ATBD_R

Algorithm Theoretical Base Document & Results

Water - Appendix

Collection 3

Version 1

General coordinator

Carlos Souza Jr

Tecnical coordinator

Juliano Schirmbeck

Team

Amazônia: Bruno Gama Ferreira; Ives Medeiros Brandão; Júlia Gabriela Ferreira Ribeiro; **Caatinga :** Diego Costa; Rodrigo Vasconcelos; Soltan Galano; Washington Franca-
Rocha;

Pampa: Eduardo Vélez Martin; Lucimara Wolfarth Schirmbeck; **Cerrado:** Joaquim José Raposo Puga Pereira; Dhemerson E. Conciani; **Mata Atlântica e Pantanal:** Eduardo Reis Rosa; Mariana Dias; **Plataforma:** Hildebrando Simões de Araújo

Neto; Iago Machado da Conceição Silva;
Rafael dos Reis Azevedo.

Support and management: Julia Shimbo; Marcos Rosa; Tasso Azevedo

Table of contents

List of Figures

1 Introduction

1.1 Scope and content of the document

The main goal of the Mapbiomas Water initiative was to map and monitor surface water in all Brazilian biomes from 1985 to 2023. Since 2020, a group of collaborators organized in the MapBiomas Water Working Group produced a water dataset, based on Landsat data, to improve the detection and monitoring of surface water dynamics in Brazil.

One of the main results is a set of layers with annual water surface data to add up as a cross-cutting theme to be integrated within the land use and land cover maps of the MapBiomas Collection 6.

Beyond the integration with Collection 6, other applications produced by the initiative and derived from the water dataset will be publicly available on a website to improve management and use of water resources in Brazil.

This document presents an overview of the initiative with a focus on the general description of the methodology specifically used to produce the annual water surface data.

1.2 Overview

The MapBiomas Water initiative comes up from a previous study conducted by Imazon and WWF-Brazil conducted in the Brazilian Amazon biome, and perfected to the Alto Paraguay Basin, in the Chaco biome, recently published (Souza et al. 2019). This study showed the possibility of improving the capacity of the MapBiomas Initiative to detect and monitor the surface water dynamics in Brazilian biomes, which corresponds to 12% of the planet’s freshwater reserves, making up 53% of South America’s water resources.

From this background, the MapBiomas Water Working Group expanded the same approach and methodology to all the Brazilian territory. The methodology to map and monitor surface water is based on sub-pixel classification of Landsat 5, 7 and 8 and spatial analysis of surface water bodies to identify anthropogenic and potential climate change impacts.

All data processing was performed within Google Earth Engine to reconstruct a monthly time-series of changes in surface water dynamics in Brazil between 1985 and 2023. Moreover, all the water bodies mapped were classified into natural and anthropic classes. All the main results related to water surface mapping, transitions and temporal tendencies were uploaded to an interactive dashboard for end-users of the surface water dataset.

After the launching of this platform the initiative plans to train end-users, including academy, private sector and government. Reliable information on water availability is an essential tool for decision-makers, helping to prioritize actions to protect and restore and sustainably manage the use of water and freshwater ecosystems.

1.3 Identification of region of interest

The mapping of surface water covers all Brazilian biomes.

1.4 Key Science and Applications

A reliable dataset on water surface dynamics is crucial to improve management and the sound use of water resources. High-quality map-based decisions combined with capacity building at the national level could lead to necessary changes towards sustainable development.

The data and information on water surface mapping can support, among others: integrated territorial planning, monitor the Sustainable Development Goals, water stewardship initiatives, monitor water concessions/small dams, freshwater ecosystem quality assessments, climate change assessment.

2 Overview and Background Information

2.1 Context and Key Informations

Along the last decades human activities have severely altered the conditions of freshwater ecosystems. Drastic changes in land use and land cover, construction of hydroelectric dams, pollution, and overuse of water resources for the production of goods and services have altered water quality and availability worldwide. Recent evidence shows that freshwater species have extinction rates twice as high as terrestrial ones. Moreover, extreme droughts and flooding events related to climate change, have augmented the pressure on water reservoirs and aquatic ecosystems.

This scenary tends to get even worse, given the increasing global population and the growth of markets, and unless integrated water management strategies are developed it will be impossible to achieve global sustainable development goals. In this perspective, continuously and historically assessing changes in water surface dynamics on continental scales is one of the major challenges in making decisions about this precious resource (Oliveira & Souza, 2019).

These same challenges applies in the context of Brazil, where there is the highest per capita proportion of water on the planet, but its distribution and quality are not homogeneous. It implies the need for specific decision making considering the different regional characteristics and the interconnected and cumulative effects of water use. This will only be possible through detailed and consistent data and information on water surface dynamic.

The novel surface water mapping methodology adopted by the MapBiomas Water initiative has previously allowed for the identification and quantification of the reduction of freshwater surface in the Amazon biome, especially in wetlands (Souza et al. 2019). These results has been corroborated by another study from NASA-JPL pointing out that the atmosphere water vapor in the Amazon Basin is reducing as well.

2.2 Historical Perspective: Existent Maps and Mapping Initiatives

The use of satellite data revolutionized the human capacity to map inland surface water and its dynamic. More recently, the combination of free access to Landsat data with cloud computing capabilities allowed the launch of a multi-decadal global surface water dataset - the Global Surface Water (GSW) (Donchyts et al. 2016, Pekel et al. 2016). This water surface mapping initiative gives information on the extent and dynamics of surface water all over the Earth surface, based on a 30-year analysis of Landsat images at the pixel level, with several scientific and management applications. However, direct use of GWSD at a country-level remains challenging for several reasons including: some methodological constraints in detecting water in floodplains, wetlands and small water bodies, lack of a rigorous validation assessment at the county level, and lack of near real-time information.

The MapBiomas Water seeks to overcome some of these limitations adopting the same general approach of combining Landsat data with cloud computing capabilities, but adding some methodological inovations to improve the surface water detections and mapping. Particularly the initiative adopts a Surface Water Subpixel Classifier (SWSC), previously applied to the Brazilian Amazon biome (Souza et al. 2019). In the next sections, details of this metohodology are presented.

3 Algorithm Descriptions, Assumptions, and Approaches

3.1 Algorithm description

The combination of the Landsat Data Archive with the cloud computing facilities provided by Google Earth Engine allowed the MapBiomes Water initiative to produce the first Brazilian Surface Water Dataset (BSWD). **Figure 1** shows the main methodological steps encompassing a Surface Water Subpixel Classifier (SWSC), decision trees, and post-classification procedures to generate annual and monthly surface water datasets.

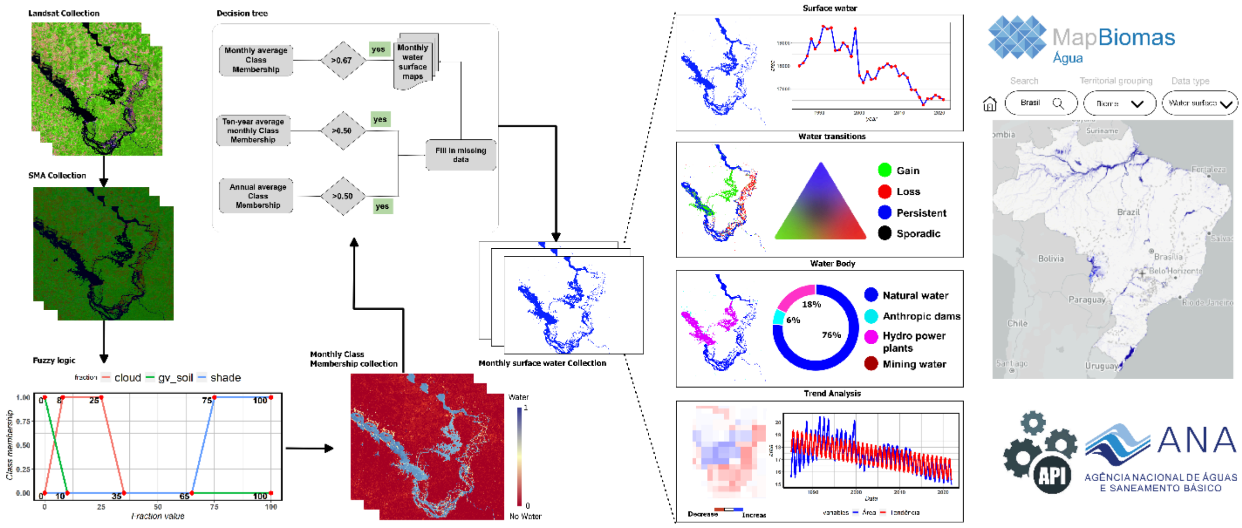

Figure 1. Methodological steps to produce the Brazilian Surface Water Dataset (BSWD), 1985-2023.

3.1.1 Landsat dataset

The project used the Landsat Data Archive (LDA) available in the Google Earth Engine platform. It includes Thematic Mapper (TM), Enhanced Thematic Mapper Plus (ETM+), and the Operational Land Imager (OLI) Landsat sensors, on board Landsat 5, Landsat 7 and Landsat 8, respectively. With a pixel spatial resolution of 30 m, we used the orthorectified Landsat surface reflectance Collection 1 Tier 1. All Landsat path-row scenes covering Brazil (n = 381) during 1985 and 2023 were used and their metadata filtered to select the scenes with cloud cover < 70%.

3.1.2 Spectral Mixture Analysis

The Spectral Mixture Analysis (SMA) allows for the estimation of pixel's fractional composition pure pixels (i.e., endmembers) of green vegetation (GV), non-photosynthetic vegetation (NPV), soil, cloud and shade. The sub-pixel information obtained with SMA is useful to characterized mixed surface water with GV and soil overcoming the limitation of whole-pixel classifiers to map wetlands, floodplains, narrow streams and small waterbodies. In this SMA model, a generic Landsat endmember library is used to calculate the percentage GV, soil, NPV, and cloud in each pixel. The SMA model is based on Earth Engine's spectral unmixing algorithm. We used photometric shade (i.e., flat zero reflectance in all bands) and calculated the shade fraction by subtracting the sum of all endmember fractions from 1.

3.1.3 Surface water Sub-pixel Classifier (SWSC)

The original Sub-pixel Surface Water Classifier (SWSC) algorithm uses three hierarchical binary decision (i.e., True, False) rules based on fractional information of Shade, GV, Soil, and Cloud. First, because water absorbs much of the electromagnetic radiation in the visible, near-infrared, and shortwave infrared Landsat bands, SWSC uses Shade fraction image > 65 percent to classify the majority of the Landsat pixels as surface water. Second, GV and Soil fractions combined (i.e., < 10%) to account for mixed surface water with vegetation and soil. The mixed surface water occurs along the water bodies' edges, on narrow streams, and floodplain and wetland ecosystems. Finally, residual Cloud fraction (i.e., < 25%) is included to detect surface water with a high sediment load. This residual Cloud fraction is due to the Cloud and Soil endmember's spectral ambiguity. In fact, the residual Cloud fraction model is due to the spectral response of the Soil endmember in Cloud free pixels.

An empirical assessment of surface water mapped with the hierarchical binary decision rules described above revealed that the threshold still excludes surface water (**Figure 1**). Because of that, we defined transition rules along the fractional thresholds using a set of linear functions. As a result, the original SWSC binary decision was transformed into three independent probability functions of a Landsat pixel to be classified as surface water. We then calculated the average probability to obtain a continuous surface water probability map with values ranging from 0 to 1 (**Figure 1**). Based on these probabilities we classified the Landsat pixels to produce monthly surface water layers.

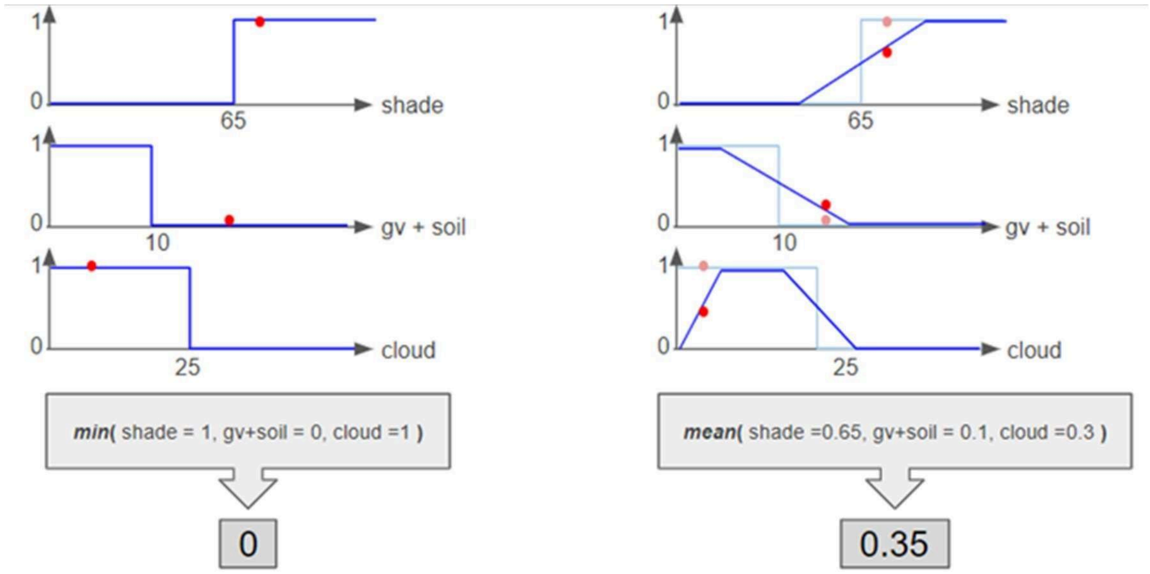

Figure 2. Surface Water Sub-pixel Classifier binary rules. The left example show an example of undetected surface water for a pixel with fraction values indicated by the red points (the value of gv + soil is above the threshold > 10 to be classified as water). The right example applies a linear probability function to variable and the mean value allows to correctly classify the pixel (red point) as surface water.

3.2 Monthly Surface Water Maps

To produce the monthly surface water maps we first produced maximum monthly probability maps, picking the pixels with maximum SWSC probability values from the available Landsat scenes within each month. Then, all pixels with mean probabilities > 0.67 were classified as surface water, resulting in monthly surface water time-series during 1985 and 2023 for all Brazil.

We complemented these monthly maps with a procedure to restore false negatives and other to remove false positives, based on temporal metrics (**Figure 2**). First, we calculated the mean monthly surface water probability for the whole year (i.e., intra-annual mean) and the decadal mean of every month. Then a gap filling was applied to reclassify as water those pixels that were eventually covered by clouds or within areas where no Landsat scenes existed for a given month, using a combination of two rules: mean probability within the year >0.6 and the decadal mean of the correspondent month >0.6. At last, the presence of cloud shadows or other dark objects in the Landsat scene can also produce false-positive surface water classifications. Then, a removal filter was applied to reclassify as no water those pixels with a mean monthly probability <0.35.

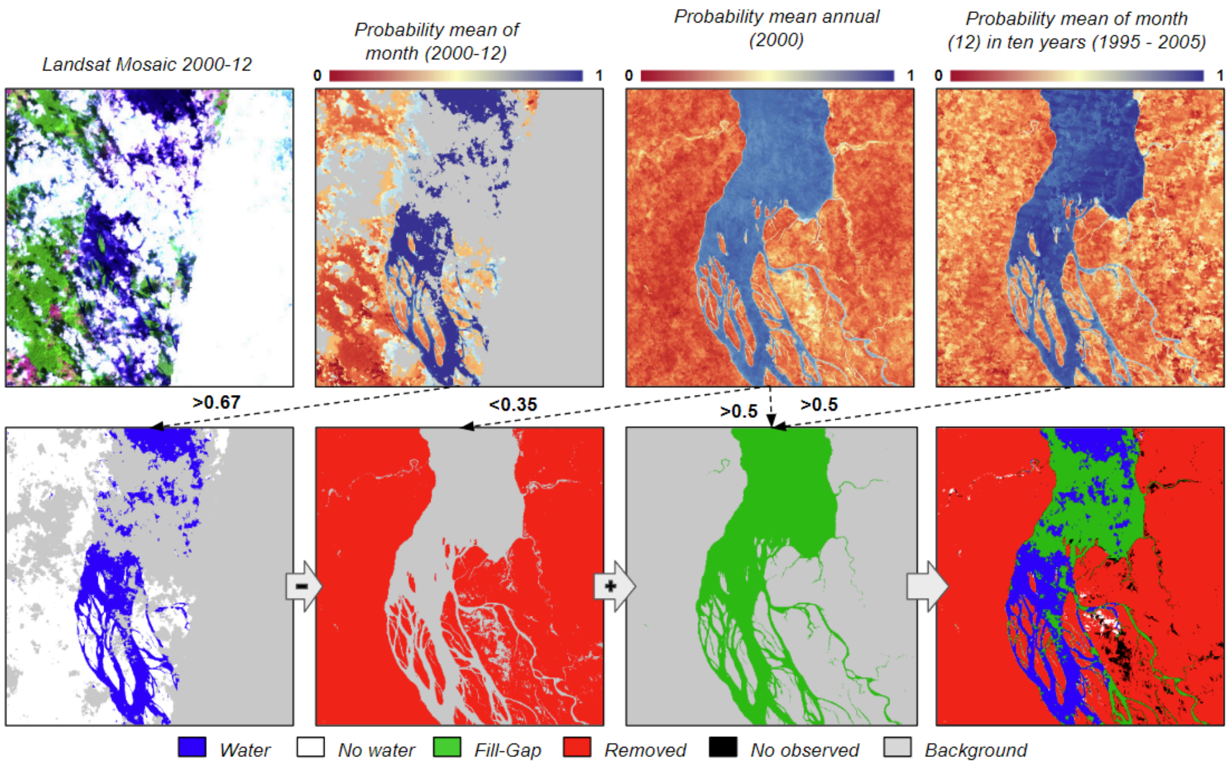

Figure 3. a: Monthly Landsat mosaic; b: Monthly SWSC mean probability; c: Annual Monthly mean SWSC probability; d: Decadal monthly mean SWSC probability; e: Month Surface water classification; f: Total decadal surface water ; g: Area likely to be surface water based on c and d thresholds. h: Final surface water map for the month, with inclusions and removals.

3.3 Annual Surface Water Maps

Two annual surface water maps versions were produced for each year based on thresholds corresponding to the number of months that a pixel was classified as water. The first one considers a frequency > 6 months and the second one, a frequency > 8 months to set a pixel in the annual surface water map. The surface water classifications includes both permanent and temporary water surfaces. The definition of the threshold was biome based. In the Pampa biome, the absence of a dry season justified the selection of a more restrictive threshold to exclude all temporary waters from the annual surface water representation. In the other biomes, the existence of a dry and a wet season justifies a more flexible threshold to encompass all permanent water bodies that are naturally more dynamic.

3.4 Post-classification biomes

We perform a post-classification on the annual and monthly maps according to the needs of each biome. The biome specialist was responsible for indicating the necessary adjustments and carrying out the implementation. Below, we present the needs and adjustments made for each biome:

Amazon:

In 2023, the Amazon suffered a severe drought, especially during the last four months of the year. To guarantee the detection of this event and avoid errors, we implemented the methodology for the Amazon biome in the year 2023 to detect drought using the Sentinel-1 and Sentinel-2 sensors. We used images from these sensors to detect the surface water in the years 2022 and 2023. Our analysis compared monthly average mosaics of these two types of satellite data, eventually identifying areas where there was a loss or gain of surface water in 2023 about the last year. We mapped surface water with Sentinel-2 at the subpixel scale, using spectral mixture analysis (SMA) (Souza Jr. et al., 2005) to estimate the proportion, within a pixel, of the purest spectral responses, that is, the " endmembers" of Green Vegetation, Non-Photosynthetic Vegetation (NPV), Soil, Shade and Cloud.

Due to cloud blocking of Sentinel-2 optical data, we also use Sentinel-1 images to obtain ground observations and detect water surface changes in cloudy conditions. We process the Sentinel-1 VV (VV) polarization bands during the same period analyzed with Sentinel-2, i.e. 2022 (t0) and 2023 (t1), using monthly mosaics and subsequently performing change detection. Areas with changes in Sentinel-1 ≤ -5 dB and Sentinel-2 ≤ -0.3 were defined as surface water loss, while the threshold for surface water gain was in Sentinel-1 ≥ 8 dB and non-Sentinel-2 ≥ 0.67. Based on visual interpretation, we eliminated false detections in the study of the Amazon Biome due to the concentration of sediments that reduced the signals in Sentinel-1 and Sentinel-2. Figure 1 summarizes the methodological steps described above.

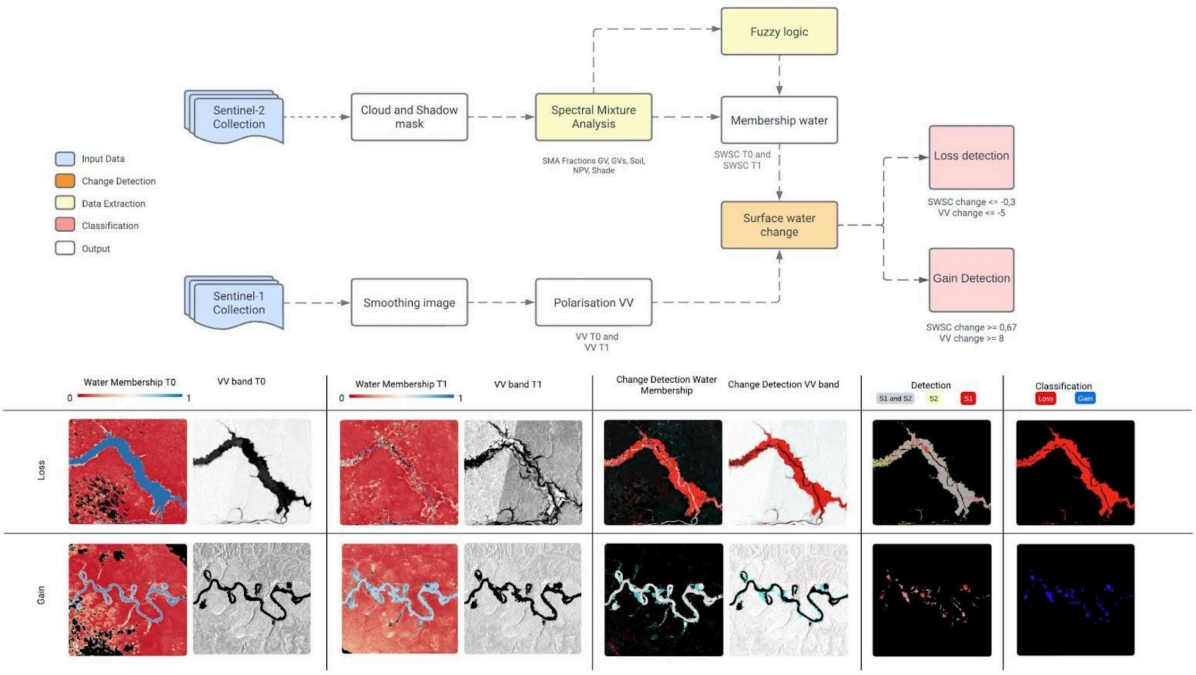

Figure 4. Remote sensing workflow to detect surface water change between 2022 and 2023 (i.e., loss and gain) using Sentinel-1 and Sentinel-2 satellite imagery and examples of surface water change detected with both sensors

Pantanal:

The Pantanal is one of the most extensive continuous wetland areas in the world, characterized by a vast alluvial plain subject to large seasonal floods due to a flood pulse, with pronounced periods of flood and drought. Monitoring this flood pulse is essential to understanding how land use and cover change over time. To improve the mapping of seasonally flooded areas this collection includes an adjustment in the data methodology referring to the Pantanal biome.

Monthly mapping of water surface and flooded fields used supervised classification with the Random Forest algorithm, as described in the ATDB of the Pantanal biome. This classification utilizes the maximum number of available images, creating monthly mosaics for each satellite and data from different Landsat satellites. Some monthly mosaics from Landsat contain pixels with no data due to the lack of images or the cloud mask filter's effect. To overcome this, we developed a gap-filling filter for the classified data. The filter considered the local precipitation pattern to fill each no-data pixel with the classification from the closest and driest available month. We think this a conservative strategy to avoid overestimating the mapping of flooded areas in the Pantanal during periods without images, as the rainiest months are also the cloudiest.

To contribute to the details of the data provided by MapBiomas Water, we selected the monthly water surface data generated by this approach with supervised classification, as well as data relating to the 'flooded field' area that aligns with that mapped by MapBiomas Water, based on a spectral mixture model.

With this monthly data, it was possible to generate annual frequency data. To integrate the annual maps, we consider pixels classified for at least six months of the year. Intra-annual flooding with a frequency of less than six months is available in the “Intermittent water” data.

Cerrado:

The Cerrado is one of the biomes with the highest population density and concentrates a significant portion of Brazil's agricultural commodity production, especially grains. The biome hosts large artificial water reservoirs, mainly for electricity generation, urban supply, and crop irrigation. Considering this context, accurate temporal mapping of reservoirs is essential for understanding the hydrological dynamics in the biome and mitigation of the impact of climate change on human populations.

In evaluating the data produced for the year 2023, we identified potential omissions in arms of large reservoirs, especially those subject to intense intra-annual seasonality or subject to water surface coverage only in atypically rainy years. We identified that part of the omission originates from the spatiotemporal filter to exclude false positives. That happens because these areas are covered by grassy-herbaceous vegetation for part of the year, having a spectral signature of organic matter and undergoing accelerated eutrophication during the rainy months. We have revised the false positive filter, recognizing that some areas may represent water surfaces.
We adopt the methodology to cross-check the false positive data each year. After cross-checking, we selected only false-positive pixels spatially connected with the water surface mapping from the same year. On these pixels, we estimate membership using the same methodology used for water surface mapping by the original. Finally, we apply the following criteria to retain “false positive” pixels and add them as water surfaces in the final map:

i) be spatially connected (queen’s case) to the water surface in the same year
ii) have membership (probability) greater than 50%

Atlantic Forest

In the assessment carried out for the Atlantic Forest biome, we observed mining commissions, urban areas, mangrove areas, and irrigated crops. To resolve this, we apply MapBiomas masks related to these classes to avoid commissions. To prevent the mining mask from omitting bodies of water in this environment, we adjusted the algorithm by adding an “and” logic where the mining mask only works where annual membership is less than or equal to 0,5.

3.5 Water bodies classification

The surface water mapping was used in a classification of water bodies scheme with the classes: 1. Natural, 2. Reservoirs, 3. Hydroelectrics, 4. Mining water. We also included a fifth class of “False positives”, as a by product of the classification to remove some cases of unwanted false positives that remained in the collections of monthly and annual water surface.

The classification of water bodies included the following steps:

1. Vector delimitation of objects
2. Assigning properties to objects
3. Classification of objects based on training samples
4. Temporal filter

The vector delimitation of objects was called annual vectorization once it corresponds to the process of converting the monthly frequency maps of the water surface (raster data) for each year into regular polygons (vector data) within the spatial delimitation of the water bodies.

This procedure was performed with a segmentation tool, in which a particular water body may have been converted into one or more polygons. The *snic* function available on *Google Earth Engine* was used to generate small and relatively regular segments. **Figure 4** shows some examples of the segmentation based on monthly frequency data for a given year.

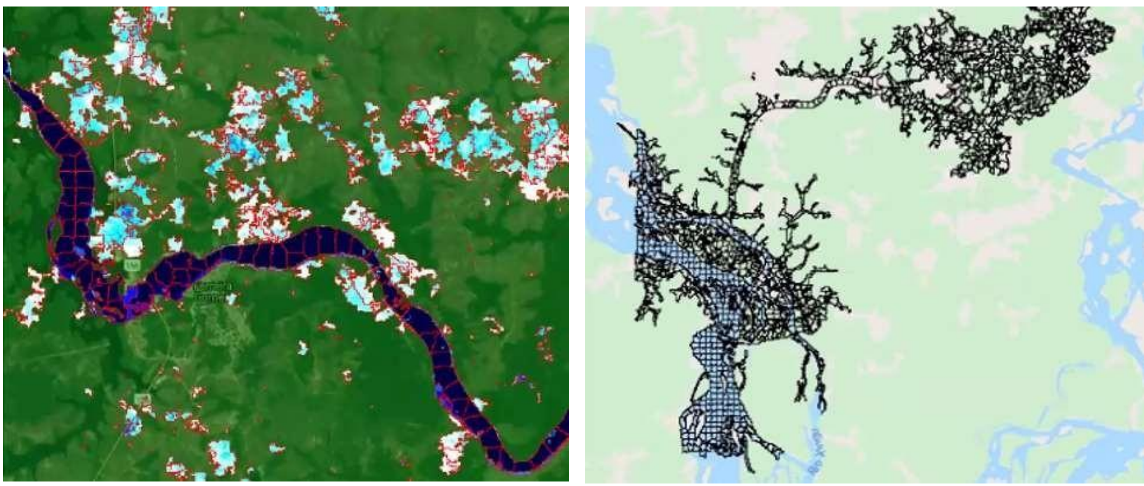

Figure 5. Examples of the segmentation process converting raster data (monthly frequency within each year) into a fishnet of regular vectors.

After generating the objects for each year, a feature dataset with new properties was assigned to each one of these objects for later use in the classification. These properties encompassed informations related to the object morphology, geomorphology and qualitative information from other studies on water bodies classification and land cover and land use mapping. The following variables were associated to each object: area, perimeter, area/perimeter ratio, compactness, roundness, degree of elongation, Laenge- Breite ratio, convexity, maximum extension, number of neighbors, number of neighbors within a 50m buffer, ANA classification – anthropic, ANA classification – hydroelectric, MapBiomas coverage and use class (urban, mining, forest, non-forest class, pasture), maximum SRTM value, mean of total frequency.

The classification of water bodies was performed using the Random Forest algorithm. The training samples were collected in the different biomes for each of the five classes. The samples were collected using a set of grids previously drawn from each biome, covering different years of the time series.

After the classification, the results were submitted to a post classification routine by applying a temporal filter. The time filter logic was to remove improbable transitions between classes of the same segment along the time series.

At last, all the polygons classified as false positives within each year were reconverted to raster format and used to filter the annual and monthly water surface data set, removing the remaining false positives.

4 Validation Strategies

The accuracy analysis of annual water classification data was conducted using MapBiomas Use and Coverage data collected by LAPIG/UFG as a reference (also on an annual scale). The class “River, Lake or Ocean” was considered as a water surface. The sample stratification method based on annual water frequency classes and distance from the next body of water was applied to reduce the sampling error of the producer's accuracy. The water frequency and distance classes used were:

- Permanent: appears more than 90% on the annual map and at least one time in the last three years;
- Intermittent: appears between 50% and 89% of frequency on the annual map;
- Infrequent: appears between 10% and 49% of frequency on the annual map;
- Soil: less than 10% of frequency appears in the time series (1985-2023);
- 250 meters away from nearest water body;
- 500 meters away from nearest water body;
- 5,000 meters away from nearest water body.

The total number of samples per biome, obtained by the sample design described above, is represented in Table 1, and the distribution of samples per stratum is in Table 2.

Biome	Current number of points
Amazon	35.260
Caatinga	9.956
Cerrado	20.851
Atlantic Forest	14.478
Pampa	2.568
Pantanal	2.002

Table 1 - Number of samples (points) per biome.

Stratum	Amazon	Caatinga	Cerrado	Atlantic Forest	Pampa	Pantanal	Total
permanent water	86.684	3.987	7.039	15.880	17.287	3.062	**133.939**
intermittent water	21.771	4.123	5.927	6.044	1.080	5.660	**44.605**
infrequent water	31.614	5.743	8.888	11.329	2.835	12.009	**72.416**
soil	18.437	3.475	5.966	15.836	2.135	10.020	**55.868**
250m meters away from nearest water body	262.046	81.902	163.883	238.056	36.116	43.191	**825.194**
500m meters away from nearest water body	243.169	120.022	241.635	251.339	42.526	26.578	**925.269**
5km meters away from nearest water body	2.056.316	584.423	1.461.881	567.817	92.008	50.116	**4.812.560**
5,000 meters away from nearest water body	1.457.222	59.192	90.103	1.129	68	412	**1.608.125**
Total	**4.177.257**	**862.866**	**1.985.322**	**1.107.430**	**194.054**	**151.047**	**8.477.977**

Table 2 - Area (km²) of each stratum in each biome.

The accuracy of the user and producer (Figure 3) is generally above 75% for the annual mapping of the water surface and can reach up to 90% in some years. The exceptions were the Caatinga since 2010, the Atlantic Forest at the beginning of the time series, and mostly Pantanal, in which the user accuracy was below 50% for all the series. These results are preliminary because the data reference used in this analysis was obtained for mapping the water surface. Divergences can exist, for example, between the dates used to generate reference data (based on the Landsat images) and the water surface annual integrated result, which is not based on a single date of the year. There is a great interannual irregularity in the accuracy of practically every biome, also pointing to seasonal effects in the water surface that were not caught by the reference data. So, the results must be considered as exploratory and preliminary.

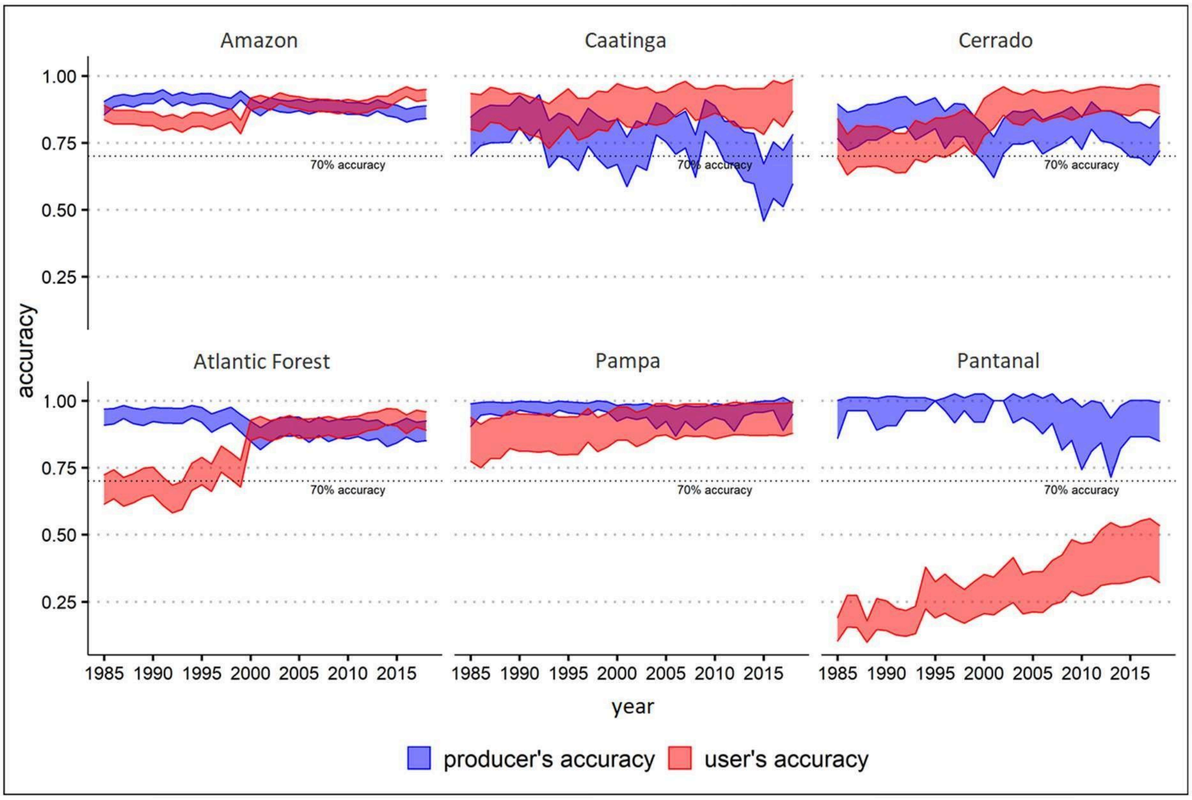

Figure 6. Examples User and producer of “water” class accuracy per biome and per year.

5 Map Collections and Analysis

The main results of collection 1 are publicly available on the web platform: [https://dev-agua.geodatin.com/,](https://dev-agua.geodatin.com/) including four data layers: surface water, transitions, trends and water bodies classification (beta).

5.1 Surface Water Area

This layer corresponds to the annual surface water data. The data presentation includes mapped water surface considering different temporal ranges within the interval 1985-2023, with the correspondent relative frequency. Different spatial limits are also available: biomes, hydrographic regions, municipalities, states of the federation, geographical regions, protected areas and watersheds. According to the temporal and spatial selection made by the user, the map and the area statistics are updated and presented in graphs (**Figure 6**).

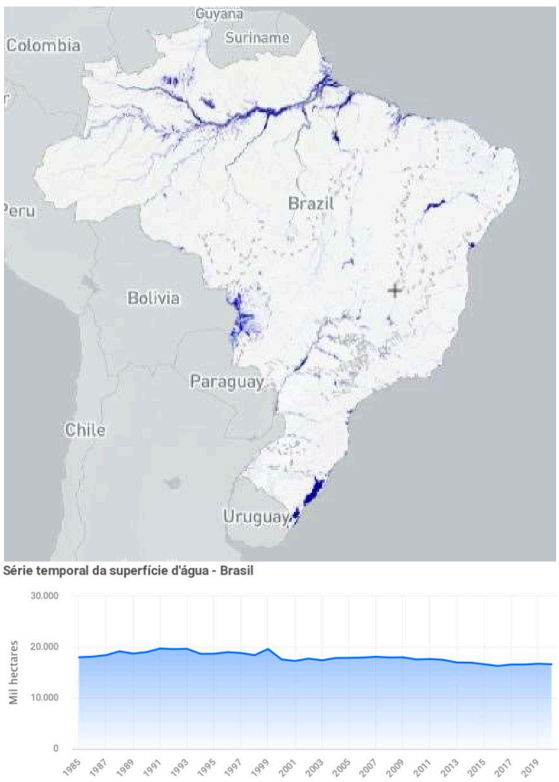

Figure 7. Example of data visualization (map and graph) of the layer Surface Water in the Water MapBiomas platform.

5.2 Surface Water Transitions

This layer shows were the water surface disappeared, appeared, and remained permanent or sporadic along the 36 years. Using the total number of surface water classifications per pixel over the entire annual time-series (i.e., 36 years) it was possible to identify such areas. An RGB color composite was built to facilitate the visual identification of these categories. (**Figure 7**). First, we assigned the total number of years classified as surface water to characterize persistency (blue). Second, we selected the number of years from the

beginning of the time-series until the first classification was surface water to characterize appearance (green). Finally, we chose the number of years from the last surface water observation to the end of the time-series, indicating the disappearance of surface water (red). Accordingly, those permanent water bodies appear predominantly blue, temporary surface water would appear black if they were sporadically reoccurring, lost surface water in red and new ones in green (**Figure 8**).

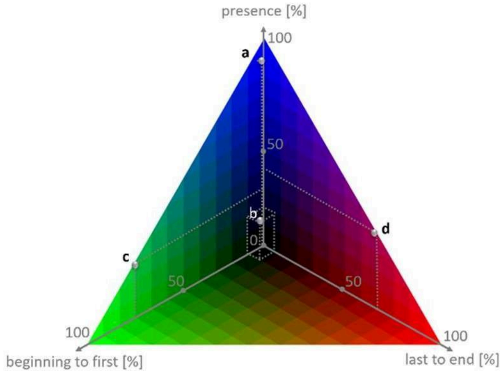

Figure 8. RGB surface water dynamics. Blue colors indicate persistency, as the total number of years classified as surface water; green appearance (the number of years from the beginning of the time-series until the first classification); and Red disappearance. Black and dark colors indicate Sporadic or non-permanent surface water.

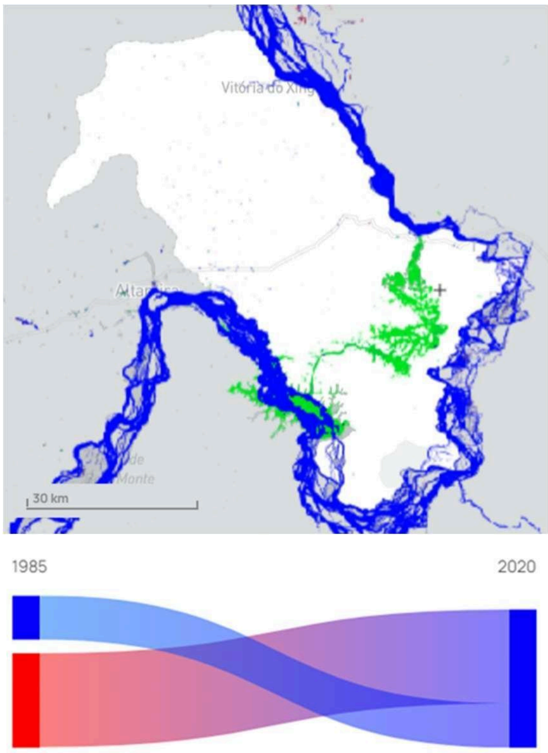

Figure 9. Example of surface water transitions for the Vitória do Xingu municipality. The green in the map indicates the water surfaced the appeared in the last years with the construction of the Belo Monte dam. A graph (red and blue) indicates the transition between the selected years, according to the mapped water surface area.

5.3 Surface Water Trends

The analysis of temporal trends in surface water was performed using the monthly surface water database.

Fitting a harmonic model

To describe and test changes in seasonal variation patterns, concerning changes in phase, amplitude, and variation rate, searching for seasonal pattern and possible trends referring to the monthly water surface mapping data for the 1985-2023 time series, we used the harmonic model described by Shumway & Stoffer(2006) (**Figure 10**).

Let x_1,x_2, ,x_n be a set of n data where x_t represents the value of the variable area in the time series of water bodies, for all t < [1,n] ∈ℵ, in data n = 432 (months). We calculate the harmonic value of each element of the series from the following equation.

$$x_t = \beta_0 + \beta_1 * t + \beta_2 * \cos(2\pi * t) + \beta_3 * \sin(2\pi * t)$$

The estimated coefficients β_0, β_1, β_2, β_3 for the equation were calculated using the ordinary least squares method. (Shumway & Stoffer, 2006).

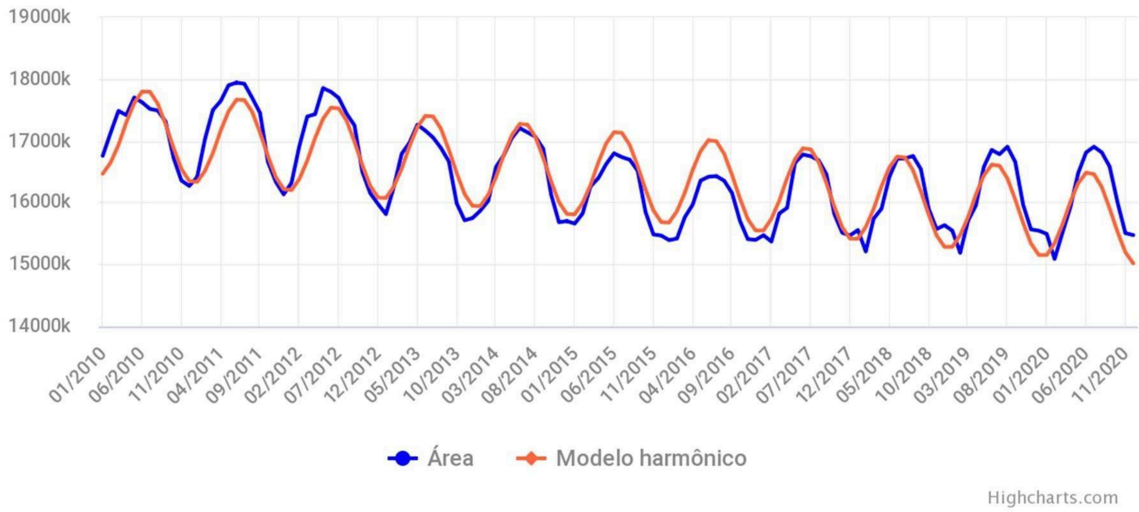

Figure 10. Example of fitting a harmonic series to the results of mapped monthly water surface data. The blue line is the mapped water surface area, the red line indicates the harmonic fitted model.

Differences between the harmonic model and observed data

To detect increased or decreased behavior along the data time series, we calculated the difference between the harmonic model and the mapped water surface area. This analysis aims to point out more clearly which monthly periods showed more intense departures according to the expected by the model (**Figure 11**).

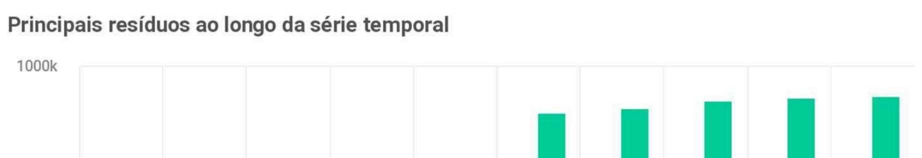

Figure 11. Example of differences between the fitted model and observed data. Green points and bars indicate increased values, red points and bars indicate decreased values in relation to the expected by the harmonic model.

Temporal trends

We searched for trends of increase, decrease or maintenance in the mapped surface water along the 1985-2023 time series. The test was performed using the Mann Kendall Seasonal Test (MK test), which is used to analyze data collected over time for increasing or decreasing trends with monotonic behavior in Y values (**Figure 12**). It is non- parametric, so the data do not need to meet the assumptions of normality, which analyzes data for monotonic trends in seasonal data (Hirsch *et al*., 1982; Hirsch *et al*., 1984; Gilbert, 1987; Helsel & Hirsch, 1995, Morell & Fried, 2009).

To perform the test on the time series, the values are considered an ordered time series. Each value in the series is compared to the rest of the subsequent elements in the series. Kendall's statistic is considered as 0, when S = 0. If the value of S for the whole set of elements is greater than 0, it means that the trend of the series data is decreasing. Otherwise, it means that the series is increasing.

The following steps were considered to calculate the Mann-Kendall statistics:

Let $x_1,x_2,....,x_n$ be a set of n data where x_j represents the value of the variable at time *j* for all *j* < [1,n] ∈ℵ, then the is calculated as:

$$S = \Sigma \quad \sum_{j=k+1}^{n} \quad sign(x_j - x_k)$$

the function is:

$$sign(x_j - x_k) = 1 \qquad se\ x_j - x_k > 0$$

$$sign(x_j - x_k) = 1 \qquad se\ x_j - x_k = 0$$

$$sign(x_j - x_k) = -1 \qquad se\ x_j - x_k < 0$$

Each value-added to S means an increase or a decrease of the value concerning the subsequent one in the series.

In the time series of area data to collect maps of water bodies, it was necessary to consider the Kendall calculation for a series with seasonal behavior. The series of areas of water bodies corresponds to a time series of 12 months over the 36 years. Given this structure for the calculation of the S statistic, the data were divided into 12 subsets, where the first subset corresponds to all values corresponding to the month of January in the series, the second to all values corresponding to February in the series, and so on, until December. Then the value of S will be the sum of all S_j where J = [1.12] (Helsel & Hirsch, 1995).

$$S = \sum_{j=1}^{12} S_j$$

Kendall's value for each month is calculated as described above.

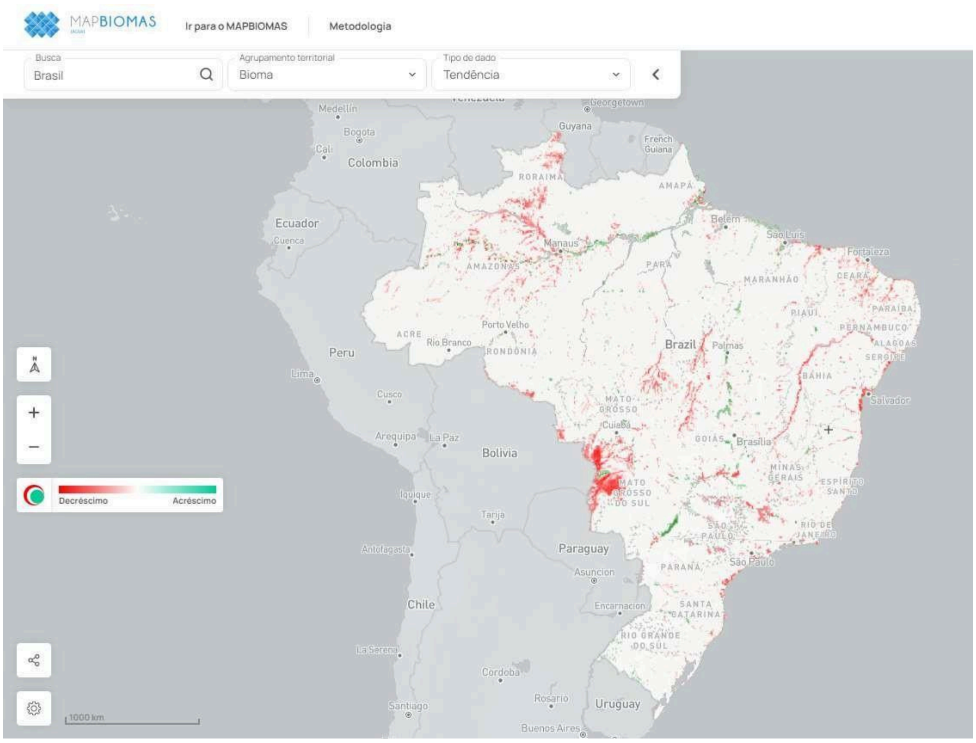

Figure 12. Example of trend calculation results for Brazil using Mann Kendall's seasonal test. Data in red indicate a decrease, data in green an increase and data in white, stability. Only data with significant values are presented.

5.4 Water bodies classification (BETA)

This layer shows the results of the classification of water bodies, including four classes:

1. Natural, 2. Reservoirs, 3. Hydroelectrics, 4. Mining water. The classification is annual, so it is possible to detect the appearance of anthropic water bodies along the time series (1985-2023). The results of the classification are promising, but still require some adjustment. This is because we consider this layer as a beta version. The available statistics include the area of water surface in each one of the water bodies class mapped for the year of interest and considering the spatial unit selected in the interface and a graph with the areas along all the time series.

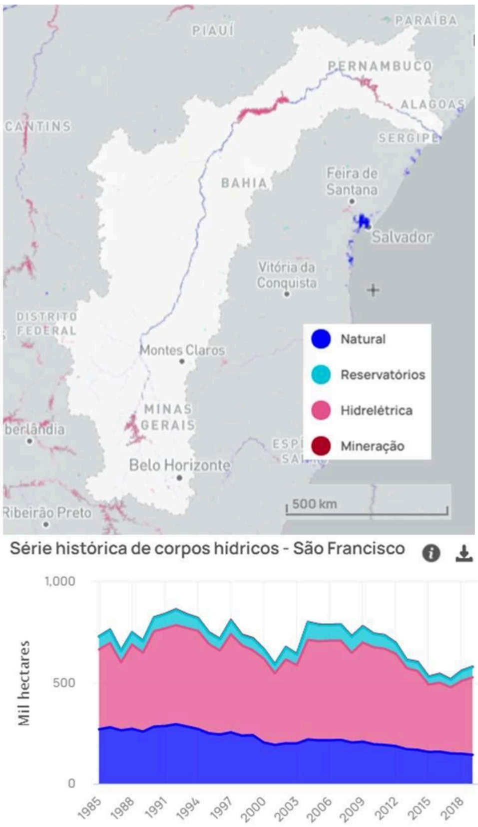

Figure 13. Example of water bodies classification for the São Francisco hydrographic region. The map shows hydroelectrics interrupting the natural water flux of the river in three different sectors. The graph show the area mapped as surface water among the different classes of water bodies.

6 Practical Considerations

Once, this is the first version of the product, it has to be considered as a starting point for future improvements and corrections for any imperfections. All data are publicly available and it is expected that multiple users can point out the hits, the need of corrections and suggestions so that the product can be improved in future versions.

The users of the database have to take into account that the use and applications of the quantitative data presented must always be confronted with the accuracy results in order to understand the degree of existing uncertainty in the data, and decide if is acceptable for the intended use.

7 Concluding Remarks and Perspectives

This collection represents an important advance in the mapping of the water surface in Brazil and for the understanding of its temporal dynamics. Within the scope of the MapBiomas initiative, the results achieved have substantially improved the mapping of the water class presented in the new Collection 6. The combination of data on temporal dynamics with the classification of water bodies is a crucial information to split the natural patterns from patterns resulting from the human action and to better understand their consequences.

8 References

Donchyts, G.; Baart, F.; Winsemius, H.; Gorelick, N.; Kwadijk, J.; van de Giesen, N. Earth's surface water change over the past 30 years. Nat. Clim. Chang. 2016, 6, 810–813.

Gilbert, R.O. 1987.Statistical Methods for Environmental Pollution Monitoring. Wiley, NY.

Helsel, D.R. and R.M. Hirsch. 1995.Statistical Methods in Water Resources. Elsevier, NY, 338-340.

Hirsch, R.M. and J.R. Slack. 1984. A nonparametric trend test for seasonal data with serial dependence. Water Resources Research 20(6):727-732.

Hirsch, R.M., J.R. Slack and R.A. Smith. 1982. Techniques of Trend Analysis for Monthly Water Quality Data. Water Resources Research 18(1):107-121.

Morell and Fried 2009. On Nonparametric Tests for Trend Detection in Seasonal Time Series

Oliveira, B.C.; Souza, C.M. 2019. A novel approach to monitor water in Brazil with satellite images - a concept note. WWF Report 2019.

Pekel, J.-F.; Cottam, A.; Gorelick, N.; Belward, A.S. High-resolution mapping of global surface water and its long-term changes. Nature 2016, 540, 418–422.

Souza, C.M., Kirchhoff, F.T., Oliveira, B.C., Ribeiro, J.G. & Sales, M.H. (2019). Long-Term Annual Surface Water Change in the Brazilian Amazon Biome: Potential Links with Deforestation, Infrastructure Development and Climate Change. Water, 11, 566.