ATBD

Algorithm Theoretical Base Document

MapBiomas Water - Brazil

Collection 5

Version 1 - september 2026

Team

Bruno Gama Ferreira
Dhemerson E. Conciani
Diego Costa
Eduardo Reis Rosa
Fernando Paternost
Helga Correa
Joaquim José Raposo Puga Pereira
Julia Shimbo
Juliano Schirmbeck

Tasso Azevedo

Lucimara Wolfarth Schirmbeck
Marcelo Latuf
Marcos Rosa
Maria Eduarda
Mariana Dias
Natália Crusco
Nerivaldo Afonso
Pedro Galve
Soltan Galano

MapBiomas General Cordination: Tasso Azevedo
MapBiomas water Coordination: Juliano Schirmbeck

“How to cite”

Table of contents

1 Introduction **4**

1.1 Scope and content of the document 4
1.2 Overview 4
1.3 Identification of region of interest 4
1.4 Key Science and Applications 4

2 Overview and Background Information **5**

2.1 Context and Key Informations 5
2.2 Historical Perspective: Existent Maps and Mapping Initiatives 5

3 Algorithm Descriptions, Assumptions, and Approaches **6**

3.1 Algorithm description 6
3.1.1 Landsat dataset 6
3.1.2 Spectral Mixture Analysis 7
3.1.3 Surface water Sub-pixel Classifier (SWSC) 7
3.2 Monthly Surface Water Maps 8
3.3 Annual Surface Water Maps 9
3.4 Post-classification biomes 9
3.5 Water bodies classification 12

4 Validation Strategies **13**

5 Map Collections and Analysis **15**

5.1 Surface Water Area 15
5.2 Surface Water Transitions 16
5.3 Surface Water Trends 17
5.4 Water bodies classification (BETA) 21

6 Practical Considerations **22**

7 Concluding Remarks and Perspectives **22**

8 References **23**

List of Figures

Figure 1. Methodological steps to produce the Brazilian Surface Water Dataset (BSWD), 1985-2024 6

Figure 2. Surface Water Sub-pixel Classifier binary rules 8

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. 9

Figure 4. Remote sensing workflow to detect surface water change between 2022 and 2024 (i.e., loss and gain) using Sentinel-1 and Sentinel-2 satellite imagery and examples of surface water change detected with both sensors 10

Figure 5. Examples of the segmentation process converting raster data (monthly frequency within each year) into a fishnet of regular vectors 12

Figure 6. Examples User and producer of "water" class accuracy per biome and per year 15

Figure 7. Example of data visualization (map and graph) of the layer Surface Water in the Water MapBiomas platform 16

Figure 8. RGB surface water dynamics. Blue colors indicate persistence, 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 16

Figure 9. Example of surface water transitions for the Vitória do Xingu municipality. The green in the map indicates the surface water that 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 surface water area 17

Figure 10. Example of fitting a harmonic series to the results of mapped monthly surface water data. The blue line is the mapped surface water area, the red line indicates the harmonic fitted model 18

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 18

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 20

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 21

Executive Summary

MapBiomas Water Collection 5 provides a comprehensive long-term dataset for mapping and monitoring surface water dynamics across Brazil. The initiative covers the entire Brazilian territory and its six official biomes—Amazon, Atlantic Forest, Caatinga, Cerrado, Pampa, and Pantanal—providing monthly and annual information on surface water occurrence from 1985 to 2025 at 30 m spatial resolution. The dataset was developed to support the understanding of spatial and temporal changes in surface water and to provide reliable information for water resources management, integrated territorial planning, freshwater ecosystem assessment, climate change analysis, monitoring of water-related activities, and other environmental and policy applications.

The core surface water mapping methodology is based on the Landsat Data Archive available in Google Earth Engine, using surface reflectance imagery from Landsat 5, 7, 8, and 9. Surface water detection is performed using Spectral Mixture Analysis (SMA) combined with the Surface Water Sub-pixel Classifier (SWSC). SMA estimates the fractional composition of Green Vegetation (GV), Non-Photosynthetic Vegetation (NPV), Soil, Cloud, and Shade within each Landsat pixel. These sub-pixel fractions are used by the SWSC to estimate the probability of surface water occurrence, improving the detection of mixed pixels associated with waterbody edges, narrow streams, floodplains, wetlands, and other environments where whole-pixel approaches may underestimate surface water.

Monthly surface water maps are generated by selecting the maximum SWSC probability observed among the available Landsat images within each month. Pixels with probabilities greater than 0.67 are initially classified as surface water. The resulting monthly classifications are subsequently refined through temporal procedures designed to recover false negatives associated with cloud cover or missing observations and to remove false positives associated with cloud shadows and other dark targets. These procedures use intra-annual and multi-year temporal information to improve the temporal consistency of the monthly surface water time series.

Annual surface water maps are derived from the monthly classifications according to the number of months in which surface water is detected during each year. Biome-specific thresholds are adopted to account for differences in hydrological regimes and seasonality across Brazil. In Collection 5, pixels detected as water for at least six months are considered annual surface water in the Amazon, Caatinga, Cerrado, and Atlantic Forest biomes, while a threshold of eight months is adopted for the Pampa and a specific criterion is used for the Pantanal to represent its pronounced seasonal flooding dynamics. The annual and monthly products are also subject to biome-specific post-classification procedures designed to address particular mapping conditions and sources of omission and commission errors.

In addition to monthly and annual surface water maps, Collection 5 provides a set of derived products designed to characterize different dimensions of surface water dynamics. These include surface water transitions, annual surface water frequency, permanence, time since the last surface water occurrence, and temporal trends based on the monthly time series. Together, these products enable the identification and analysis of areas characterized by persistence, gain, loss, recurrence, recent or historical occurrence, and

statistically significant temporal changes in surface water.

Collection 5 also includes an annual classification of water bodies, distinguishing natural and anthropogenic water bodies and their respective types. The classification is based on objects derived from the surface water maps and incorporates spatial, temporal, morphological, geomorphological, and contextual attributes. These objects are classified using a Random Forest algorithm and subsequently subjected to temporal consistency rules. Aquaculture information is also incorporated into the water bodies product.

The quality of the annual surface water mapping was assessed using reference information derived from the MapBiomas Land Use and Land Cover validation dataset collected by LAPIG/UFG. The assessment uses a stratified sampling design based on surface water frequency and distance from mapped water bodies, with samples distributed across the six Brazilian biomes. User's and producer's accuracies are generally above 75% and may exceed 90% in some biomes and years, although important spatial and temporal variations occur. The current accuracy results should be interpreted as exploratory and preliminary because temporal differences between the reference observations and the annual integrated surface water product may affect the comparison, particularly in highly seasonal environments.

Compared with the previous collection, Collection 5 extends the historical time series through 2025 and incorporates updates and adjustments to the post-processing filters applied to annual surface water data. The collection also expands the set of derived variables and products available for characterizing surface water dynamics, including annual frequency, permanence, and time since the last occurrence of surface water. In addition, auxiliary datasets used to support the classification of water bodies were updated. These developments expand the analytical capabilities of MapBiomas Water while maintaining the long-term methodological framework based on Landsat imagery, sub-pixel surface water detection, temporal analysis, and biome-specific post-processing.

The Collection 5 products are publicly available through the MapBiomas platform, where users can explore maps and statistics for different temporal periods and territorial units. The platform provides access to the monthly and annual surface water time series and associated products for investigating the spatial and temporal dynamics of surface water throughout Brazil.

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 2025. Since 2020, a group of collaborators organized in the MapBiomas Water Working Group has been producing a water dataset, based on Landsat data, to improve the detection and monitoring of surface water dynamics in Brazil.

One of the main results of this collaborative effort is a set of layers with annual surface water data to add up as a cross-cutting theme to be integrated within the land use and land cover (LULC) maps of the MapBiomas. Beyond the integration with the current LULC Collection, 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 surface water data.

1.2 Overview

The MapBiomas Water initiative comes up from a previous study conducted by Imazon and WWF-Brazil in the Brazilian Amazon biome, which was subsequently adapted and applied to the Alto Paraguay Basin, in the Chaco biome, as documented by Souza et al. (2019). This study showed the possibility of improving the MapBiomas Initiative's capacity to detect and monitor the surface water dynamics in Brazilian biomes, which encompass 12% of the planet's freshwater reserves, and 53% of those in South America.

Based on these findings, the MapBiomas Water Working Group expanded the same approach and methodology to the entire Brazilian territory. The methodology to map and monitor surface water is based on sub-pixel classification of Landsat 5, 7, 8 and 9 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 2025. Moreover, all the water bodies mapped were classified into natural and anthropic classes. All the main results related to surface water mapping, transitions and temporal tendencies were uploaded to an interactive dashboard for end-users of the surface water dataset.

1.3 Identification of region of interest

The mapping of surface water covers the entire Brazilian territory, thus covering all the six official biomes of the country: Amazon, Atlantic Forest, Caatinga, Cerrado, Pampa, and Pantanal. A biome is a geographic region defined based on vegetation types associated with geomorphological and climatic conditions.

1.4 Key Science and Applications

A reliable dataset on surface water dynamics is crucial for improving the management and the use of water resources. High-quality, map-based decision-making, combined with national capacity-building efforts, could lead to necessary changes towards sustainable development.

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

Parallel to data production, the initiative conducts ongoing training for end-users, including academia, private sector and government. Providing reliable information on water availability is an essential tool for decision-makers, helping to prioritize actions for protecting, restoring, and sustainably managing the use of water and freshwater ecosystems.

2 Overview and Background Information

2.1 Context and Key Informations

Over 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 weather events like droughts and floods , have augmented the pressure on water reservoirs and aquatic ecosystems.

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

These same challenges apply to Brazil's context: a country with one of the highest per capita proportion of water on the planet, but with uneven water quality and distribution. It implies the need for specific decision making considering the different regional characteristics and the interconnected and cumulative effects of water use. This is possible through detailed and consistent data and information on surface water dynamics.

The surface water mapping methodology adopted by the MapBiomas Water initiative has previously allowed the identification and quantification of the reduction of fresh surface water in the Amazon biome, especially in wetlands (Souza et al. 2019). These results have been corroborated by another study from NASA-JPL pointing out that the atmospheric 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 dynamics. 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 surface water mapping initiative gives information on the extent and dynamics of surface water across the Earth, 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 country 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 innovations to improve surface water detection and mapping. Particularly the initiative adopts a Surface Water Subpixel Classifier (SWSC), as previously applied to the Brazilian Amazon biome (Souza et al. 2019). In the next sections, details of this methodology 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.

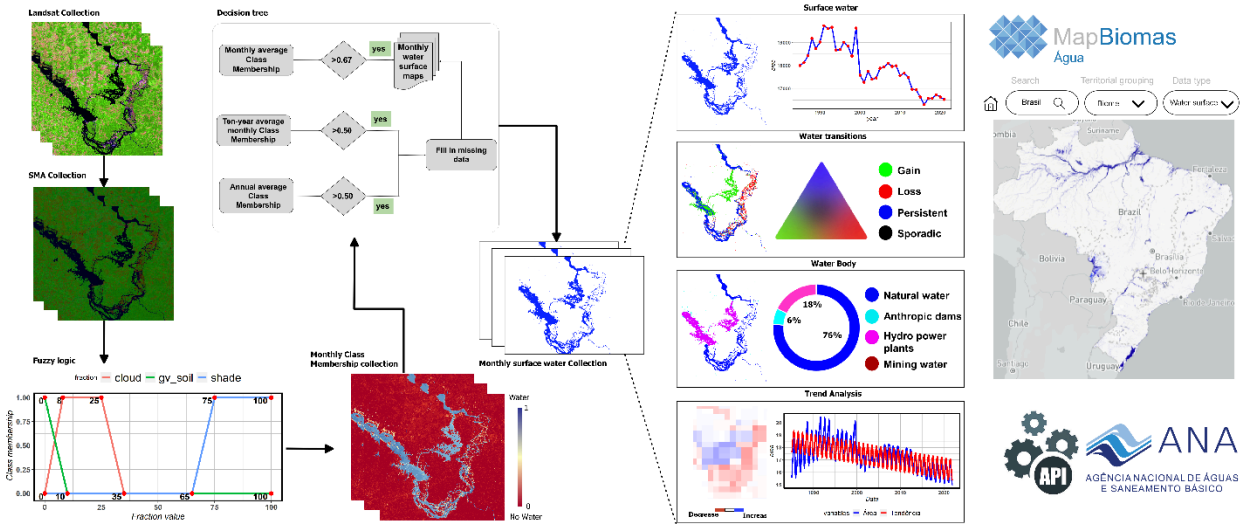


Figure 1. Methodological steps to produce the Brazilian Surface Water Dataset (BSWD), 1985-2025.

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, Landsat 8 and Landsat 9, respectively. With a pixel spatial resolution of 30 m, we used the orthorectified Landsat surface reflectance Collection 2 Tier 1. All Landsat path-row scenes covering Brazil (n = 381) during 1985 and 2025 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 fractional composition of 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 characterize 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% 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 underestimate surface water (**Figure 2**). 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 2**). Based on these probabilities we classified the Landsat pixels to produce monthly surface water layers.

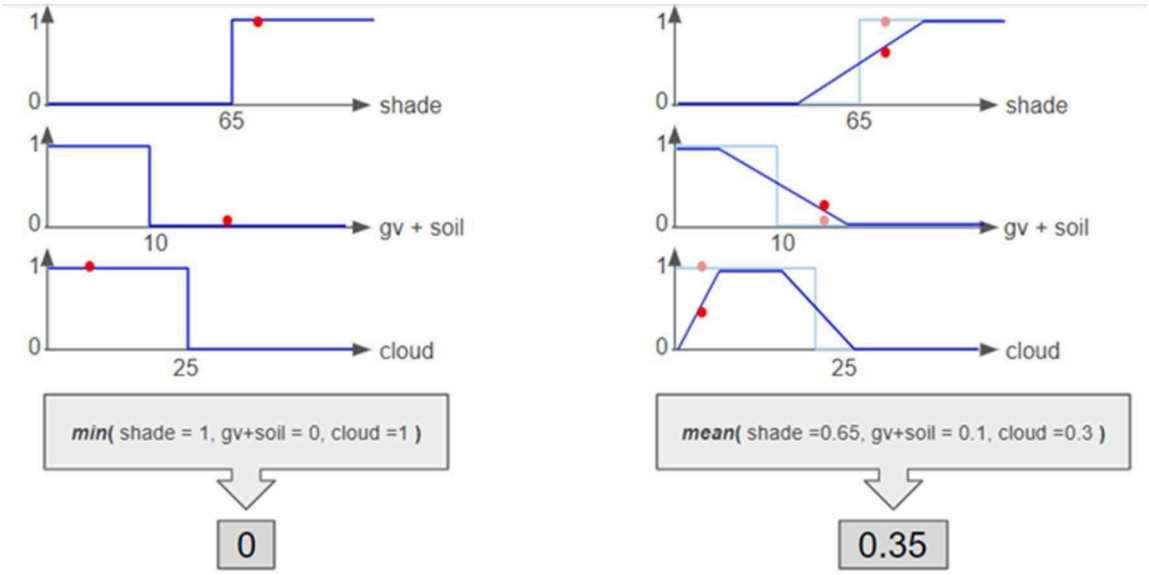


Figure 2. Surface Water Sub-pixel Classifier binary rules. The left example shows an example of undetected surface water for a pixel with fraction values indicated by the red points (the value of gv + soil is above the threshold > 10 to be classified as water). The right example applies a linear probability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

3.2 MapBiomas Água Legends

The legends and corresponding pixel values are presented individually for each MapBiomas Water product, together with their respective methodological descriptions. This approach reflects the diversity of products included in the collection, which represent different dimensions of surface water occurrence and temporal dynamics and, therefore, have distinct data structures, value domains, and interpretation criteria.

3.3 Monthly Surface Water Maps

To produce the monthly surface watermaps 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 2025 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 3**). 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.

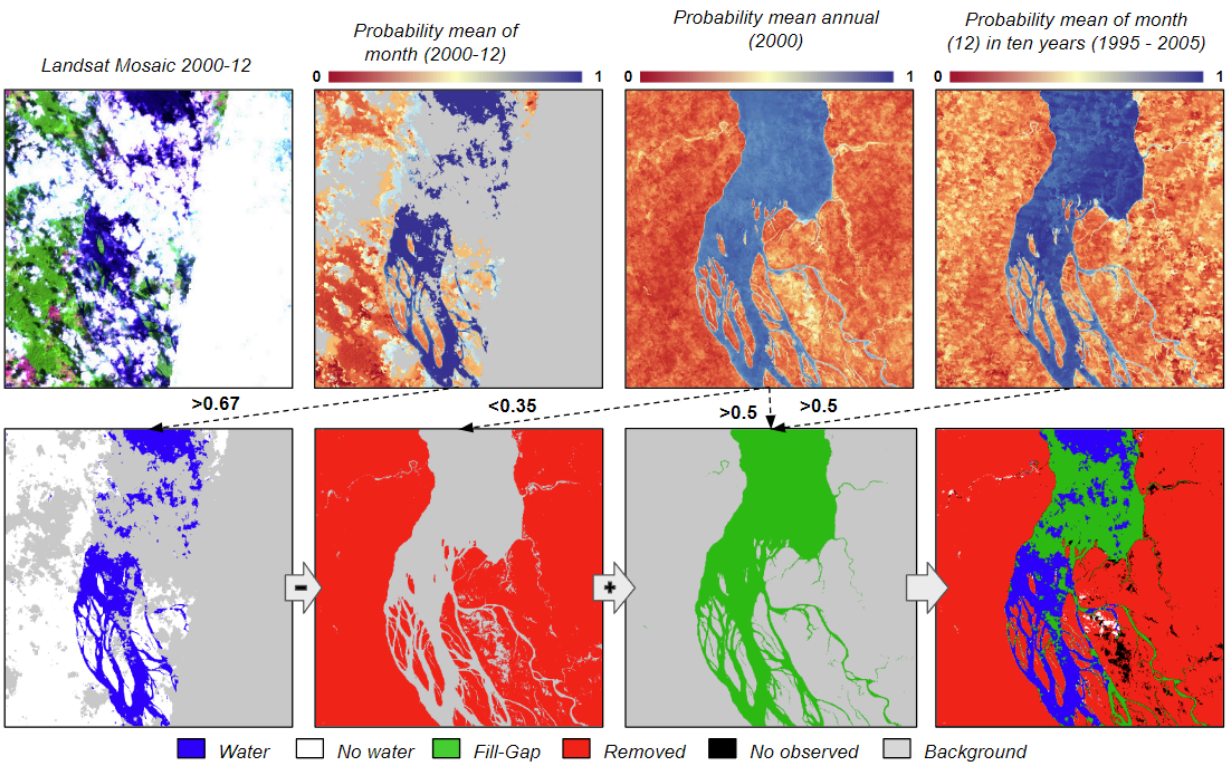


Figure 3. a: Monthly Landsat mosaic; b: Monthly SWSC mean probability; c: Annual Monthly mean SWSC probability; d: Decadal monthly mean SWSC probability; e: Month Surface water classification; f: Total decadal surface water ; g: Area likely to be surface water based on c and d thresholds; h: Final surface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

The Monthly Surface Water product is organized by month, with pixel values ranging from 1 to 12 corresponding to the calendar months from January to December. Table X presents the pixel values used to identify each month in the monthly surface water dataset.

Table X. Monthly Surface Water Maps legend.

Pixel Value	Surface Water Month
1	January
2	February
3	March
4	April
5	May
6	June
7	July
8	August
9	September
10	October
11	November
12	December

3.4 Annual Surface Water Maps

Annual surface water maps were produced for each year by integrating the monthly surface water classifications and applying biome-specific temporal criteria based on the number of months in which each pixel was classified as water. For the Amazon, Caatinga, Cerrado, and Atlantic Forest biomes, pixels classified as surface water for at least six months of the year were included in the annual surface water map. In the Pampa biome, a more restrictive criterion of at least eight months was adopted because of the absence of a well-defined dry season, reducing the inclusion of temporary surface waters in the annual representation. In the Pantanal biome, where seasonal flooding is a dominant characteristic, a specific post-classification procedure is applied as described in Section 3.5.

The Annual Surface Water product represents pixels classified as annual surface water according to the biome-specific temporal criteria described above. Water pixels are represented by a pixel value of 1. Table X presents the class, pixel value, and corresponding color used for visualization.

Table X. Annual Surface Water Maps legend.

Class	Pixel Value	Color (Hexadecimal)	Color
Water	1	#0000FF	

3.5 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 and 2024, 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 e 2024 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 2023/2024. 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 2023/2024 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 surface water 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 2023/2024 (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 4 summarizes the methodological steps described above.

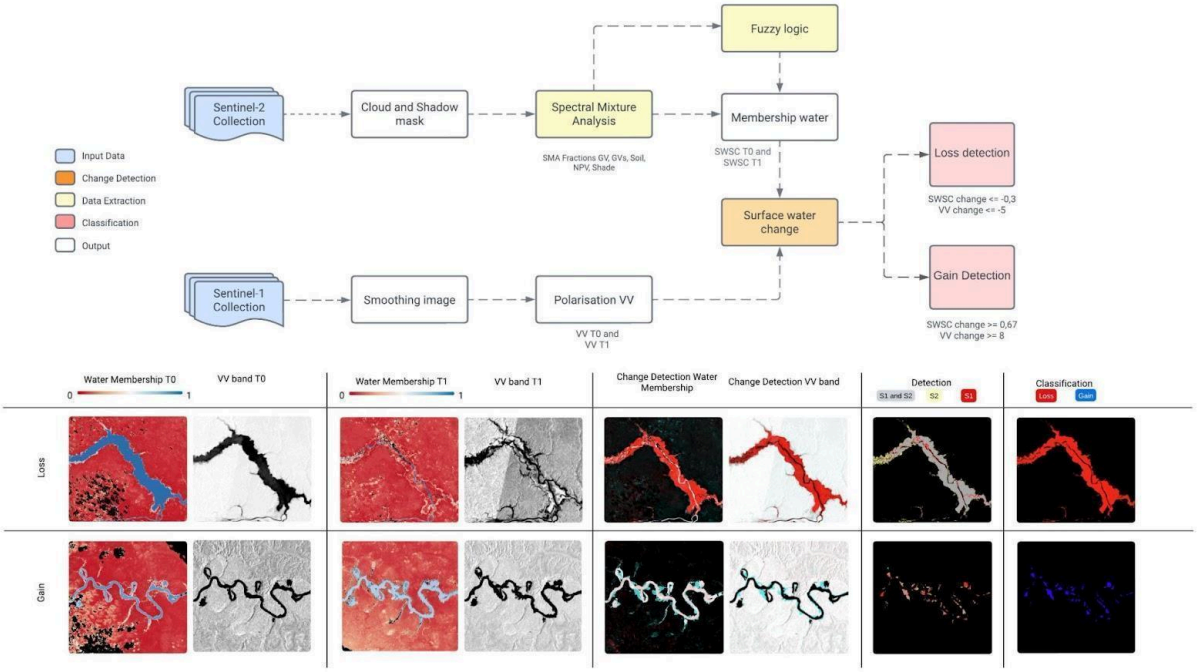


Figure 4. Remote sensing workflow to detect surface water change between 2022 and 2023/2024 (i.e., loss and gain) using Sentinel-1 and Sentinel-2 satellite imagery and examples of surface water change detected with both sensors

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 surface water and flooded fields used supervised classification with the Random Forest algorithm, as described in the ATBD 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 surface water data generated by this approach with supervised classification, as well as data relating to the 'flooded field' or ‘wetland’ 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 twice or more, representing the maximum flooded area in the year. The 'Wetland' data is a complement to the water data and is available only in the land cover and land use 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 and 2024, we identified potential omissions in arms of large reservoirs, especially those subject to intense intra-annual seasonality or subject to surface water 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 surface waters.
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 surface water mapping from the same year. On these pixels, we estimate membership using the same methodology used for surface water mapping by the original. Finally, we apply the following criteria to retain “false positive” pixels and add them as surface waters in the final map:

i) be spatially connected (queen’s case) to the surface water 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.

3.6 Water bodies classification

TThe annual surface water maps were used as the basis for the Water Bodies Classification product. The classification workflow distinguishes natural and anthropogenic water bodies

and uses ancillary spatial datasets and temporal information to identify specific anthropogenic water-body types, including reservoirs, hydroelectric-related water bodies, and aquaculture areas. The final Collection 5 product is organized into the classes Natural, Anthropic, Reservoirs, Hydroelectric, and Aquaculture, according to the legend presented at the end of this section

A false-positive mask is also used as an intermediate component of the classification workflow. This mask identifies areas that may be incorrectly mapped as surface water, such as aquaculture areas, mining-related surface waters, and coastal regions affected by tidal variation. The false-positive information is used to support the refinement and attribution of the final water-body classes and does not constitute a separate class in the final Collection 5 product legend.

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 surface water (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 5** shows some examples of the segmentation based on monthly frequency data for a given year.

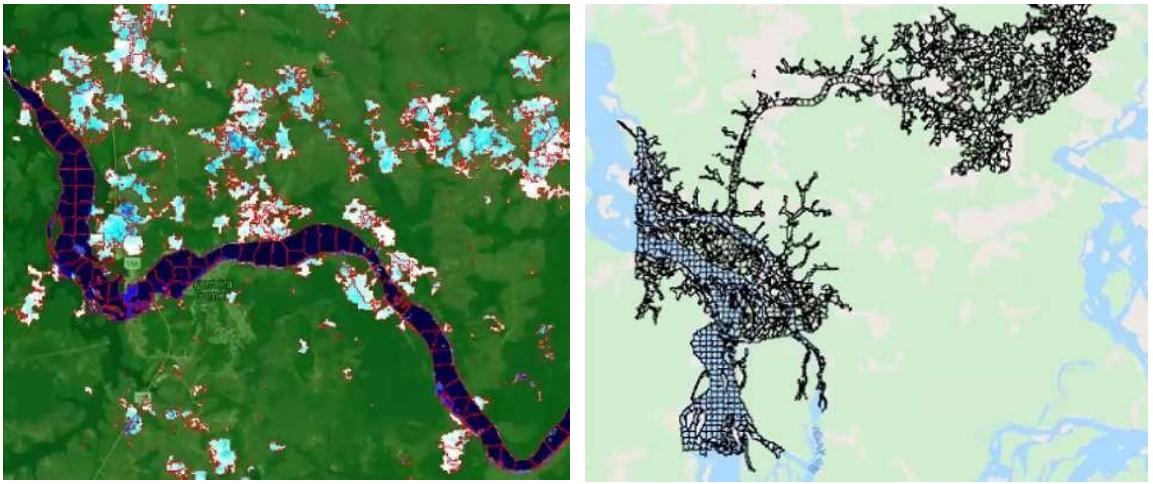


Figure 5. Examples of the segmentation process converting raster data (monthly frequency within each year) into a fishnet of regular vectors.

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 information 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 surface water data set, removing the remaining false positives.

Additionally, we incorporate Aquaculture as another class of water bodies. We used the historical MapBiomas maps produced by the coastal zone theme. The method for generating the aquaculture class is detailed in https://brasil.mapbiomas.org/download-dos-atbds-com-metodo-detalhado/ (Coastal Zone Appendix)

The Water Bodies Classification product distinguishes natural and anthropogenic water bodies and their respective types according to the Collection 5 legend. The classification includes Natural and Anthropic water bodies, with Reservoirs, Hydroelectric, and Aquaculture represented as specific classes in the product. Table X presents the classes, pixel values, and corresponding colors used for visualization.

Table X. Water Bodies Classification legend.

Class	Pixel Value	Color (Hexadecimal)	Color
Natural	1	#0000FF	
Anthropic	2	#00C4D9	
Reservoirs	3	#00C4D9	
Hydroelectric	4	#E6528B	
Aquaculture	6	#091077	

3.7 Derived Surface Water Variables

In addition to the monthly and annual surface water maps, MapBiomas Water Collection 5 includes a set of derived variables designed to characterize different dimensions of surface water dynamics throughout the 1985–2025 time series. These variables are derived from the monthly or annual surface water maps, depending on the temporal process being analyzed, and provide complementary information on changes, recurrence, persistence, and temporal behavior of surface water across Brazil.

The derived variables include Surface Water Transitions, Annual Surface Water Frequency, Surface Water Permanence, Time Since the Last Surface Water Occurrence, and Surface Water Trends. Together, these variables provide different and complementary perspectives on surface water dynamics, ranging from direct comparisons between individual years to long-term patterns observed throughout the historical time series.

3.7.1 Surface Water Transitions

Surface water transitions represent changes in annual surface water occurrence between two selected years. The analysis is performed through a pixel-by-pixel comparison of the annual surface water maps corresponding to the initial (T0) and final (T1) years.

For each pixel, the surface water condition at T0 is compared with its condition at T1, resulting in three transition classes:

Persistence (Water → Water): pixels classified as surface water in both T0 and T1;
Loss (Water → Non-water): pixels classified as surface water in T0 and as non-water in T1;
Gain (Non-water → Water): pixels classified as non-water in T0 and as surface water in T1.

The transition product therefore allows the spatial identification and quantification of surface water gains, losses, and persistence between any two selected years of the historical time series.

Because this analysis compares two discrete annual maps, the transition classes represent the difference between the conditions observed in the selected years and do not describe the surface water behavior during the intervening years.

The Surface Water Transitions product is represented by three classes corresponding to the possible changes in surface water condition between the initial (T0) and final (T1) years: persistence, loss, and gain. Table X presents the transition represented by each class and the corresponding color used for visualization.

Table X. Surface Water Transitions legend.

Class	Transition	Color (Hexadecimal)	Color
Persistence	Water → Water	#2A6FB0	
Loss	Water → Non-water	#BD1830	
Gain	Non-water → Water	#1C7A42	

3.7.2 Annual Surface Water Frequency

Annual surface water frequency represents the number of years in which each pixel was classified as annual surface water throughout the 1985–2025 historical time series.

For each pixel, the binary annual surface water classifications are summed over the entire 41-year period. Therefore, for pixels with at least one occurrence of annual surface water, frequency values range from 1 to 41, where a value of 1 indicates that surface water was mapped in only one year and a value of 41 indicates that surface water was mapped in every year of the time series.

Higher frequency values indicate greater recurrence of annual surface water throughout the historical period, while lower values indicate less frequent or episodic occurrences.

This variable describes the cumulative recurrence of surface water and does not require occurrences to be consecutive. It therefore differs from Surface Water Permanence, which measures continuous occurrence up to the most recent year of the collection.

The Annual Surface Water Frequency product is represented by integer values ranging from 1 to 41, corresponding to the number of years in which each pixel was classified as annual surface water during the 1985–2025 time series. Table X presents the frequency value and the corresponding color used for visualization, with the color scale representing increasing recurrence of surface water throughout the historical period.

Table X. Annual Surface Water Frequency legend.

Surface Water Frequency (years)	Pixel Value	Color (Hexadecimal)	Color
1	1	#03FFFF	
2	2	#03F9FD	
3	3	#03F2FA	
4	4	#03ECF8	
5	5	#03E6F6	
6	6	#03DFF4	
7	7	#03D9F1	
8	8	#02D2EF	
9	9	#02CCED	
10	10	#02C6EA	
11	11	#02BFE8	
12	12	#02B9E6	
13	13	#02B3E3	
14	14	#02ACE1	
15	15	#02A6DF	
16	16	#029FDD	
17	17	#0299DA	
18	18	#0293D8	

Surface Water Frequency (years)	Pixel Value	Color (Hexadecimal)	Color
19	19	#028CD6	
20	20	#0286D3	
21	21	#0280D1	
22	22	#0179CF	
23	23	#0173CC	
24	24	#016CCA	
25	25	#0166C8	
26	26	#0160C6	
27	27	#0159C3	
28	28	#0153C1	
29	29	#014CBF	
30	30	#0146BC	
31	31	#0140BA	
32	32	#0139B8	
33	33	#0133B5	
34	34	#012DB3	
35	35	#0026B1	
36	36	#0020AF	
37	37	#001AAC	
38	38	#0013AA	
39	39	#000DA8	
40	40	#0006A5	
41	41	#0000A3	

3.7.3 Surface Water Permanence

Surface Water Permanence represents the number of consecutive years in which a pixel has remained classified as annual surface water up to the most recent year of the collection, using 2025 as the reference year.

The calculation starts with pixels classified as surface water in 2025 and proceeds backwards through the annual time series. For each pixel, consecutive annual surface water classifications are counted until the first year classified as non-water is encountered. The resulting value therefore represents the duration, in years, of the current uninterrupted period of annual surface water occurrence.

For example, a pixel classified as surface water in 2025, 2024, and 2023, but not in 2022, has a permanence value of 3 years. A pixel classified as surface water continuously throughout the entire 1985–2025 time series has a permanence value of 41 years.

This metric differs conceptually from Annual Surface Water Frequency. Frequency represents the total number of years with surface water occurrence throughout the historical series, regardless of interruptions, whereas permanence represents the duration of the continuous surface water condition ending in 2025.

The Surface Water Permanence product is represented by integer values ranging from 1 to 41, corresponding to the number of consecutive years in which each pixel remained classified as annual surface water up to the reference year 2025. Table X presents the permanence values and the corresponding colors used for visualization, with the color scale representing increasing duration of continuous surface water occurrence.

Table X. Surface Water Permanence legend.

Surface Water Permanence (years)	Pixel Value	Color (Hexadecimal)	Color
1 Year	1	#FFFF80	
2 Years	2	#F6FC74	
3 Years	3	#EBFA69	
4 Years	4	#DEF75E	
5 Years	5	#D2F556	
6 Years	6	#C3F24B	
7 Years	7	#B7F241	
8 Years	8	#A9F037	
9 Years	9	#9BED2F	
10 Years	10	#89EB28	
11 Years	11	#7AE820	
12 Years	12	#68E619	
13 Years	13	#54E312	
14 Years	14	#3DE00B	
15 Years	15	#39DE18	
16 Years	16	#3BD62D	
17 Years	17	#3DD13B	
18 Years	18	#3EC948	
19 Years	19	#3FC453	
20 Years	20	#3CBD5C	
21 Years	21	#3DB868	
22 Years	22	#3BB371	
23 Years	23	#37AD7C	
24 Years	24	#34A885	
25 Years	25	#2FA190	
26 Years	26	#2A9C9A	
27 Years	27	#2196A3	
28 Years	28	#1A90AB	
29 Years	29	#1E84A6	
30 Years	30	#1F79A3	
31 Years	31	#216E9E	
32 Years	32	#22659C	
33 Years	33	#215896	
34 Years	34	#215194	
35 Years	35	#1F458F	
36 Years	36	#1F3E8C	

Surface Water Permanence (years)	Pixel Value	Color (Hexadecimal)	Color
37 Years	37	#1C3487	
38 Years	38	#192B85	
39 Years	39	#162280	
40 Years	40	#11197D	
41 Years	41	#0D1178	

3.7.4 Time Since the Last Surface Water Occurrence

Time Since the Last Surface Water Occurrence represents the number of years elapsed since each pixel was most recently classified as annual surface water, using 2025 as the reference year.

For each pixel, the annual time series is analyzed to identify the most recent year in which surface water was mapped. The number of years between this most recent occurrence and 2025 is then calculated.

A value of 0 therefore indicates pixels classified as surface water in 2025. A value of 1 indicates that the most recent surface water occurrence was in 2024, a value of 2 indicates that the most recent occurrence was in 2023, and progressively higher values indicate increasingly older surface water occurrences.

Considering the 1985–2025 time series, values range from 0 to 40 for pixels with at least one mapped surface water occurrence. This variable makes it possible to distinguish areas with recent surface water occurrence from areas where surface water has not been detected for longer periods.

Together with Annual Surface Water Frequency and Surface Water Permanence, this variable provides complementary information about the temporal history of surface water occurrence at the pixel level.

The Time Since the Last Surface Water Occurrence product is represented by integer values ranging from 0 to 40, corresponding to the number of years elapsed since each pixel was most recently classified as annual surface water, using 2025 as the reference year. A value of 0 represents surface water occurrence in 2025, while increasing values indicate progressively older occurrences. Table X presents the number of years since the last surface water occurrence, the corresponding pixel values, and the colors used for visualization.

Table X. Time Since the Last Surface Water Occurrence legend.

Time Since Last Occurrence (years)	Pixel Value	Color (Hexadecimal)	Color
0 Years	0	#FFFF80	
1 Year	1	#FFFB7D	
2 Years	2	#FFF678	

3 Years	3	#FFF373	
4 Years	4	#FFEE70	
5 Years	5	#FCE96A	
6 Years	6	#FCE365	
7 Years	7	#FCE062	
8 Years	8	#FCDD5D	
9 Years	9	#FCD75B	
10 Years	10	#FAD155	
11 Years	11	#FACD52	
12 Years	12	#FAC94D	
13 Years	13	#FAC54B	
14 Years	14	#F7BF45	
15 Years	15	#F7BE43	
16 Years	16	#F7BA3E	
17 Years	17	#F5B43B	
18 Years	18	#F5AF36	
19 Years	19	#F2AC33	
20 Years	20	#F2A72E	
21 Years	21	#ED9F2B	
22 Years	22	#E69729	
23 Years	23	#DE8E26	
24 Years	24	#D68522	
25 Years	25	#CF7A1F	
26 Years	26	#C9741E	
27 Years	27	#C26C1B	
28 Years	28	#BA6418	
29 Years	29	#B35915	
30 Years	30	#AD5313	
31 Years	31	#A64C11	
32 Years	32	#9E4410	
33 Years	33	#993C0E	
34 Years	34	#91360C	
35 Years	35	#8A2E0A	
36 Years	36	#852708	
37 Years	37	#7D2006	
38 Years	38	#781704	
39 Years	39	#700F02	

40 Years	40	#6B0601	

3.7.5 Surface Water Trends

Surface Water Trends are derived from the monthly surface water time series and are designed to identify long-term patterns of increase, decrease, or stability in surface water occurrence. Unlike the previous derived variables, which are calculated at the pixel level from annual surface water maps, the trend analysis is based on monthly surface water information aggregated into 5 × 5 km territorial grid cells.

The analysis considers the seasonal behavior of surface water throughout the historical time series. A harmonic model is initially fitted to the monthly surface water observations to characterize the expected seasonal pattern and its variation through time. Differences between the fitted model and the observed surface water values provide information on positive and negative deviations from the expected seasonal behavior.

Long-term temporal trends are evaluated using the Seasonal Mann-Kendall test, a non-parametric statistical approach suitable for detecting monotonic trends in time series characterized by seasonal behavior. The monthly observations are organized according to their corresponding month, allowing the trend analysis to account for the seasonal structure of surface water dynamics.

The resulting product identifies statistically significant patterns of surface water increase or decrease, while areas without a significant temporal trend are interpreted as stable. The magnitude and direction of the detected trends are used to represent different levels of surface water increase and decrease.

The Surface Water Trends product is represented by six classes that indicate the direction and magnitude of statistically significant changes in surface water over time. The classes distinguish three levels of decrease and three levels of increase according to the magnitude of the estimated trend. Table X presents the trend classes, their corresponding intervals, and the colors used for visualization.

Table X. Surface Water Trends legend.

Class	Trend Interval	Color (Hexadecimal)	Color
Strong decrease	< -30%	#8C0000	
Medium decrease	-30% to -20%	#FF0000	
Slight decrease	-20% to -10%	#FF8080	
Slight increase	10% to 20%	#C5FF80	
Medium increase	20% to 30%	#80FF00	
Strong increase	> 30%	#008000	

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 surface water. 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% on the annual map;
- Infrequent: appears between 10% and 49% on the annual map;
- Soil: less than 10% frequency in the time series (1985 - 2024);
- 250 meters away from the nearest water body;
- 500 meters away from the nearest water body;
- 5,000 meters away from the 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
Total	**85.115**

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 6) is generally above 75% for the annual mapping of the surface water 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 surface water. Divergences can exist, for example, between the dates used to generate reference data (based on the Landsat images) and the surface water 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 surface water that were not caught by the reference data. So, the results must be considered as exploratory and preliminary.

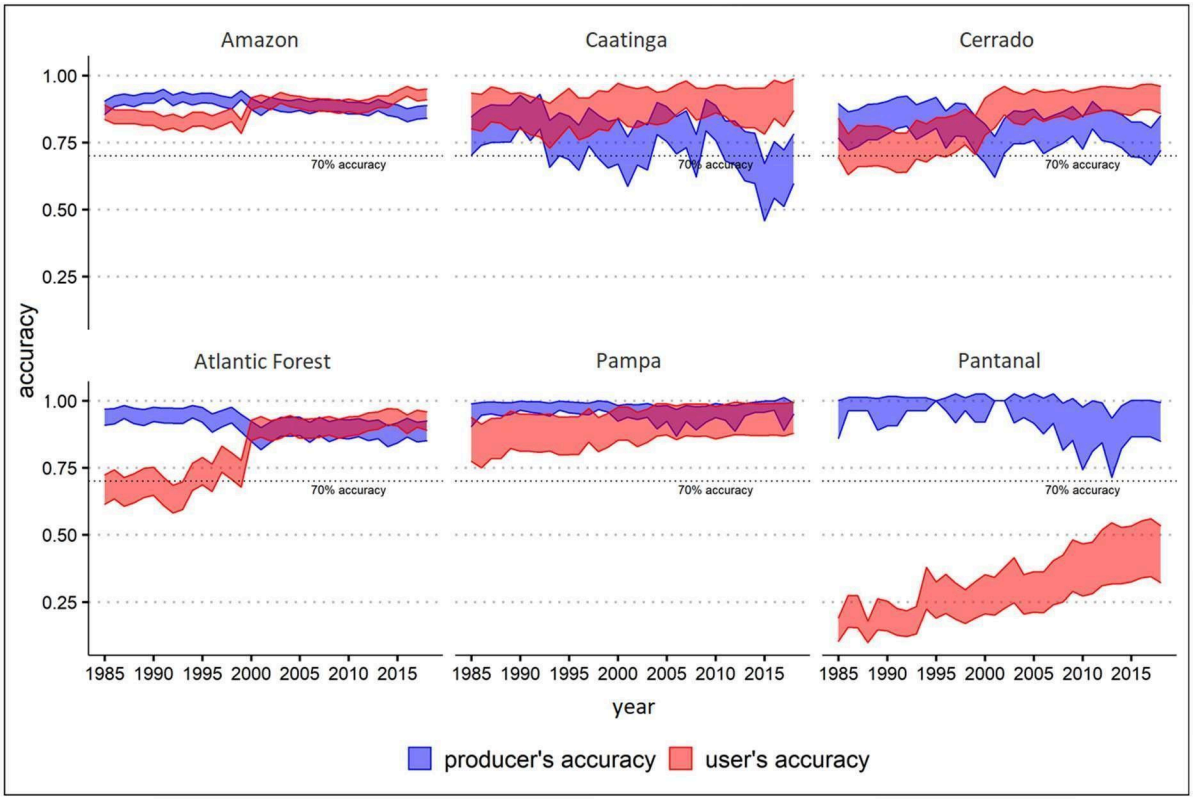


Figure 6. Examples User and producer of "water" class accuracy per biome and per year.

5 Practical Considerations

MapBiomas products are developed through a continuous improvement process, in which each new collection incorporates methodological advances, updated input and ancillary data, improvements in processing and classification procedures, and corrections identified from the evaluation and use of previous collections. Consequently, Collection 5 represents the current stage of development of MapBiomas Water and may be further refined in future collections as new data, methods, and validation information become available. The data are publicly available, allowing users and partners to evaluate the products, identify potential inconsistencies, and provide feedback that can contribute to the continuous improvement of subsequent collections.

6 Concluding Remarks and Perspectives

This collection represents an important advance in mapping surface water in Brazil and understanding its spatial and temporal dynamics. Collection 5 extends the historical time series through 2025 and incorporates updated annual post-processing procedures, new derived variables for characterizing surface water dynamics, and updated ancillary datasets used in water-body classification. These improvements enhance the characterization of changes in surface water occurrence, frequency, permanence, and temporal trends across the country. The integration of information on temporal dynamics with water-body classification provides an important basis for distinguishing natural patterns from those associated with human activities and for improving the understanding of their changes and potential consequences.

7 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.