# Algorithm Theoretical Basis Document (ATBD)

**Collection 4, Version 1**

**Peru team:**

Joaquin Romualdo, Cristhian Vega, Ricardo Valverde

*June 2026*

# Content

## List of Tables



## List of Figures

# 1. Introduction

## 1.1. Scope and content of the document

The purpose of this Algorithm Theoretical Basis Document (ATBD) is to provide users with a detailed guide to the methodological steps, regionalization framework, and threshold justification used in the generation of Peru's surface water maps, corresponding to Collection 4 of the MapBiomas Water Peru initiative. For the glacier module specifically, the methodological procedures are described in Turpo et al. (2022).

This document also presents an overview of the MapBiomas Water initiative, a historical context of the influence of climate change on water resource dynamics, and a review of previous approaches for surface water detection.

## 1.2. General Description

MapBiomas Water is an initiative of the MapBiomas Network whose main objective is to map surface water bodies throughout the Peruvian territory (Figure 1) on a monthly and annual basis, from 1985 to 2025.

This initiative originated from a previous study conducted by Imazon and WWF-Brazil in the Brazilian Amazon biome, later expanded to the Upper Paraguay Basin in the Chaco biome, and published in recent years (Souza et al., 2019). This study demonstrated the potential to improve the capacity of the MapBiomas Land Cover and Land Use Initiative to detect and monitor surface water dynamics across all Brazilian biomes. Based on this context, the MapBiomas Water working group expanded the same approach and methodology to the entire Brazilian territory.

Thus, in 2022, through a partnership between MapBiomas and the Amazon Network of Georeferenced Socio-Environmental Information (RAISG), it was decided to adapt, extend, and apply the MapBiomas Water methodology to the full territories of the Amazonian countries in order to obtain and better understand water dynamics across the different biomes of the region.

The methodology for mapping and monitoring surface water is based on the sub-pixel classification of Landsat 5, 7, 8, and 9 imagery, combined with the spatial analysis of surface water bodies to identify anthropogenic impacts and the potential effects of climate change. All data processing was carried out using the Google Earth Engine cloud-computing platform.

As a result of this initiative, a collection of maps was produced, including monthly and annual surface water extent, transitions, trends, and the classification of surface water into natural and anthropogenic water bodies.

Another important outcome is that the annual surface water maps will be incorporated as a cross-cutting layer into Collection 4 of the MapBiomas Peru Land Cover and Land Use project.

All these data are available through the Water module of the MapBiomas Peru web platform, which is publicly accessible. Following the launch of this platform, the initiative plans to provide training to end users, including the academic sector, the private sector, and government institutions.


**Figure 1.** Peruvian territory.

## 1.3. Science and key applications

The dataset on surface water body dynamics helps improve the understanding of aquatic systems and their interaction with other components of the environment. It is crucial for decision-making processes and contributes to water resource management with a sustainable development perspective.

Information derived from surface water mapping can support integrated territorial planning, monitoring of the Sustainable Development Goals (SDGs), sustainable water management initiatives, monitoring of water concessions and small reservoirs, assessment of freshwater ecosystem quality, and research and evaluation of changes in water bodies and their relationship with climate change.

# 2. General information and background

## 2.1. Context and key information

The condition of many freshwater ecosystems has deteriorated due to human activities over recent decades. Major changes in land use and land cover, the construction of hydropower dams, pollution, and the excessive use of water resources for the production of goods and services have altered water quality and availability worldwide. Recent evidence indicates that freshwater species face extinction rates twice as high as those of terrestrial species (WWF, 2020). In addition, extreme droughts and floods associated with climate change have increased pressure on water reserves and aquatic ecosystems.

This situation is expected to worsen further as the global population continues to grow—surpassing 8 billion people in 2022—and resource consumption increases. Unless integrated water management strategies are developed, achieving the global Sustainable Development Goals will be impossible. From this perspective, the continuous and historical assessment of changes in surface water dynamics at a continental scale represents one of the major challenges for decision-making regarding this valuable resource (Oliveira and Souza, 2019).

These challenges also apply to the Amazonian countries, which contain the largest proportion of freshwater resources per capita on the planet, yet with uneven distribution and variable quality. This creates the need for specific decision-making processes that take into account regional differences and the interconnected and cumulative effects of water use. Such efforts are only possible through the availability of detailed and consistent data and information on surface water dynamics.

The innovative surface water mapping methodology adopted by the MapBiomas Water initiative has previously enabled the identification and quantification of freshwater surface water dynamics in Amazonian countries, particularly in wetlands (Souza et al., 2019). These findings have been corroborated by a NASA-JPL study, which also points to a reduction in atmospheric water vapor across the Amazon Basin (Barkhordarian et al., 2019).

## 2.2. Historical perspective: existing maps and cartographic initiatives

The use of satellite data revolutionized the human capacity to map continental surface waters and their dynamics. The combination of free access to Landsat data and cloud-computing capabilities enabled the development of a multi-decadal global surface water dataset:

Global Surface Water (GSW) (Donchyts et al., 2016; Pekel et al., 2016). This initiative provides information on the extent and dynamics of surface water across the Earth's surface, based on the analysis of more than 30 years of Landsat imagery at the pixel level, supporting a wide range of scientific and management applications. However, the direct use of GSW at the national level remains challenging due to methodological limitations in detecting water in floodplains, wetlands, and small water bodies.

MapBiomas Water seeks to overcome some of these limitations by adopting the same general approach of combining Landsat data with cloud-computing capabilities, while incorporating methodological innovations to improve the detection and mapping of surface waters. In particular, the initiative adopts a Surface Water Subpixel Classifier (SWSC), initially applied in the Brazilian Amazon biome (Souza et al., 2019). Details of this methodology are presented in the following sections.

# 3. Water surface detection methodology

The combination of the Landsat historical image archive and the cloud-processing capabilities provided by the Google Earth Engine platform enabled the MapBiomas Water initiative to produce the first surface water dataset covering all Amazonian countries. Figure 2 illustrates the main methodological steps, which include a Surface Water Subpixel Classifier (SWSC), decision trees, and post-classification procedures used to generate monthly and annual surface water datasets.

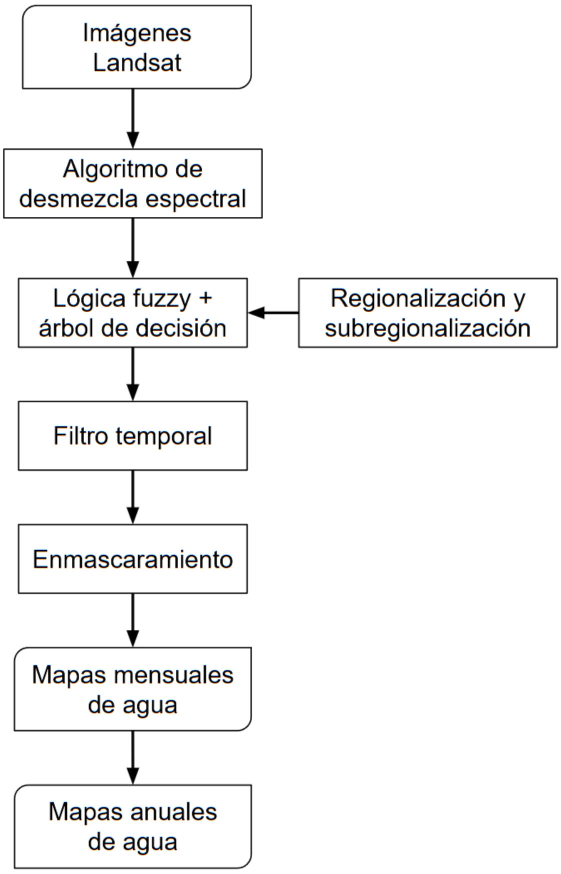


**Figure 2.** Methodological process for producing water surface data, 1985–2025.

## 3.1. Landsat imagery

The project used Landsat satellite imagery, which has a spatial resolution of 30 meters and has been available on the Google Earth Engine platform since 1972. The pixel size and extensive temporal coverage support the objective of the MapBiomas Water initiative to build a historical surface water dataset spanning more than 40 years, making Landsat imagery an appropriate input source for this initiative.

Accordingly, Collection 2 Level-1 Surface Reflectance (SR) imagery, orthorectified to surface reflectance, was used together with data acquired by the Thematic Mapper (TM), Enhanced Thematic Mapper Plus (ETM+), and Operational Land Imager (OLI) sensors onboard the Landsat 5, Landsat 7, Landsat 8, and Landsat 9 satellites, respectively.

To mitigate the effects of cloud cover in the scenes, the following procedures were applied:

- Filtering of scenes with less than 70% cloud cover, based on the "Cloud Cover" metadata available for each image.
- Application of the Cloud Score algorithm using a "Cloud Thresh" threshold of 10, 30, 50, and 70 for the lower Amazon, Andes and upper Amazon, Coast, and Northern Coast regions, respectively.
- Application of masks based on the quality assessment bands ("QA_PIXEL"), selecting bits 3 and 4, corresponding to Cloud and Cloud Shadow, respectively.

### 3.2. Spectral mixture analysis

Spectral Mixture Analysis (SMA) is a technique that estimates the fractional composition of pixels as a combination of endmembers based on the spectral bands of a satellite image. Endmembers are pure spectral signatures of the materials that compose a pixel. The subpixel information obtained through SMA is useful for characterizing surface water mixed with other components, such as soil and vegetation. This overcomes the limitations of traditional full-pixel classifiers and enables the mapping of wetlands, floodplains, narrow rivers, and small water bodies.

For the MapBiomas Water initiative, six endmembers were used: Green Vegetation (GV), Non-Photosynthetic Vegetation (NPV), Soil, Cloud, Shade, and Snow. These endmembers were derived from a generic spectral library for Landsat imagery based on the values of the red, green, and blue bands (visible bands), near-infrared (NIR), shortwave infrared 1 (SWIR1), and shortwave infrared 2 (SWIR2). The spectral values of each band for each endmember are presented in Table 1.

| **Endmembers** | **Bandas** | | | | | |
|---|---|---|---|---|---|---|
| | **Blue** | **Green** | **Red** | **Nir** | **Swir1** | **Swir2** |
| GV | 119 | 475 | 169 | 6250 | 2399 | 675 |
| NPV | 1514 | 1597 | 1421 | 3053 | 7707 | 1975 |
| Soil | 1799 | 2479 | 3158 | 5437 | 7707 | 6646 |
| Cloud | 4031 | 8714 | 7900 | 8989 | 7002 | 6607 |
| Shade* | 7800 | 7910 | 7950 | 6750 | 310 | 380 |
| Snow* | 810 | 650 | 100 | 0 | 0 | 0 |

**Table 1.** Visible and infrared band values for each endmember.

Due to spatial differences between regions, two cases are presented:

i) In the Lower Amazon region, SMA is applied to each pixel using four endmembers to calculate the GV, Soil, NPV, and Cloud fractions. For the shade fraction, photometric shade (zero reflectance in all bands) is used and calculated by subtracting the sum of GV, Soil, Cloud, and NPV from 1.

ii) In the Andes, Coast, and upper Amazon regions, fractions are calculated directly by using six endmembers in the SMA (GV, Soil, NPV, Cloud, Snow, and Shade).

## 3.3. Regionalization and subregionalization

### 3.3.1. Classification regions

The regionalization adopted in Peru (shown in Figure 3) was developed in response to the heterogeneity and unique characteristics of the country's geography. Consequently, the surface water classification methodology applied by the MapBiomas Water initiative was adapted for specific regions.

The delineation of these regions was initially based on the country's major biomes: Coast, Andes, upper Amazon, and lower Amazon. However, spectral heterogeneity among water bodies was still observed even within the same biome (an example is presented in Table 2). Therefore, the regionalization was further refined according to the following criteria:

- Classification regions previously proposed by the MapBiomas Land Cover and Land Use initiative.
- Additional criteria based on the geographic distribution of water bodies.
- Specific criteria related to particular types of water bodies, such as the Northern Coast Shrimp Farms region and the Glacier region.

| Region | Code | Cloud | GV | NPV | Shade | Snow | Soil |
|---|---|---|---|---|---|---|---|
| Central Coast 01 | 1330 | 5.51 | 6.72 | 0.03 | 82.50 | 0.72 | 2.54 |
| Northern Coast Shrimp Farms | 1327 | 2.24 | 4.65 | 0.00 | 88.66 | 1.03 | 1.73 |
| Western Andes 01 | 1317 | 4.47 | 7.61 | 0.01 | 75.42 | 0.12 | 10.17 |
| Andes glaciers | 1315 | 1.06 | 0 | 0.02 | 84.19 | 12.96 | 0.01 |
| Northern High Amazon 03 | 1304 | 1.48 | 12.13 | 0.09 | 81.20 | 0.28 | 3.05 |
| Southern High Amazon 03 | 1309 | 1.20 | 8.35 | 0.01 | 87.91 | 0.13 | 0.87 |

**Table 2.** Example of parameterization of some classification regions in different biomes.

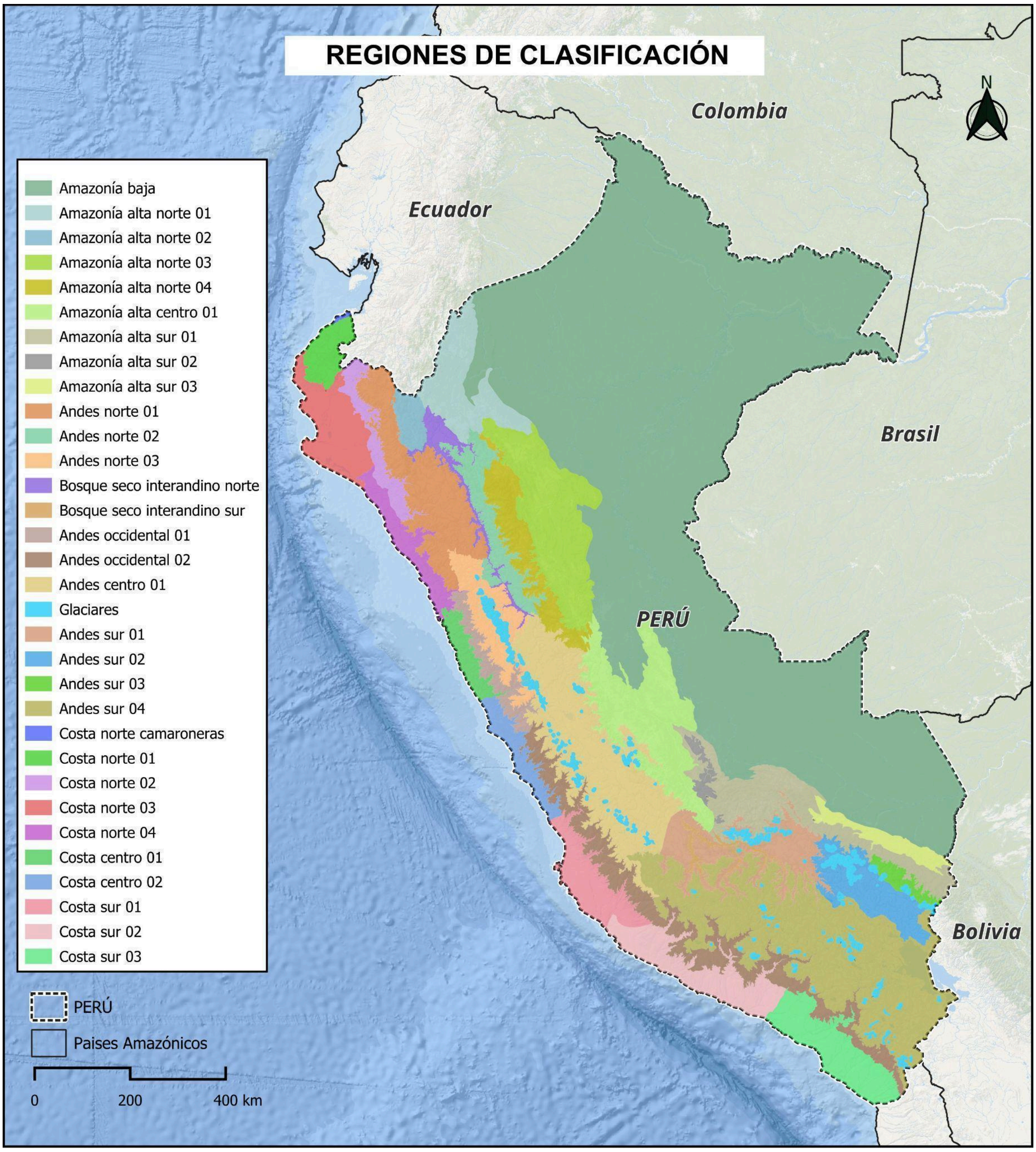


**Figure 3.** Regionalization of Peru.

For Collection 4, a total of 32 regions were defined. Their distribution across biomes is presented in Table 3.

| Bioma | # de regions |
|---|---|
| Lower Amazon | 1 |
| Upper Amazon | 8 |
| Andes | 13 |
| Coast | 10 |

**Table 3.** Number of classification regions in Peru.

### 3.3.2. Specific subregions

To adapt the classification parameters to the heterogeneity of water bodies, specific subregions were defined within each classification region. This strategy allows the application of differentiated thresholds according to the characteristics of each environment, particularly for water bodies such as wetlands, narrow rivers, and water bodies mixed with dark soils, which require special treatment. In these cases, either less restrictive thresholds (for wetlands and narrow rivers) or higher thresholds (for dark soils) are needed to ensure accurate detection. Applying these parameters locally improves the detection of such environments without increasing the occurrence of false positives or underestimating surface water in the rest of the classification region.

For the construction and delineation of these subregions, the following steps were carried out:

- Collection 3 processing: The accumulated water raster from MapBiomas Water Collection 3 was used and converted into vector format. Subsequently, geometric properties were calculated for the resulting polygons in order to differentiate and separate rivers from other water bodies, such as lagoons and reservoirs.
- Integration of auxiliary layers: The generated vector layer was complemented with official river cartography produced by the National Water Authority (ANA), as well as geometries specifically developed for the detection of other water bodies, such as wetlands and lagoons.

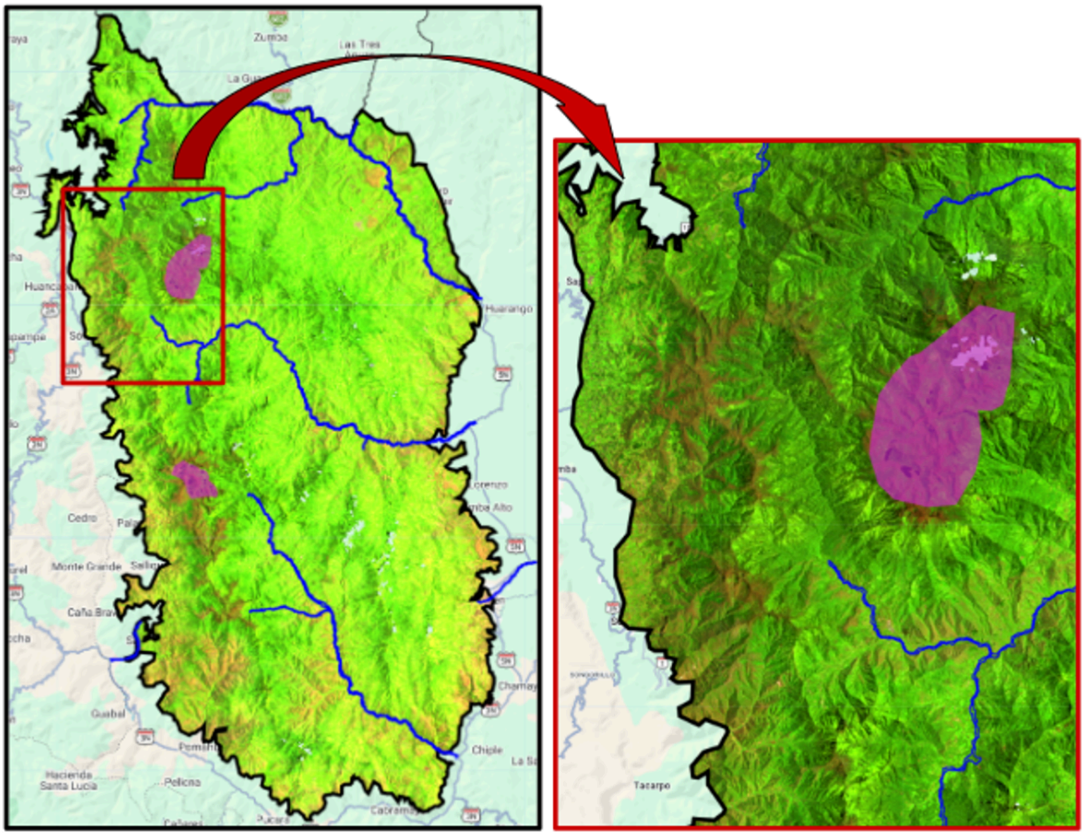

**Figure 4.** Specific subregions within classification region 1303, upper Amazon.

## 3.4. Algorithm for the Detection of Water Surface

#### 3.4.1. Surface Water Subpixel Classifier (SWSC)

The Surface Water Subpixel Classifier (SWSC) algorithm uses transition rules (Figure 5) across the fractional thresholds of the endmembers, employing a set of independent linear functions based on fuzzy logic (fuzzy rules) to determine the degree of truth/certainty (membership) that a Landsat pixel is classified as water.

The endmembers used for water detection are Shade, GV, Soil, Cloud, and Snow. Because water absorbs a large portion of electromagnetic radiation in the visible, near-infrared, and shortwave infrared Landsat bands, the SWSC primarily relies on the Shade fraction to classify pixels as surface water. In addition, the GV and Soil fractions are used to quantify the mixing of surface water with vegetation and soil. This mixed surface water typically occurs along the edges of water bodies, in narrow rivers, and within floodplain and wetland ecosystems.

The residual Cloud fraction is also incorporated to detect surface water with high sediment loads. This residual Cloud fraction arises from the spectral ambiguity between the Cloud and Soil endmembers. In fact, the residual Cloud fraction model results from the spectral response of the Soil endmember in cloud-free pixels. Finally, the Snow fraction is used to detect glacial lagoons, which, compared to non-glacial lagoons, exhibit lower absorption in the visible bands.

Due to the regionalization scheme, each subregion uses the endmembers and thresholds that are best suited for water detection. To define these parameters, water and non-water sample points were collected within each subregion, and statistical metrics such as mean, median, standard deviation, and variance were calculated. Based on this analysis, the endmembers showing the greatest separation between the water and non-water classes were selected. Thresholds were then defined using the mean values in order to maximize the detection of the water class while minimizing the detection of the non-water class, thereby reducing the number of false positives (erroneous water detections).

Based on these endmembers, two types of thresholds are defined for each classification region:

- **General thresholds:** These are used to classify water bodies with relatively pure spectral reflectance, that is, water bodies exhibiting lower levels of spectral mixing with other components. They are applied across the entire classification region and are intended to ensure the detection of large water bodies (primarily wide rivers and lagoons), while also limiting the generation of false positives throughout the region.

- **Specific thresholds:** These may be less restrictive in order to detect water bodies with greater spectral mixing, such as narrow or sinuous rivers, wetlands, and small lagoons, among others. Alternatively, they may be more restrictive than the general thresholds in areas where soils exhibit low reflectance (dark soils) and tend to generate false positives.

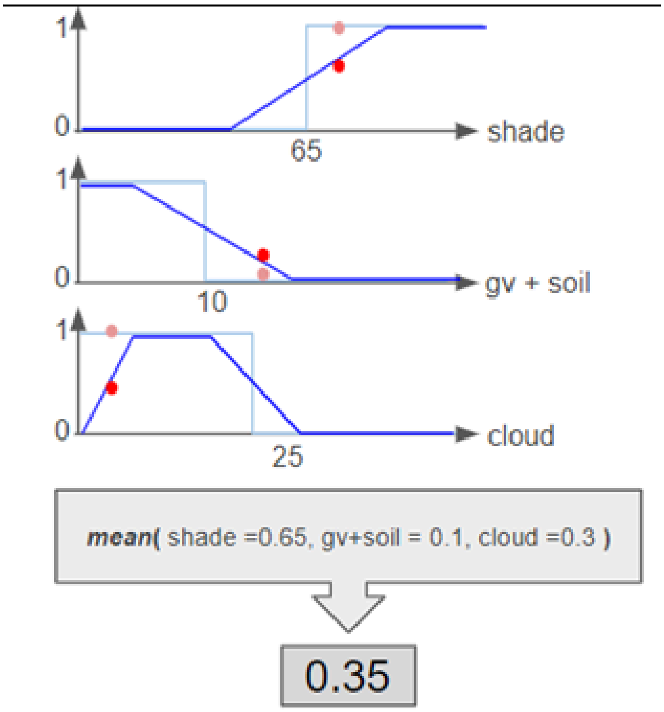


**Figure 5.** Transition rules of the water surface subpixel classifier. In this example, a fuzzy logic approach is applied to each endmember, and based on the pixel values and the selected thresholds, the final probability that the pixel represents surface water is obtained.

An example illustrating the distinction between general and specific threshold parameters for an Andes region is presented in Table 4.

**Table 4.** Example of general and specific thresholds.

| Endmembers | Threshold | | |
|---|---|---|---|
| | General | specific 1 | specific 2 |
| Minimum shade | 72 | 50 | 55 |
| Maximum shade | 92 | 60 | 65 |
| Minimum ascending GV | 0 | 5 | 0 |
| Maximum ascending GV | 5 | 10 | 5 |
| Minimum descending GV | 13 | 35 | 30 |
| Maximum descending GV | 23 | 45 | 40 |
| Minimum descending cloud | 8 | 10 | 4 |
| Maximum descending cloud | 12 | 15 | 14 |

The degree of truth (membership) is then calculated by applying a specific weight to each endmember, ultimately producing a continuous membership map with values ranging from 0 to 1.

### 3.4.2. Auxiliary mosaics

Data gaps, whether caused by cloud and shadow masking or by the exclusion of images with high cloud cover, represent a challenge across all regions of Peru. To address these information gaps, three auxiliary mosaics were generated:
- **Five-month moving mosaic:** Based on a five-month moving temporal window (a seven-month window is used for the upper Amazon region), consisting of the target month plus the two preceding and two subsequent months. This window is calculated independently for each month in the time series. Within this temporal window, Landsat images are filtered by month and monthly mosaics are generated using the median reducer (resulting in a total of five monthly mosaics). The Surface Water Subpixel Classifier (SWSC) is then applied to each mosaic to produce a continuous membership map. Finally, a weighted average of the five monthly membership maps is calculated, assigning weights according to the temporal proximity of each month to the central month.

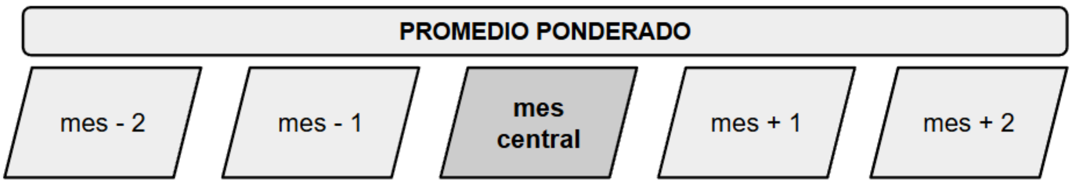


**Figure 6.** Logic for calculating the 5-month moving mosaic.

**Decadal mosaic:** This mosaic is based on an interannual temporal window composed of the target month and the same month from the five preceding and five subsequent years (for the upper Amazon region, a window of two preceding and two subsequent years is used). Landsat images are filtered by month and year, and monthly mosaics are generated using the median reducer (resulting in a total of 11 mosaics). For each month-year combination, the Surface Water Subpixel Classifier (SWSC) is applied to generate a membership probability map. Subsequently, the average membership probability is calculated considering only the years with valid observations. Finally, the resulting map is masked using a condition that requires the number of valid observations from years after the target year to be greater than or equal to the number of valid observations from years preceding the target year.

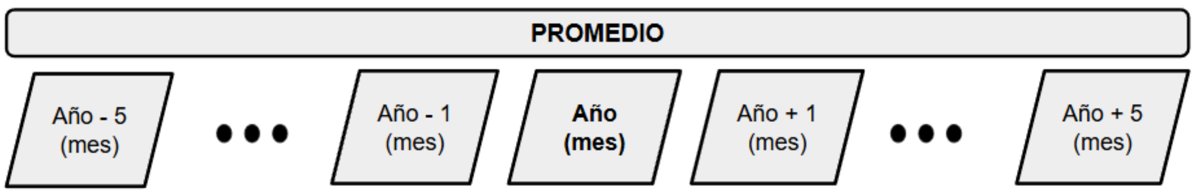


**Figure 7.** Decadal mosaic calculation logic.

- **Seasonal mosaic:** To generate the seasonal mosaic, all Landsat images available for a target year are used, and the band corresponding to the Soil endmember is selected. These images are then reduced using a specific percentile, which varies according to the seasonality (wet, dry, or transitional season) and the classification region. The resulting raster is used as a threshold to mask pixels in each image of the target year, retaining only those whose values satisfy a condition defined relative to the selected percentile. For example, for a wet-season mosaic, pixels with lower Soil values are retained, whereas for a dry-season mosaic, pixels with higher Soil values are preserved. Finally, the seasonal mosaic is generated by applying the median reducer to the set of masked images, and the Surface Water Subpixel Classifier (SWSC) is applied to obtain a membership probability map. The specific seasonal mosaic used depends on both the month being analyzed and the classification region.

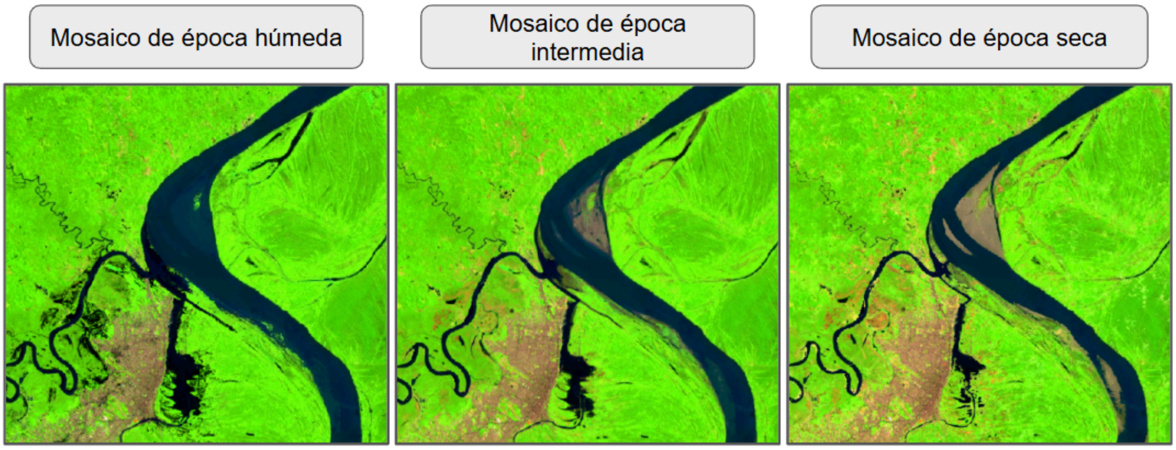


**Figure 8.** Comparison of seasonal mosaics in Iquitos for the year 2024.

Finally, the five-month moving mosaic and the seasonal mosaic are integrated into an annual mosaic (Figure 9), where the five-month mosaic predominates (because it is closer to the target month and therefore more closely reflects the actual conditions), while the seasonal mosaic is used as a background layer to fill missing pixels (since, by using all available images from the target year, it tends to be more complete than the five-month mosaic).

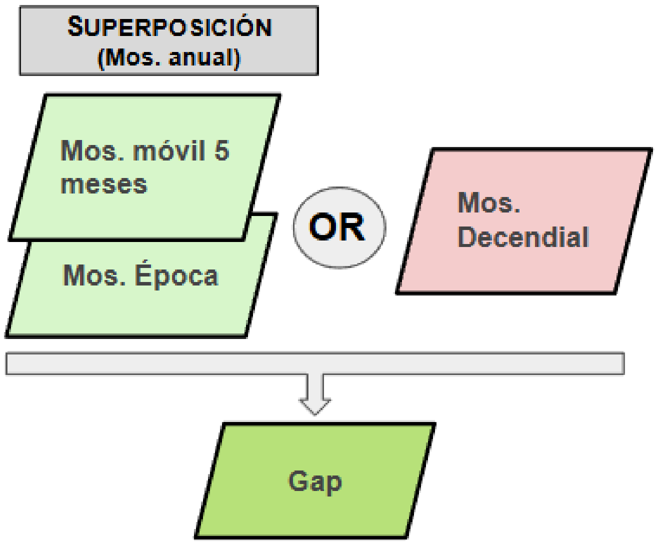


**Figure 9.** Integration of auxiliary mosaics.

#### 3.4.3. Monthly water data extraction

##### 3.4.3.1. Water surface classification

The monthly maps were complemented with procedures to restore false negatives and remove false positives based on temporal metrics (Figure 11). A distinction is also made between two contexts: when monthly information is available, detection is based exclusively on the monthly membership, whereas when monthly information is not available (empty pixels), detection is based on the membership generated from the auxiliary mosaics (Figure 10).

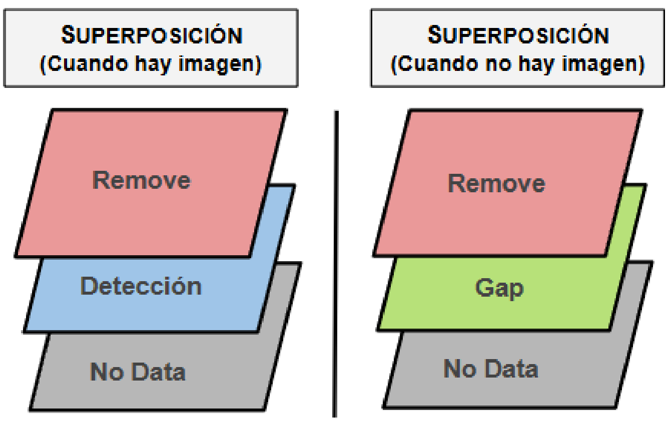


**Figure 10.** Logical flow of spatial overlay under presence and absence of monthly data.

Once the distinction between contexts has been made, the Water Surface Subpixel Classifier (SWSC) is applied to the monthly images, producing a raster in which the probability value for each pixel ranges from 0 to 1. A threshold is then applied to this raster under two conditions: when the pixel has a value lower than a defined slope threshold, a water detection threshold ranging from > 0.6 to 0.67, depending on the region, is applied; whereas if the pixel has a value higher than the defined slope threshold, a detection threshold ranging from > 0.8 to 0.9 is applied. This distinction is made because it is less likely for a water body to occur on steep slopes, thereby removing false positives generated by mountain shadows.

Subsequently, for pixels with no available information, gaps are filled by applying the SWSC to the auxiliary mosaics (annual and dekadal), using an inclusion threshold of 0.5 for both mosaics when the pixels have values below the slope threshold, and an inclusion threshold of 0.8 when the pixels have values above the slope threshold. Additionally, a conditional rule is applied so that gap filling can be performed even when only one of the two auxiliary mosaics is available. If both mosaics are available, both must exceed the detection threshold for the pixel to be classified as water.

The presence of cloud shadows or other dark objects in the Landsat scene can also produce false positives in the water classification. Therefore, a removal filter was applied to reclassify as non-water those pixels with a probability < 0.35 based on the annual mosaic.

Finally, the classification raster (Figure 11) is composed of the monthly detection (based on the available monthly mosaics) + the inclusion of pixels with no information (based on the auxiliary mosaics) – the removal process (based on the annual mosaic).

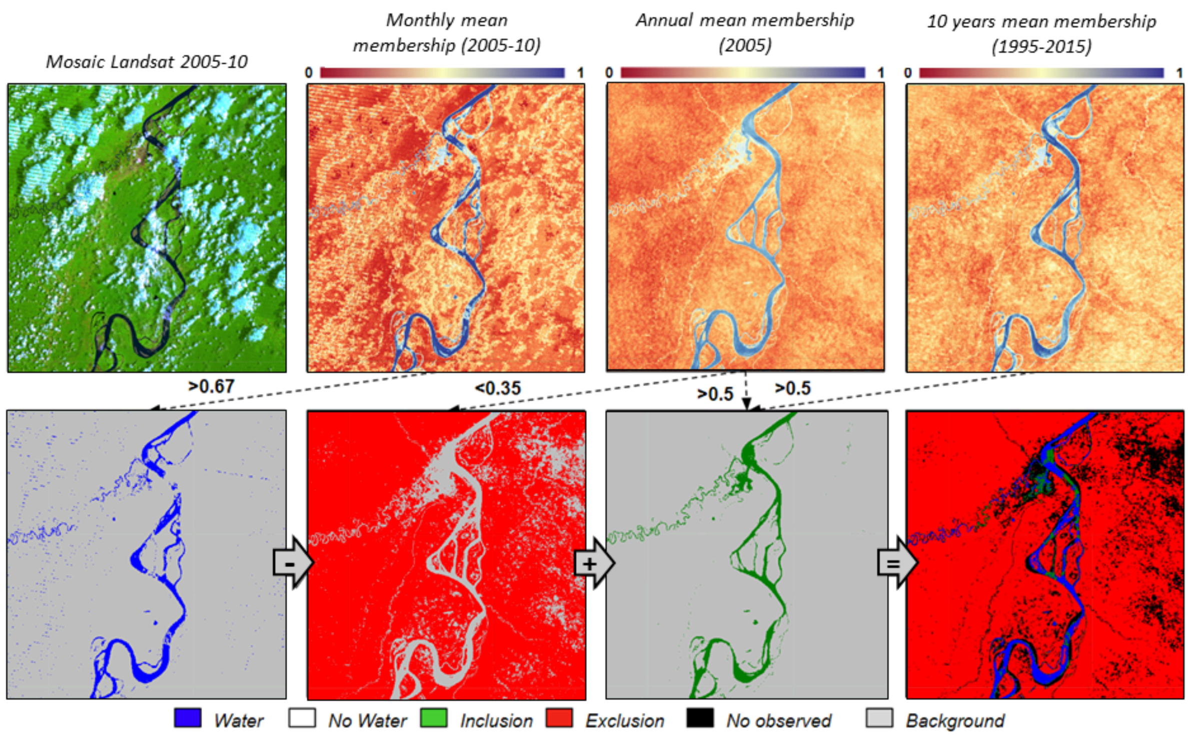


**Figure 11.** Methodological process for water surface classification. a: Monthly Landsat mosaic; b: Monthly median SWSC membership; c: Annual monthly median SWSC membership; d: Decadal monthly median SWSC membership; e: Monthly surface water classification; f: Total surface water per decade; g: Area likely to be surface water based on the thresholds of c and d. h: Final surface water map for the month, with corresponding inclusion and removal.

#### 3.4.3.2. Slope-based mask

Once the monthly probability raster is obtained, pixels that meet the water detection threshold are masked based on two variables: HAND and slope (derived from the SRTM

DEM).

- **HAND:** This is a terrain model that normalizes topography by measuring the vertical distance between any point in the landscape and its nearest drainage channel or stream. This product is available in Google Earth Engine and has a spatial resolution of 30 meters, making it compatible with Landsat imagery. Threshold values ranging from 40 to 60 are used, depending on the classification region.

- **Topographic slope:** Slope derived from the SRTM DEM available in Google Earth Engine is used as a mask to help reduce false positives. Like HAND, it has a spatial resolution of 30 meters. Threshold values ranging from 10 to 20 are used, depending on the classification region.

### 3.5. Temporal filter

To improve the classified surface water map, a post-processing step was developed consisting of a series of rules designed to remove false positives and correct false negatives based on temporal criteria. The use or omission of each rule, as well as the order in which they are applied, depends on the biome.

a) Rule 1: Identifies the date at which a pixel first becomes, or is detected as, water in a stable manner. To achieve this, the temporal persistence of a pixel classified as water immediately after its first appearance is evaluated using a moving window of generally 24 months. In addition, a dynamic threshold is incorporated, corresponding to the minimum percentage of detections required within this window for a pixel to be considered a valid first detection.

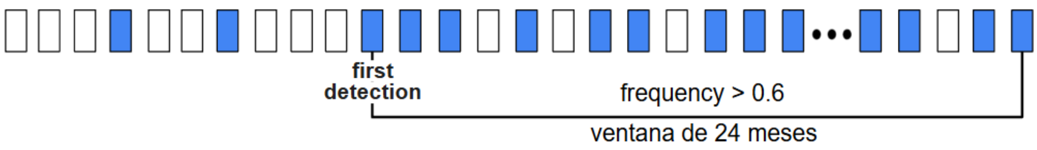


**Figure 12.** Logical scheme of Rule No. 1

b) Rule 2: Corrects false negatives caused by pixels with reflectance altered by clouds or by the lack of information in the monthly and auxiliary mosaics. The algorithm evaluates water persistence by analyzing the immediate temporal context of the pixel (both backward and forward in time) using moving windows. A pixel originally classified as "non-water" is corrected to "water" if its presence is confirmed at least once both before and after the target date.

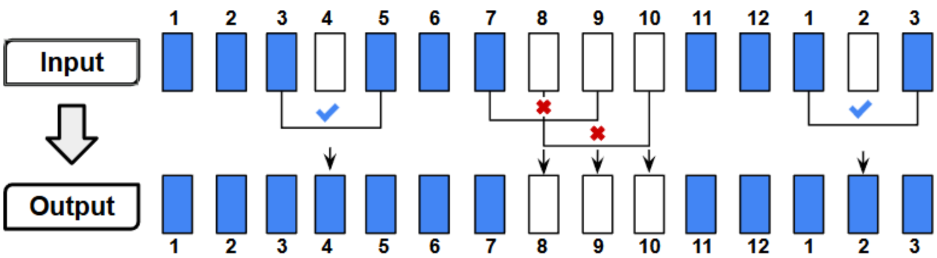


**Figure 13.** Logical scheme of Rule No. 2.

c) Rule 3: Seeks to correct inconsistencies between seasons. The algorithm assumes that a pixel cannot exhibit more water detections during the dry or transitional season than during the wet season. Therefore, detections occurring only during the dry or transitional season are reclassified as "non-water".

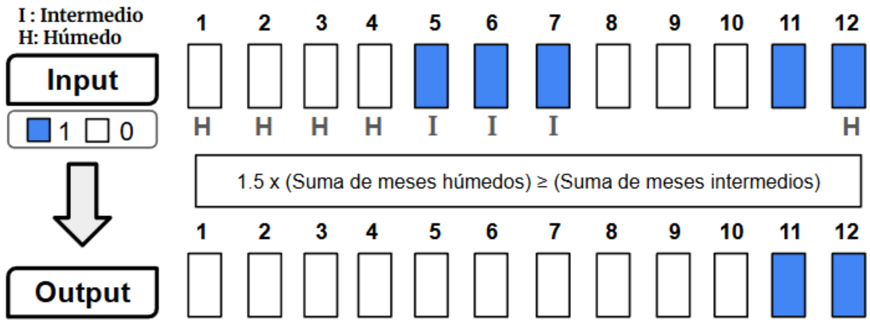


**Figure 14.** Logical scheme of Rule No. 3: comparison of wet and intermediate months.

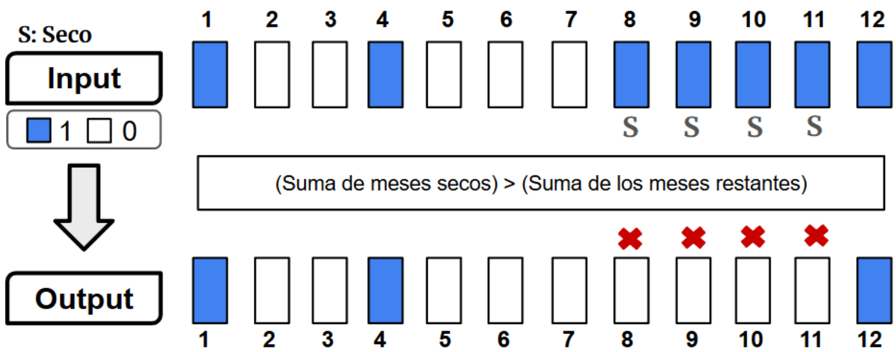


**Figure 15.** Logical scheme of Rule No. 3: comparison of dry months and remaining months.

d) Rule 4: Evaluates the persistence of a pixel forward in time, taking into account its first valid detection. The algorithm assesses the persistence rate by calculating the proportion of water detections between the first detection and the last month of the historical time series. If a pixel exceeds a defined threshold (0.8–0.9), it is assumed to be permanent water, and all months classified as "non-water" between the first detection and the last month of the series are reclassified as "water".

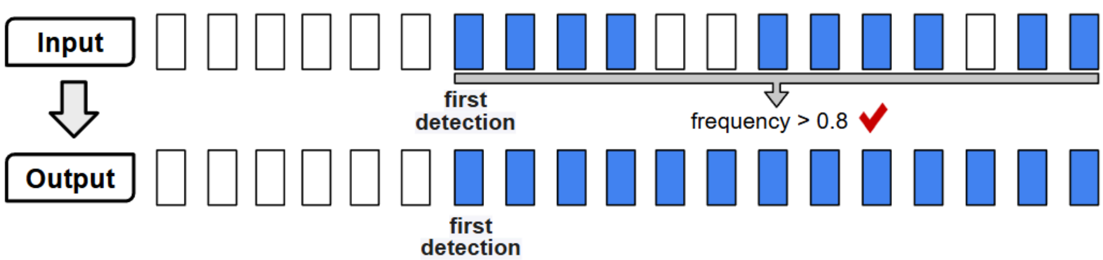


**Figure 16.** Logical scheme of Rule No. 4.

e) Rule 5: Removes false positives occurring before the first valid detection. The algorithm evaluates the interval between the first month of the time series and the first valid detection (established in Rule 1). It then calculates the proportion of water detections within this interval, and if a defined threshold is not exceeded, the detections are assumed to be errors and are reclassified as "non-water".

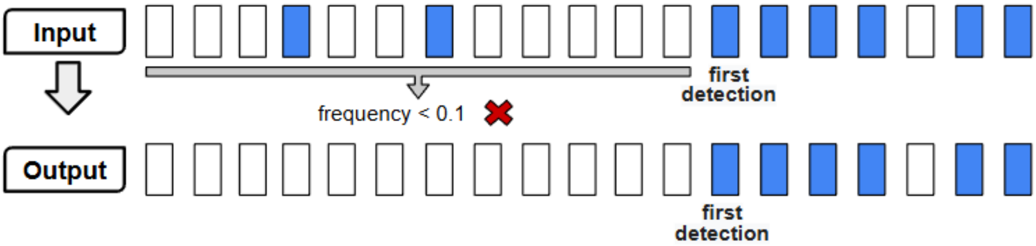


**Figure 17.** Logical scheme of Rule No. 5.

f) Rule 6: Because the recurrent cloud cover during the wet season limits pixel quality and leads to underestimation of water dynamics, a correction criterion based on annual frequency was implemented to address these omissions at a broad scale. This rule seeks to compensate for the lack of detection during critical periods of high cloud cover and low-quality pixels. To achieve this, the algorithm evaluates the annual sum of water detections; if a pixel exhibits persistent water presence throughout the year (>= 4 detections), omissions occurring between January and April are automatically filled as "water". Following the same logic, if annual persistence is more robust (>= 8 detections), the month of December is also corrected.

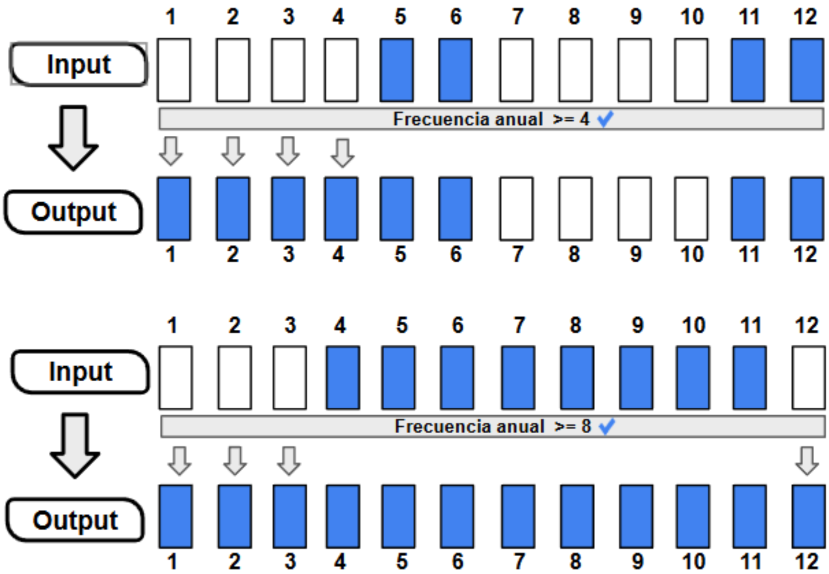


**Figure 18.** Logical scheme of Rule No. 6.

Due to the spatial heterogeneity among biomes, the rules were regionalized, as detailed in Table 5.

**Table 5.** Temporal filter rule sequence by biome.

| Regions | Order of application | | | | |
|---|---|---|---|---|---|
| | 1° | 2° | 3° | 4° | 5° |
| Lower Amazon | Rule 1 | Rule 2 | Rule 4 | Rule 5 | – |
| Upper Amazon | Rule 1 | Rule 2 | Rule 6 | Rule 3 | Rule 4 |
| Andes | Rule 3 | Rule 1 | Rule 2 | Rule 4 | Rule 5 |
| Coast | Rule 3 | Rule 1 | Rule 2 | Rule 4 | Rule 5 |

### 3.6. Masking

As the final processing step, a spatial mask was designed and implemented to permanently remove false positives that persist after the temporal filtering stage while preserving areas with the presence of water bodies. This control mask integrates four key geospatial inputs:

- The accumulated historical water frequency from MapBiomas Water Peru Collection 3.
- An optimized raster map derived from the river network generated during the subregionalization phase.
- A map of specific regions delineated in areas where water bodies omitted by MapBiomas Water Collection 3 were identified.
- The inclusion of the "Rice" class from MapBiomas Peru Land Cover and Land Use Collection 3, used to remove these agricultural areas because they exhibit reflectance characteristics similar to water.

The integration of the first three inputs allows the preservation of actual water areas, while the rice layer helps remove false positives.

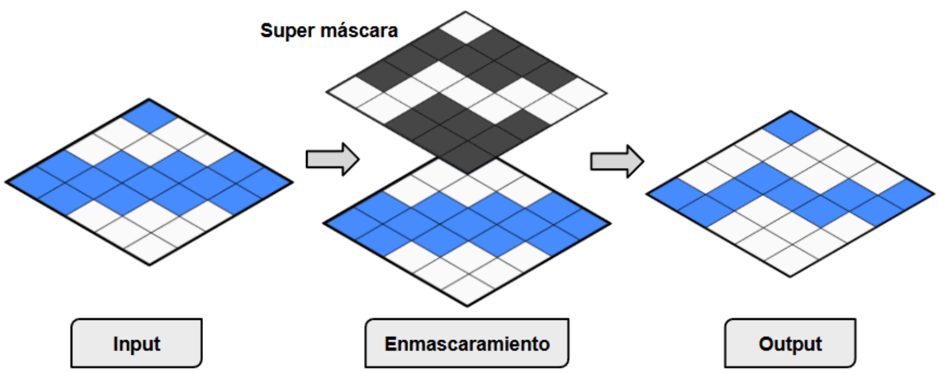


**Figure 19.** Masking process.

### 3.7. Annual water data extraction

The annual surface water maps include a distinction between permanent and seasonal water. This classification is based on thresholds corresponding to the number of months in which a pixel is classified as water. For the Amazon biome, a frequency of >= 6 months is considered, whereas for the remaining biomes a frequency of >= 3 months is used. The definition of the permanent water threshold is based on the existence of both a dry and a wet season across much of the country, thereby encompassing permanent water bodies that exhibit natural seasonal dynamics.

## 4. Water body categorization methodology

The classification of water bodies included the following steps:

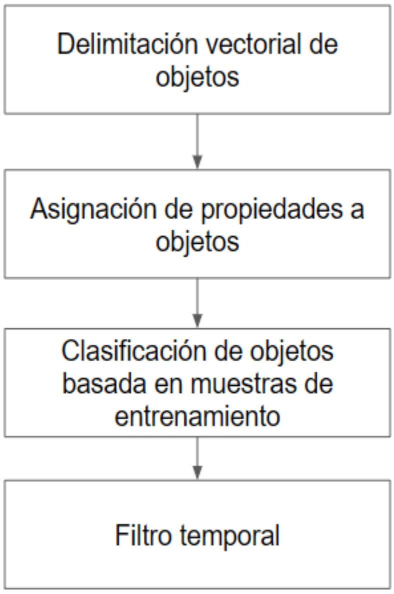


**Figure 20.** Sequence of steps for classification.

### 4.1. Vector object delineation

The vector delineation of objects consists of converting the annual surface water maps (raster data) for each year into regular polygons (vector data) within the spatial extent of water bodies.

This procedure was performed using a segmentation tool, in which a particular water body could be converted into one or more polygons. The SNIC function available in Google Earth Engine was used to generate small and relatively regular segments. Figure 21 shows some examples of the segmentation based on monthly frequency data for a given year.

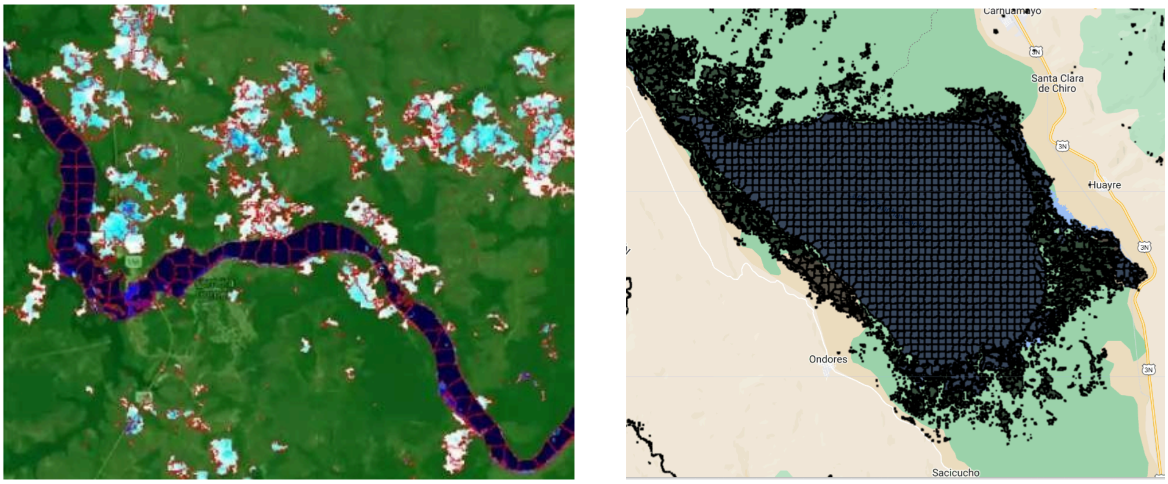


**Figure 21.** Raster data segmentation process.

## 4.2. Property assignment to objects

After generating the objects for each year, a feature dataset containing new properties was assigned to each object for subsequent use in the classification process. These properties include information related to object morphology, geomorphology, and qualitative information derived from other studies on water body classification and land cover and land use mapping. The following variables were associated with each object: area, perimeter, area-to-perimeter ratio, compactness, roundness, elongation degree, length–width ratio (Laenge–Breite), convexity, maximum extent, number of neighbors, number of neighbors within a 50-meter buffer, ANA classification – anthropogenic, ANA classification – hydropower, MapBiomas land cover and land use classes (urban, mining, forest, non-forest class, pasture), maximum SRTM value, and mean total frequency.

```
▾properties: Object (28 properties)
    count: 202
    data_anaAnt: 0
    data_anaHid: 0
    data_anaNat: 1
    data_mpbCls: 33
    data_mpbFor: 0.2154718095820371
    data_mpbMin: 0
    data_mpbNnf: 0.0025129006373729268
    data_mpbPas: 0
    data_mpbUrb: 0
    data_srtmMx: 0
    freq_meanAn: 11.999961165802608
    freq_meanTo: 419.9939809331496
    freq_tempPr: 1
    freq_tempUl: 0
    label: -546964717
    morp_aream2: 181092.90434147738
    morp_convex: 0.7218790322491963
    morp_maxExt: 598.7157217544923
    morp_nrVi50: 7
    morp_nrViim: 6
    morp_perime: 2031.732752721085
    morp_r_aPeC: 1.3468491326060188
    morp_r_aPer: 89.13224640344109
    morp_r_circ: 0.5512874862582181
    morp_r_elon: 0.801750342496893
    morp_r_laCu: 0.6967265601438797
    morp_r_shpf: 1.979428828420977
```


**Figure 22.** Properties assigned to vectors.

### 4.3. Object classification based on training samples

The classification of water bodies was performed using different methodologies depending on the target class:

The Random Forest algorithm was used to classify three specific classes: “Rivers, Lakes, and Lagoons”, “Agricultural”, and “Mining”. Training samples were collected across the different biomes (Figure 23), covering multiple years of the time series. In addition, the “Mining” class was complemented with information generated by MapBiomas Land Cover and Land Use Collection 3.

For the “Aquaculture” and “Hydropower” classes, reference vector datasets from government authorities, such as the Ministry of Production and the Ministry of Energy and Mines, respectively, were used.

For the “Glacial Lakes” class, two detection criteria were applied: lakes had to be located within a 3-kilometer buffer of the 1985 glacier extent and exhibit an increase in water surface area greater than 30%.

For the “Regulated Lakes” class, detection was performed using an algorithm that identified expansions in pre-existing water bodies and was subsequently cross-referenced with reference information on hydraulic infrastructure (reservoirs, dams, and related structures).

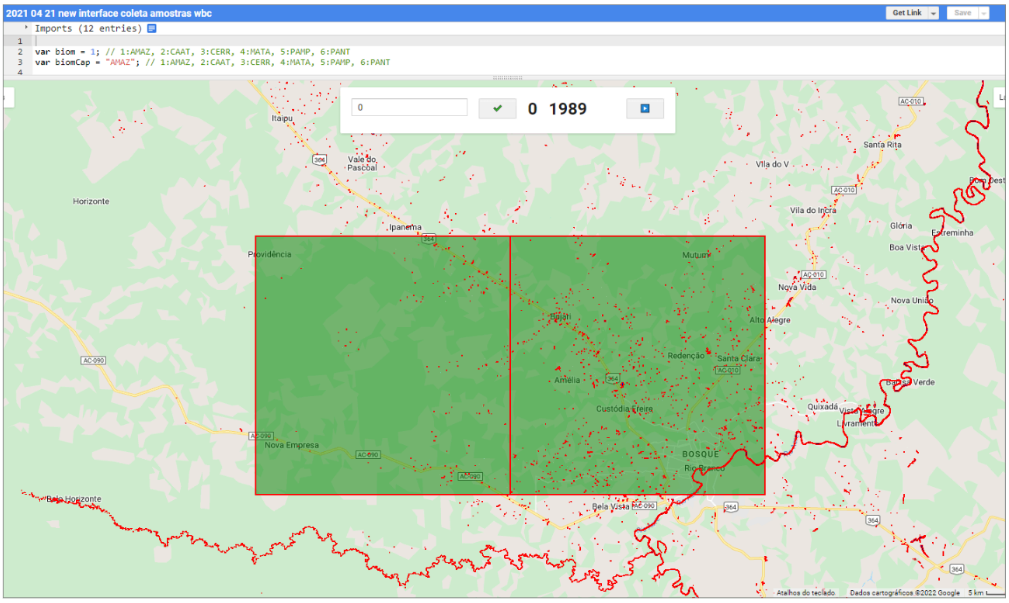


**Figure 23.** Sample collection interface.

### 4.4. Temporal filter

After classification, the results were subjected to a post-classification routine by applying a temporal filter. The logic of the temporal filter was to eliminate unlikely class transitions within the same segment throughout the time series.

Finally, all polygons classified as false positives within each year were converted back to raster format and used to filter the annual and monthly surface water datasets, removing the remaining false positives.

### 4.5. Water body categorization

Through these processes, surface water is hierarchically classified into two levels according to its origin: Natural and Anthropic. This methodological approach enabled the identification of seven specific classes: 1) Rivers, lakes, and lagoons, 2) Regulated Lagoons, 3) Glacial Lagoons, 4) Mining water, 5) Hydroelectric, 6) Aquaculture, and 7) Agricultural and Other Uses, ensuring a precise typification of water resources at the national level. A description of each category is provided in Table 6.

| **CLASE NIVEL 1** | **CLASE NIVEL 2** | **DESCRIPCIÓN** |
|---|---|---|
| Natural | Rivers, lakes y lagoons | Cuerpos de agua superficiales naturales de origen fluvial o lacustre, permanentes o estacionales, distribuidos a lo largo de todo el territorio nacional. Incluyen cursos principales y secundarios, así como espejos de agua de origen diverso no intervenidos estructuralmente. |
| | Lagoons regulates | Cuerpos de agua de origen natural que presentan un aumento significativo en su superficie como resultado de intervenciones |

| | | |
|---|---|---|
| | | humanas, principalmente mediante infraestructura hidráulica como diques o represas. |
| | Glacial lagoons | Cuerpos de agua formados por el retroceso de glaciares a partir de 1985, característicos de zonas altoandinas. |
| Anthropic | Mining water | Cuerpos de agua artificiales originados por actividades extractivas, como tajos abiertos, relaveras, sedimentadores, salares u otras infraestructuras asociadas al sector minero. |
| | Hydroelectric | Embalses o reservorios creados para la generación de energía a partir del aprovechamiento de caudales fluviales. |
| | Aquaculture | Cuerpos de agua diseñados para el cultivo intensivo de organismos acuáticos, tales como estanques, piscigranjas o jaulas flotantes. |
| | Agricultural and other uses | Cuerpos de agua artificiales vinculados al almacenamiento o uso de agua para actividades agrícolas, ganaderas, urbanas o industriales. |

**Table 6.** Classification scheme and description of water bodies.

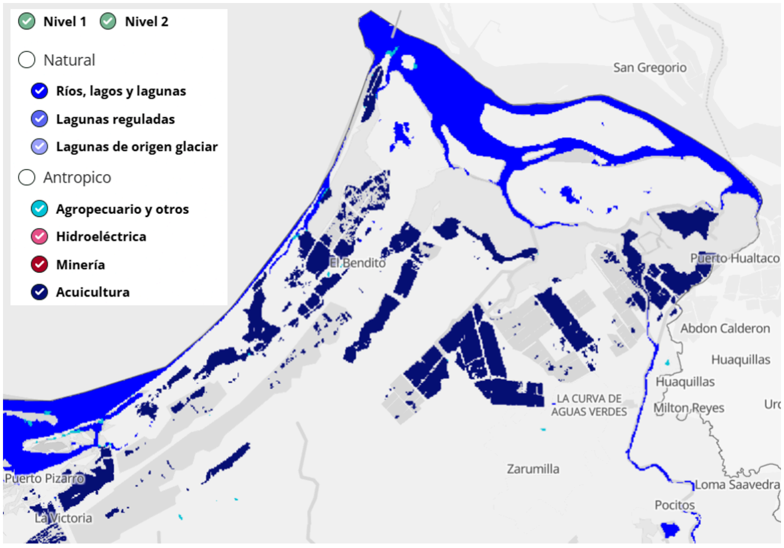


**Figure 24.** Example of water body categories in Tumbes, Peru.

## 5. Map collection and analysis

The main results of Collection 4 are available through the public web platform: https://plataforma.peru.mapbiomas.org/agua, where seven products are provided:

- Annual Surface Water
- Monthly Surface Water
- Water Body Types
- Surface Water Trend
- Surface Water Transition
- Annual Surface Water Frequency
- Time of Last Occurrence

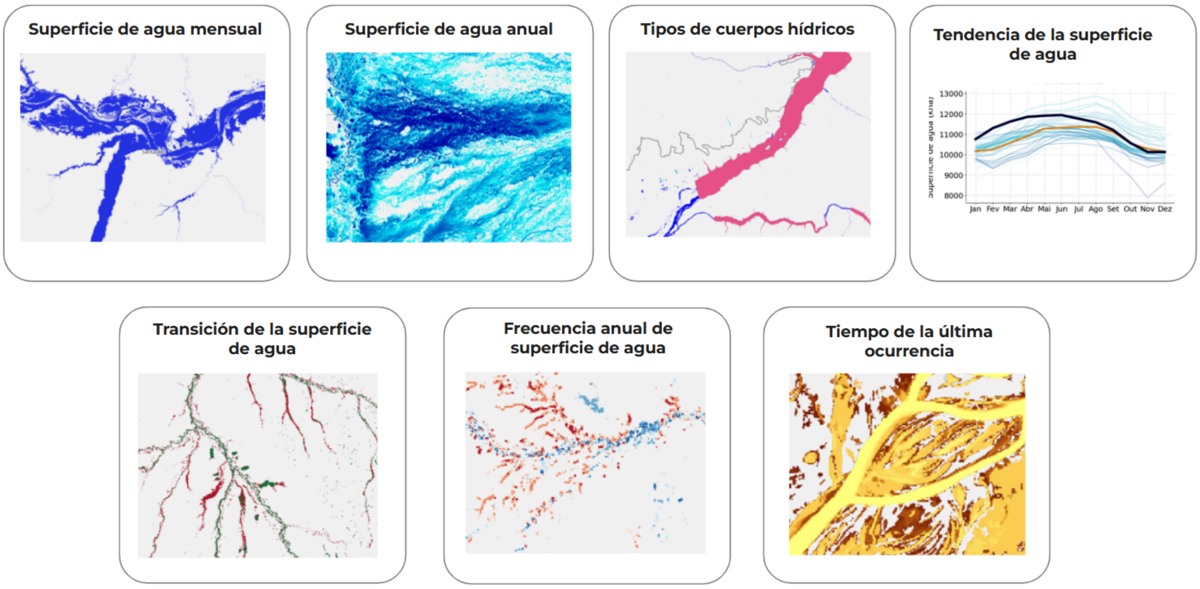


**Figure 25.** Products obtained from monthly water surface data, available on the platform.

## 5.1. Annual water surface

This variable corresponds to the annual surface water dataset, which can be visualized for any year within the 1985–2025 period (Figure 26).

The available data are organized by different territorial units, such as biomes, river basins, departments, and others. Users can select different territories and time ranges, allowing the map, charts, and statistics to be updated dynamically.

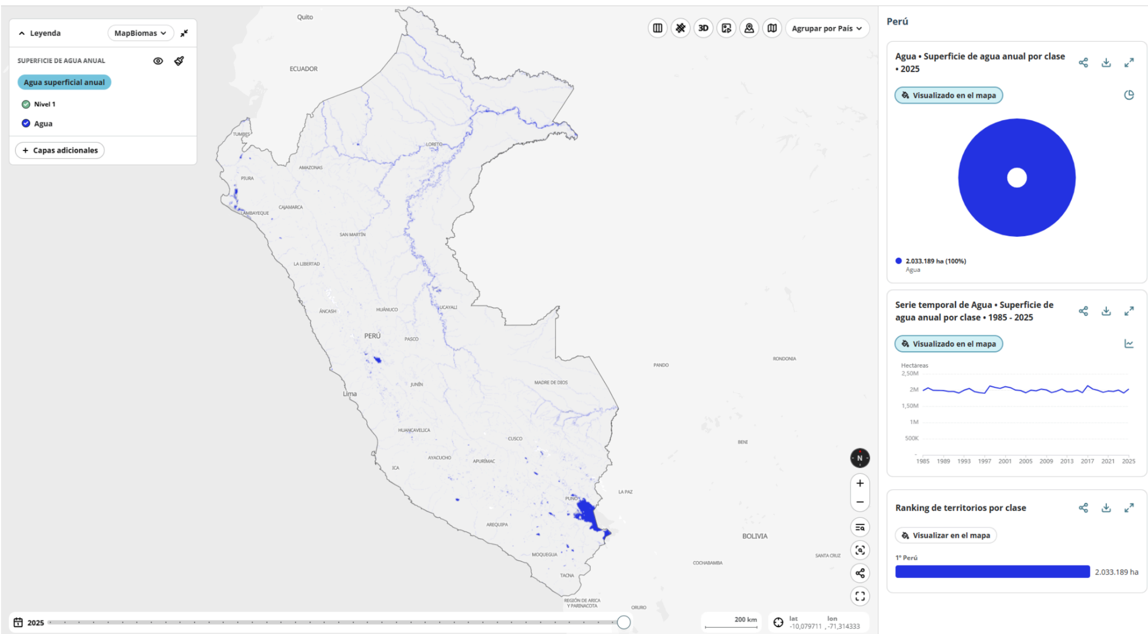


**Figure 26.** Example of annual water surface data visualization (map and chart) on the MapBiomas Agua platform.

## 5.2. Monthly water surface

This variable corresponds to the monthly surface water dataset, in which all months can be visualized simultaneously using a distinct color palette for each month. In addition to the monthly map, charts and statistics are provided, including the month with the largest and smallest surface water extent in a given year, the average surface water extent across all months of the year, the deviation from the historical monthly average, and monthly surface water trends by year.

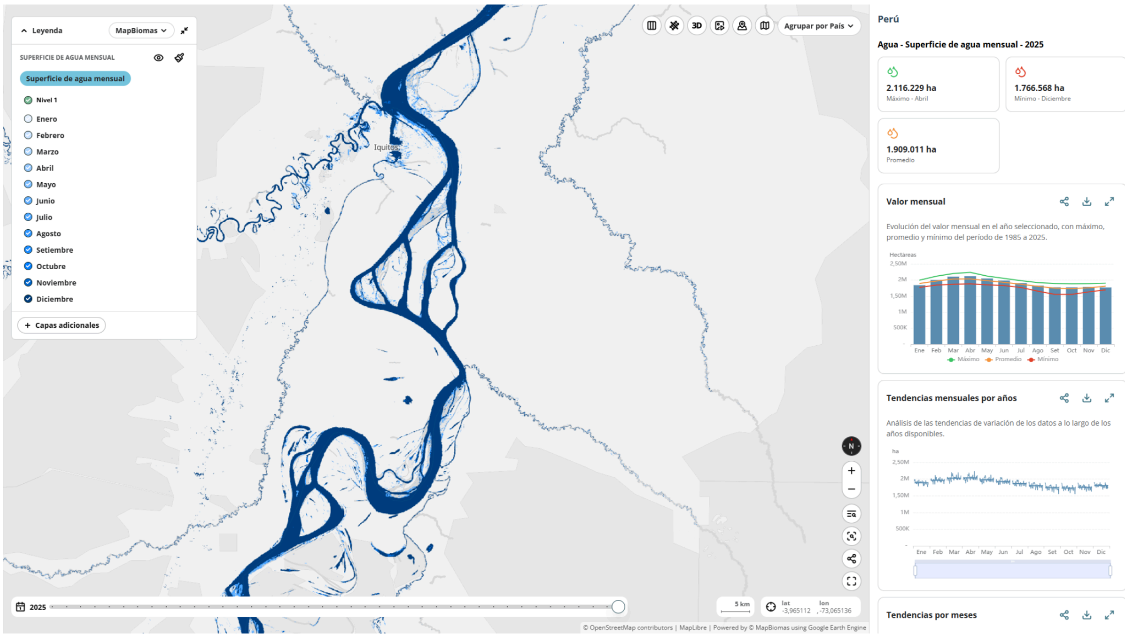


**Figure 27.** Example of monthly water surface data visualization (map and chart) on the MapBiomas Agua platform.

### 5.3. Water body types

This variable corresponds to the monthly surface water dataset, in which all months can be visualized simultaneously using a distinct color palette for each month. In addition to the monthly map, charts and statistics are provided, including the month with the largest and smallest surface water extent in a given year, the average surface water extent across all months of the year, the deviation from the historical monthly average, and monthly surface water trends by year.

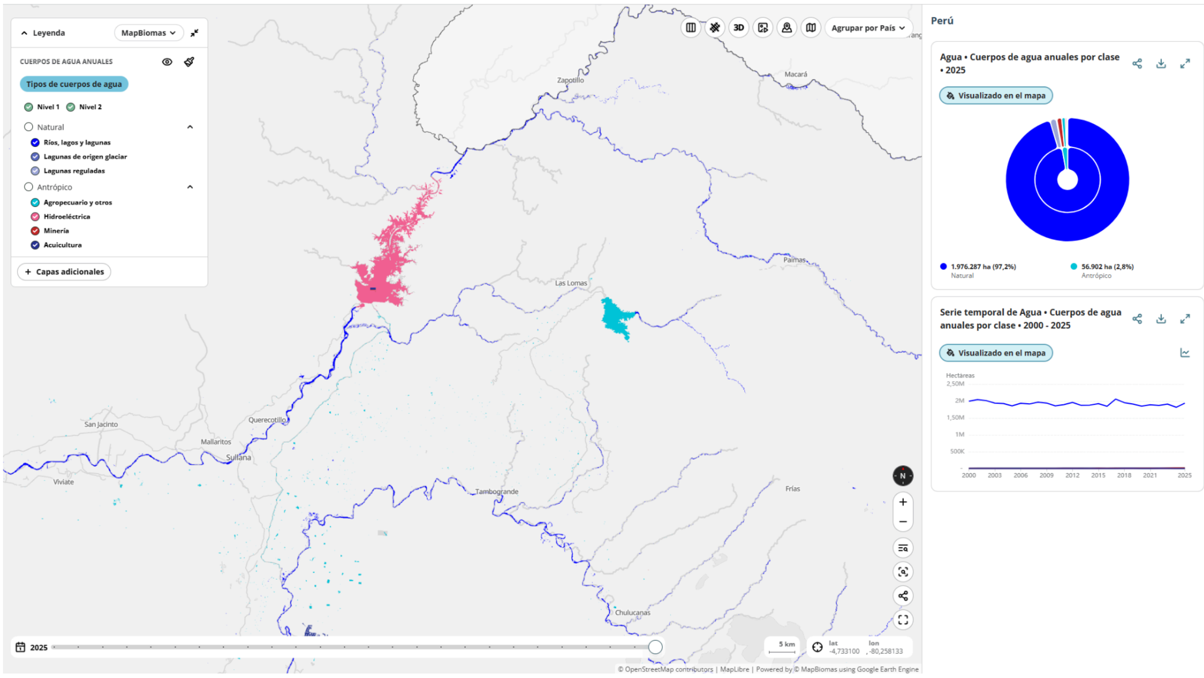


**Figure 28.** Example of water body types visualization (map and chart) on the MapBiomas Agua platform.

## 5.4. Water surface trend

The surface water trend analysis was performed using the monthly surface water database. Trends of increase, decrease, or persistence in mapped surface water extent were assessed throughout the 1985–2025 time series. The analysis was conducted using the Seasonal Mann–Kendall test (MK test), which is used to analyze data collected over time for increasing or decreasing trends with a monotonic behavior in the Y-axis values. It is a non-parametric test, meaning that the data are not required to meet normality assumptions, and it evaluates seasonal data for the presence of monotonic trends (Hirsch et al., 1982; Hirsch et al., 1984; Gilbert, 1987; Helsel and Hirsch, 1995; Morell and Fried, 2009).

On the platform, the map is displayed using different color intensities according to the magnitude of the trend (positive or negative). Only statistically significant results are presented.

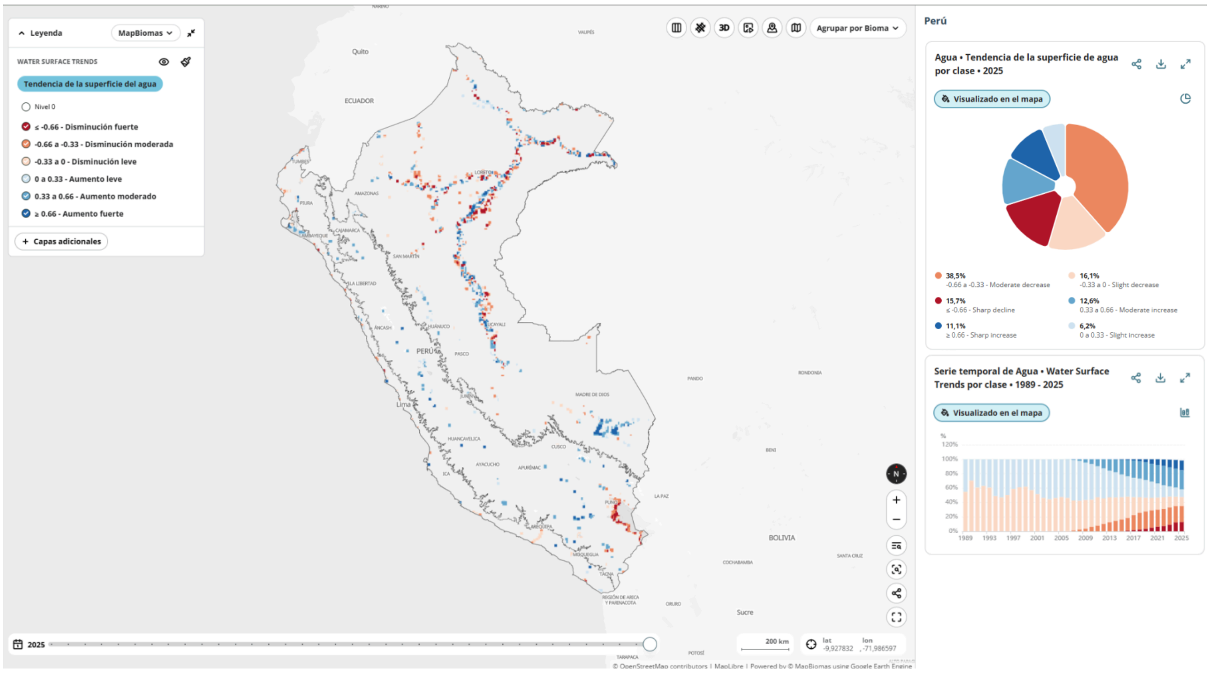


**Figure 29.** Example of water surface trend visualization on the MapBiomas Agua platform.

## 5.5. Water surface transition

This layer represents surface water areas that have decreased, increased, or remained stable throughout the time series. These areas were identified using the total number of pixels classified as water across the entire annual time series (i.e., 41 years). An RGB color composite was generated to facilitate the visualization and identification of these categories.

First, to characterize persistence, the blue channel was assigned to the total number of years classified as surface water. Second, the number of years from the beginning of the time series to the first water classification was selected and assigned to the green channel to characterize surface water expansion. Finally, the number of years from the last observation of surface water to the end of the time series was selected and assigned to the red channel to indicate surface water decline. As a result, permanent water bodies appear predominantly blue, temporary water bodies appear black when they occur sporadically, areas of surface water loss appear red, and newly formed water areas appear green (Figure 30).

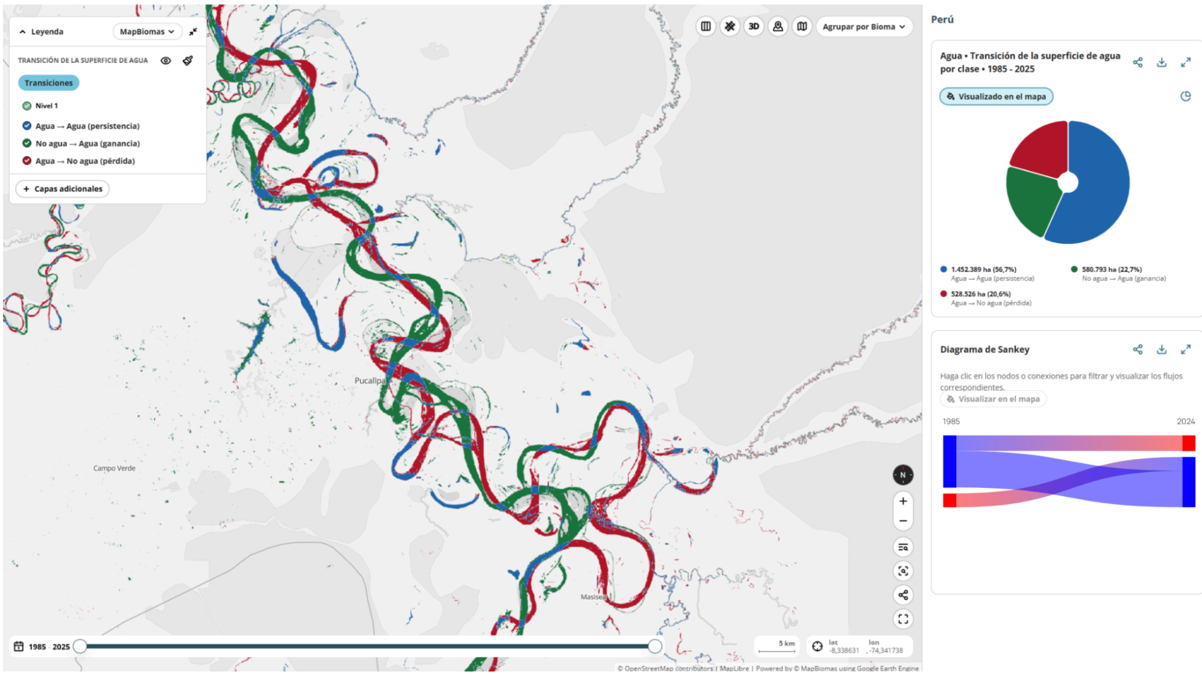


**Figure 30.** Example of water surface transitions visualization on the MapBiomas Agua platform.

## 5.6. Water surface frequency

This variable corresponds to the frequency derived from the annual surface water dataset. It displays, using different color intensities, the number of times a pixel has been mapped as water throughout the time series.

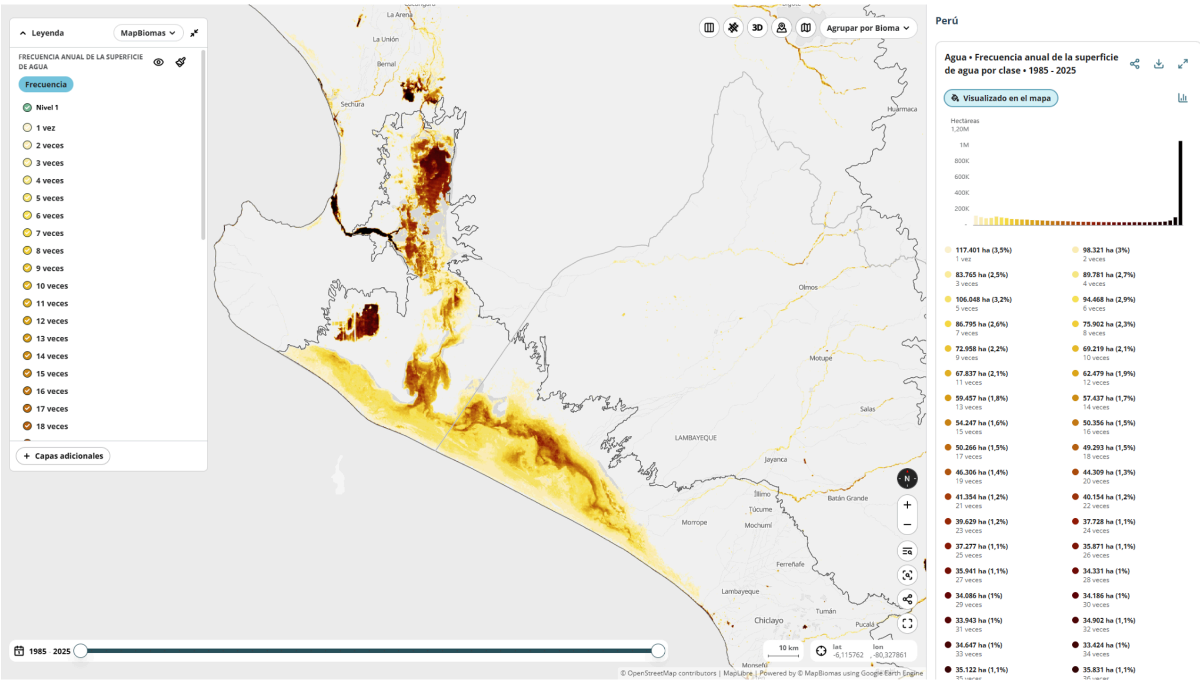


**Figure 31.** Example of water surface frequency visualization on the MapBiomas Agua platform.

### 5.7. Time since last occurrence

This variable represents the number of years that have elapsed since the last detection of surface water for a given year. It is useful for visualizing areas where surface water was present in the past but is no longer observed at present.

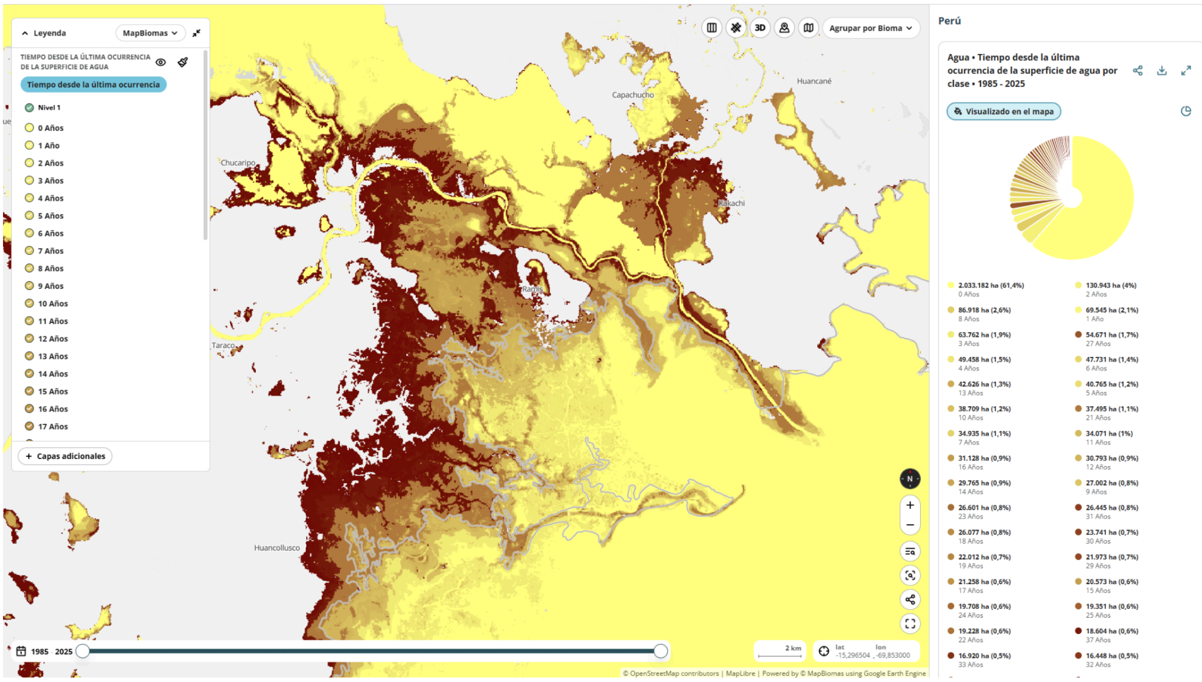


**Figure 32.** Example of time since last occurrence visualization on the MapBiomas Agua platform.

## 6. Practical considerations

The fourth version of this product represents a significant step forward in the continuous improvement process and responds to observations and suggestions made during the use and analysis of the first version. This new product is conceived as a dynamic tool, designed not only to provide relevant data but also to evolve in response to user needs and available technological advancements.

All generated data are publicly available, promoting transparency and open access. This approach enables users to review the content, identify strengths, and propose necessary adjustments. User contributions and feedback are essential to ensure that future versions continue through an iterative refinement process and that the product remains up to date and relevant.

It is important for users to consider that the application of the quantitative data presented should be accompanied by a thorough review of the associated accuracy levels and uncertainty metrics. This will enable a more robust interpretation and an appropriate assessment of their applicability across different contexts. The data should be compared with other sources or complementary studies to confirm their validity or identify potential limitations.

This continuous improvement approach reinforces the developers' commitment to the quality and usefulness of the product, promoting collaboration between the scientific community and other users to maximize the positive impact of the dataset.

## 7. Conclusions and perspectives

The MapBiomas Water Collection 4 has become a fundamental mapping product for wáter bodies in the Amazonian countries. Its contribution is significant for understanding the spatial and temporal dynamics of freshwater ecosystems, providing detailed data that allows for the analysis of change patterns, cycles of increase or decrease, and the estimation of wáter surface gains or losses. This product has also proven to be a key tool for linking these transformations to human activities, generating valuable knowledge about the interactions between natural systems and anthropogenic pressures. This achievement strengthens the objectives set forth by the MapBiomas Agua and RAISG initiative, standing out as a collaborative effort that contributes to both scientific understanding and the development of sustainable management strategies. The public availability of the data encourages its use by researchers, decision-makers, and local communities, who can use it to plan actions for the conservation, restoration, and rational use of water resources.

# 8. References


Barkhordarian, A., Saatchi, S.S, Behrangi, A., Loikith, PC. and Mechoso, C.R. 2019. A Recent Systematic Increase in Vapor Pressure Deficit over Tropical South America. Sci Rep 9, 15331.

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

WWF. 2020. Living Planet Report 2020: Bending the curve of biodiversity loss. Almond, R.E.A., Grooten, M., & Petersen, T. (Eds). World Wide Fund for Nature, Gland, Switzerland.