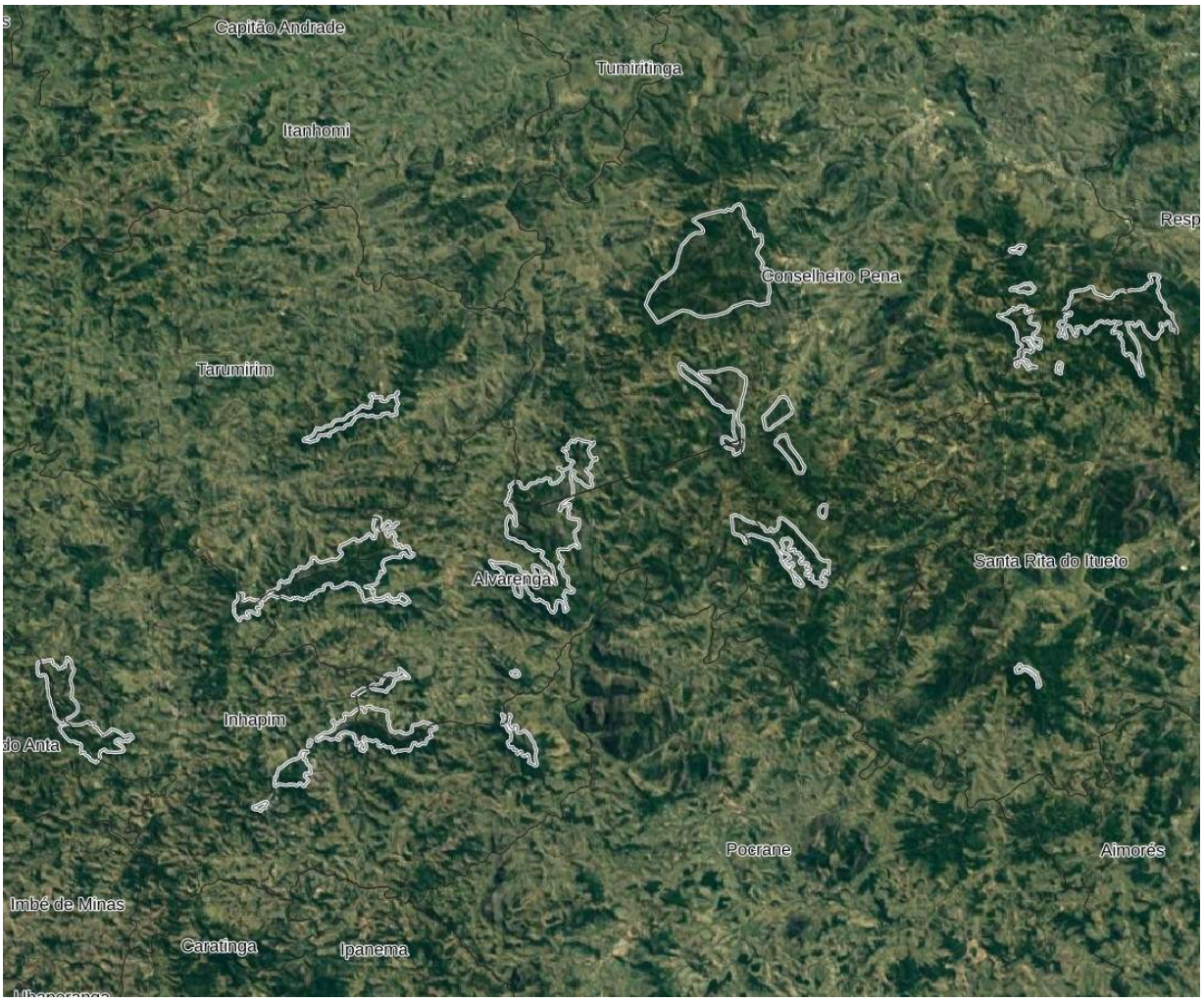
Part 3 - Sankey diagrams for temporal visualization of spatial land use change



Pixel by pixel change of the region, considering 12,851 hectares of areas with 1 or more hectares, out of a total of 14,570 hectares calculated for the set of rocky fields. The main changes were observed 427ha classified as Forest in 2020 out of 4360hs of Agriculture/Pastures and Pastures mosaic in 1985 compared to 631ha that made the opposite way in other points. 383ha of these three classes became rocks and 121ha of forest was converted to forestry. The observed total of natural areas was 8455ha in 1985 and 8317ha in 2020, with 970ha changing from natural to anthropic areas while 675ha were in the opposite direction. From an ecological and biodiversity point of view, despite the relative stability in the total of natural areas with 138ha of net loss of natural areas (1.6% of the original), in reality 1645ha (19.5%) of forest areas were impacted natural suppressed and regeneration in other points.



Class change dynamics between 1985 and 2020 with 5-year intervals between data collection for each pixel. Starting from 5,489ha of forest, it ended up having only 4,783 in 1995 and 4,712 in 2015, which by 2020 had returned to 5,019, closer to the original. The migrations between the Pasture classes and Agriculture/Pasture mosaic are due to the different cultivars and classification of classes by the mapbioma.



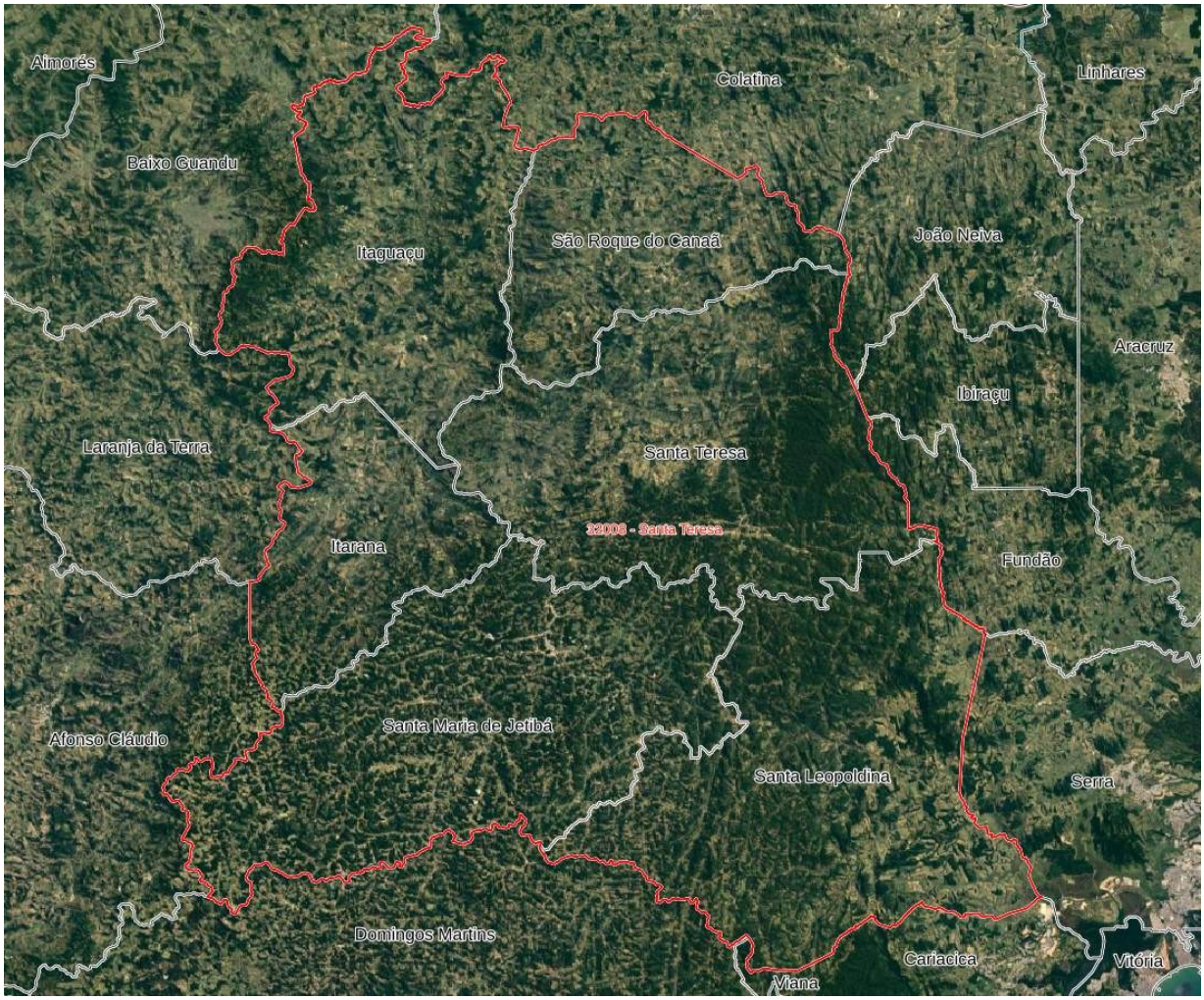
Area of sampled rupestrian fields

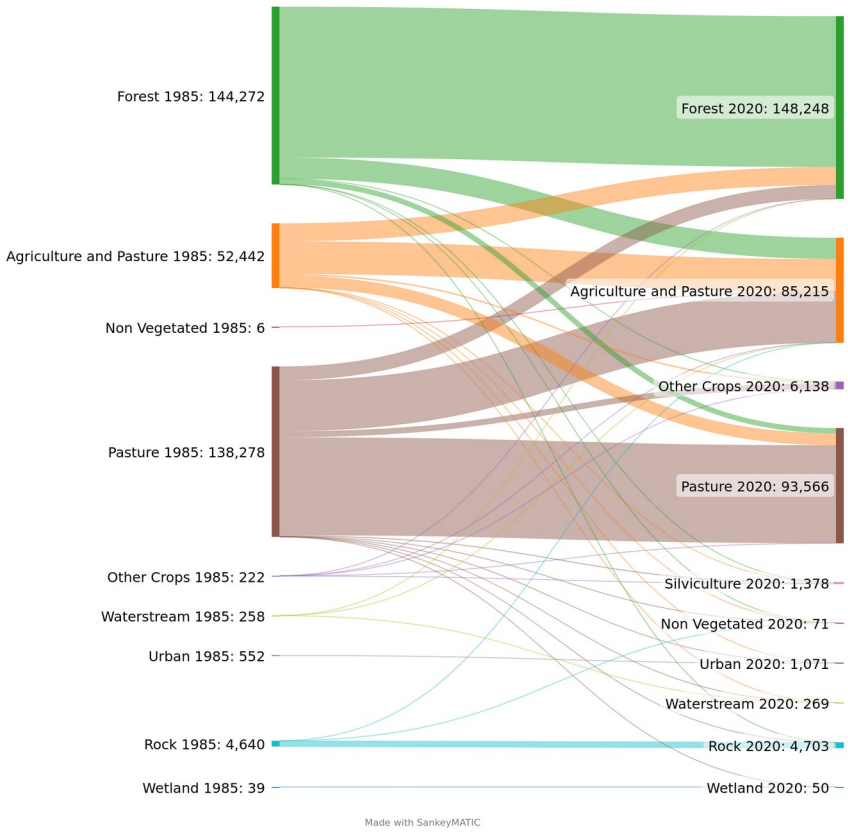
Summary table for the rupestrian areas of the 1985 and 2020 mapbioma classes, according to the count of the same pixels in both data and the observed change.

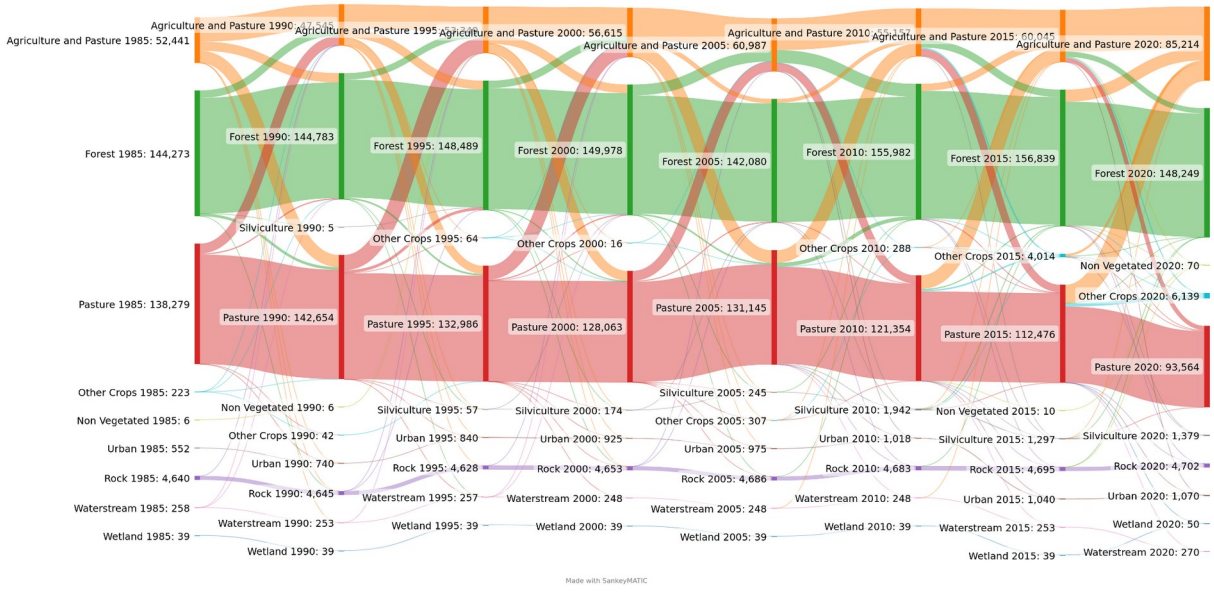
1985	2020	Sum (ha)
Agriculture and Pasture	Agriculture and Pasture	1134
	Forest	**323**
	Pasture	151
	Rock	136
	Silviculture	39
	Other Crops	30
Coffe	Agriculture and Pasture	4
	Pasture	1
Forest	Forest	4569
	Agriculture and Pasture	**563**
	Rock	157
	Silviculture	121
	Pasture	67
	Other Crops	11
	Coffe	1
Other Crops	Agriculture and Pasture	30
	Pasture	11
	Forest	5
	Silviculture	1
Pasture	Agriculture and Pasture	1731
	Pasture	564
	Forest	104
	Rock	90
	Silviculture	29
	Other Crops	22
	Coffe	6
Rock	Rock	2916
	Agriculture and Pasture	35
	Forest	15
Silviculture	Silviculture	6
	Forest	2
Total Result		**12876**

The same process was carried out in the microregion of Santa Teresa (IBGE BC250 2021), which

includes 6 municipalities in ES, including the location of INMA's headquarters. This area is inserted in the initial section of the Atlantic Forest that was converted to the database of points and classes of land use.




Forest 1985: 144,272
Agriculture and Pasture 1985: 52,442
Non Vegetated 1985: 6
Pasture 1985: 138,278
Other Crops 1985: 222
Waterstream 1985: 258
Urban 1985: 552
Rock 1985: 4,640
Wetland 1985: 39
Forest 2020: 148,248
Agriculture and Pasture 2020: 85,215
Other Crops 2020: 6,138
Pasture 2020: 93,566
Silviculture 2020: 1,378
Non Vegetated 2020: 71
Urban 2020: 1,071
Waterstream 2020: 269
Rock 2020: 4,703
Wetland 2020: 50
Made with SankeyMATIC


Agriculture and Pasture 1985: 52,441
Forest 1985: 144,273
Pasture 1985: 138,279
Other Crops 1985: 223
Non Vegetated 1985: 6
Urban 1985: 552
Rock 1985: 4,640
Waterstream 1985: 258
Wetland 1985: 39
Agriculture and Pasture 1990: 47,545
Agriculture and Pasture 1995: 52,249
Agriculture and Pasture 2000: 56,615
Agriculture and Pasture 2005: 60,987
Agriculture and Pasture 2010: 58,142
Agriculture and Pasture 2015: 60,045
Agriculture and Pasture 2020: 85,214
Forest 1990: 144,783
Forest 1995: 148,489
Forest 2000: 149,978
Forest 2005: 142,080
Forest 2010: 155,982
Forest 2015: 156,839
Forest 2020: 148,249
Silviculture 1990: 5
Other Crops 1995: 64
Other Crops 2000: 16
Other Crops 2010: 288
Other Crops 2015: 4,014
Non Vegetated 2020: 70
Other Crops 2020: 6,139
Pasture 1990: 142,654
Pasture 1995: 132,986
Pasture 2000: 128,063
Pasture 2005: 131,145
Pasture 2010: 121,354
Pasture 2015: 112,476
Pasture 2020: 93,564
Non Vegetated 1990: 6
Other Crops 1990: 42
Urban 1990: 740
Rock 1990: 4,645
Waterstream 1990: 253
Wetland 1990: 39
Silviculture 1995: 57
Urban 1995: 840
Rock 1995: 4,628
Waterstream 1995: 257
Wetland 1995: 39
Silviculture 2000: 174
Urban 2000: 925
Rock 2000: 4,653
Waterstream 2000: 248
Wetland 2000: 39
Silviculture 2005: 245
Other Crops 2005: 307
Urban 2005: 975
Rock 2005: 4,686
Waterstream 2005: 248
Wetland 2005: 39
Silviculture 2010: 1,942
Urban 2010: 1,018
Rock 2010: 4,683
Waterstream 2010: 248
Wetland 2010: 39
Non Vegetated 2015: 10
Silviculture 2015: 1,297
Rock 2015: 4,695
Urban 2015: 1,040
Waterstream 2015: 253
Wetland 2015: 39
Silviculture 2020: 1,379
Rock 2020: 4,702
Urban 2020: 1,070
Wetland 2020: 50
Waterstream 2020: 270
Made with SankeyMATIC

Summary table for the Santa Teresa microregion of the 1985 and 2020 mapbioma classes, according to the count of the same pixels in both data and the observed change.

1985	**2020**	Sum - area_ha
Agriculture and Pasture	Agriculture and Pasture	26543
	Forest	14604
	Pasture	9693
	Other Crops	1107
	Silviculture	343
	Urban	141
	Non Vegetated	6
	Waterstream	5
Forest	Forest	122349
	Agriculture and Pasture	17228
	Pasture	4229
	Silviculture	221
	Other Crops	183
	Rock	50
	Non Vegetated	12
Non Vegetated	Agriculture and Pasture	6
Other Crops	Pasture	119
	Agriculture and Pasture	49
	Forest	29
	Other Crops	19
	Silviculture	6
Pasture	Pasture	79525
	Agriculture and Pasture	41342
	Forest	11256
	Other Crops	4829
	Silviculture	808
	Urban	378
	Non Vegetated	47
	Waterstream	46
	Rock	36
	Wetland	11
Rock	Rock	4617
	Agriculture and Pasture	17
	Non Vegetated	6
Urban	Urban	552
Waterstream	Waterstream	218
	Agriculture and Pasture	30
	Forest	10
Wetland	Wetland	39
Total Result		**340709**