MAPBIOMAS
[AGUA]

PERÚ

Glacier Surface ATBD

Collection 4

Version 1

June 2026

Content

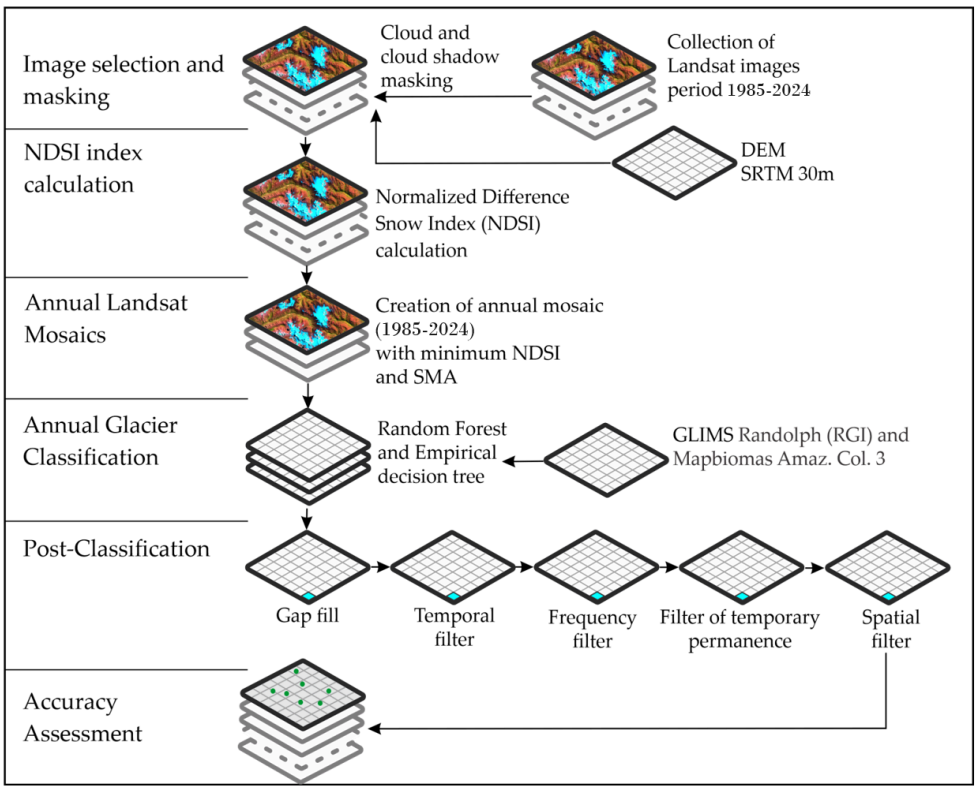
1. Introduction 3
   1.1. Study Area 4
2. Image mosaics 4
3. Classification 5
   3.1. Classification variables 6
   3.2. Reference maps 7
4. Post-classification 7
   4.1. Gap Filling 8
   4.2. Temporal filter 8
   4.3. Base year filter 10
   4.4. Frequency filter 12
   4.5. Masking of gaps 13
   4.6. Temporary stay filter 14
   4.7. Spatial filter 14
   4.8. Integration with other MapBiomas Amazonia classes 15
5. Final considerations 15
6. References 16

1. Introduction

Mapping glaciers is difficult and dangerous due to the remoteness and inaccessibility of the terrain and the challenges of conducting extensive fieldwork(Racoviteanu et al., 2009)In this context, satellite data obtained through remote sensing provide valuable information about glaciers and associated landforms. It is important to note that the judicious selection of spectral bands is fundamental in mapping glacial features.(Huang et al., 2021)Glacial features in the Amazon basin, in the Andes biome, have been delineated using selected bands from the Landsat program data(Philip & Ravindran, 1998)The vast majority of high mountains with glaciers have experienced an accelerated reduction of glaciers in recent decades.(Baťka et al., 2020; Turpo Cayo et al., 2022)

Mountain glaciers, particularly in the tropics, are currently rapidly decreasing in surface area.(Turpo Cayo et al., 2022)and are considered good indicators of climate change, due to their relatively fast response time to disturbances in climatic variables such as precipitation, air temperature and atmospheric humidity(Kaser & Osmaston, 2002)Many tropical glaciers, such as those in Peru and Bolivia, are critical buffers against reduced rainfall during the dry season. Over 99% of tropical glaciers are located in the Andes Mountains of South America, including Venezuela, Colombia, Ecuador, Peru, Bolivia, Chile, and Argentina. Much of the research has been conducted using remote sensing data, as the difficult terrain presents challenges for extensive fieldwork.(Veettil & Kamp, 2019).

In this document, we present a description of the methodology applied for mapping the tropical glaciers of Peru. The complete process was divided into 6 stages (Figure 1).



Figurea 1– *Methodological flowchart for obtaining glacier cover maps*

1.1. Study Area

The study area (Figure 2) was defined based on the Randolph glacier inventory(RGI Consortium, 2017, p. 0)This has been cut off by the Raisg boundary, over which a 1.5 km buffer has been applied to consider the classification area.

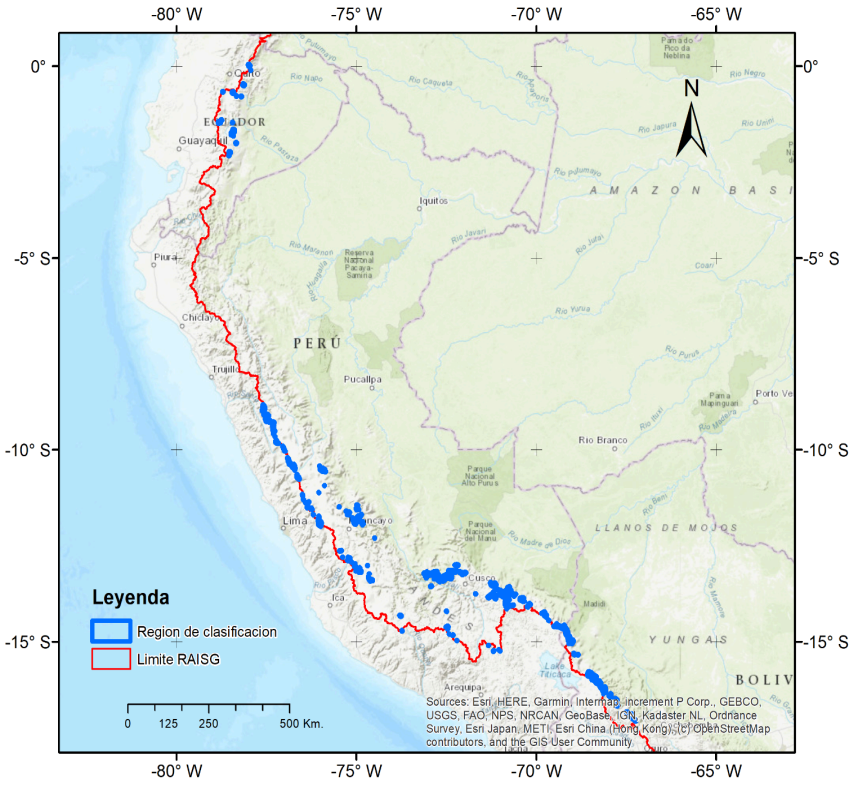


Figure 2 *–Work area*

2. Image mosaics

The classification of the cross-cutting theme “Glaciers” used Landsat image mosaics specifically generated for glacier mapping. These mosaics included images with minimum annual glacier area, based on the minimum NDSI pixel quality (Figure 3).

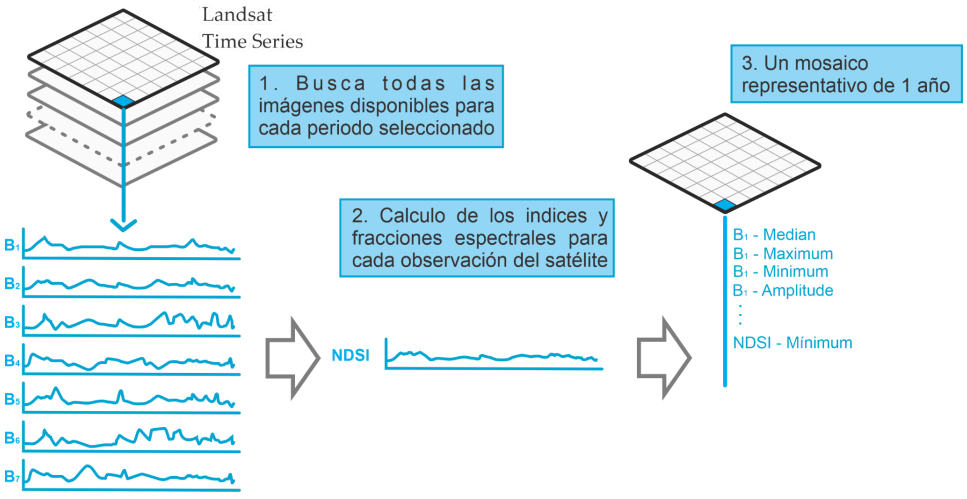


Figure 3 *–Creation of annual mosaics for glacier mapping*

3. Classification

The classification of the Landsat mosaics was carried out entirely on the Google Earth Engine platform, based on an empirical tree (Figure 4).

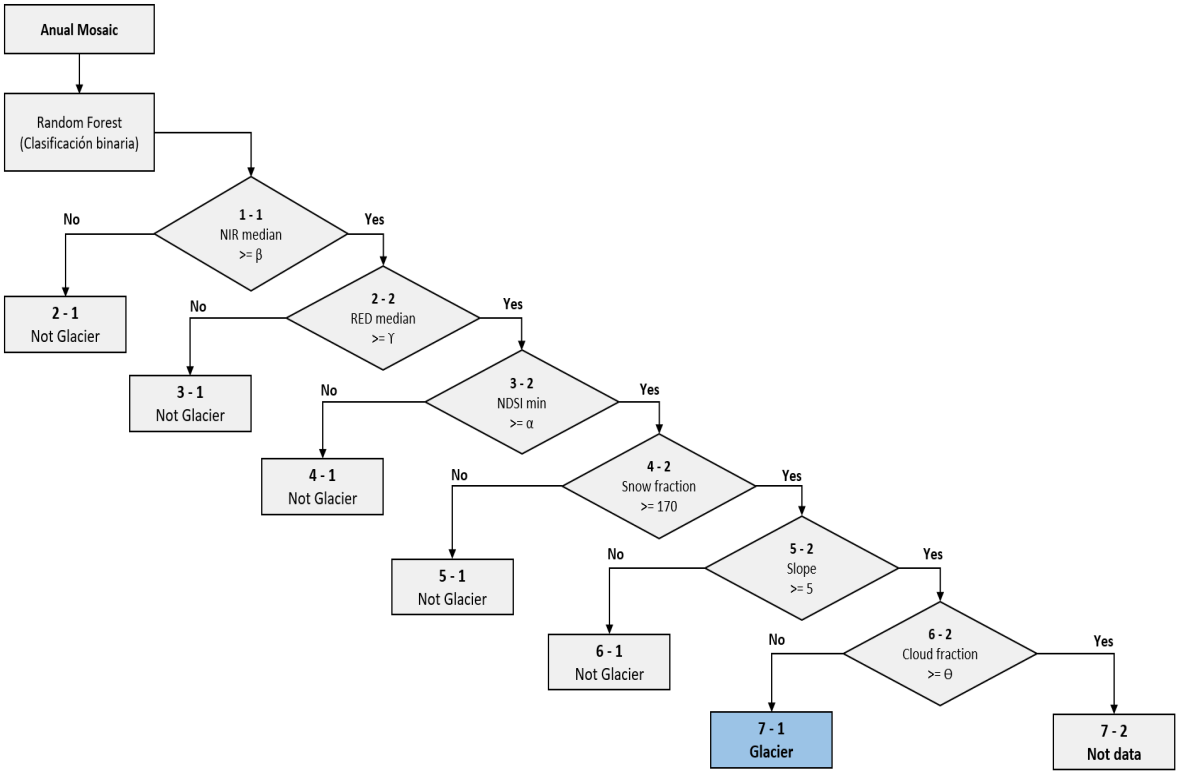


Figure 4 –*Empirical tree combined with Random Forest for glacier classification.*

The glacier classification system implements a hybrid approach that combines machine learning with decision rules specific to different satellite sensors. The classification process begins with the generation of cloud-free annual mosaics for each year of study, using Landsat images from different sensors (L5, L7, L8, and L9) depending on the time period.

The Random Forest classification is refined with a set of hierarchical decision rules specific to each sensor type. For Landsat 5 and Landsat 7, median (1-1) NIR reflectance thresholds are applied, with values close to 0.2114, while for Landsat 8 and Landsat 9 these values are close to 0.1730. For the median (2-2) RED thresholds, the estimated value for Landsat 5 and Landsat 7 is 0.2497, while for Landsat 8 and 9 the estimated value is 0.2304.

The final identification of glaciers is based on a combined condition that considers the minimum NDSI (3-2), the snow fraction (4-2), and the slope (5-2). This dual classification allows us to capture data for glaciers with high near-infrared spectral reflectance, such as those on steep terrain with significant glacial cover. Finally, the classification process applies a condition based on cloud fraction values (6-2), where pixels exceeding 170 (Θ) are classified as "no data" and discarded from the final classification, ensuring that only pixels with ideal observations are included.

3.1. Classification variables

The classification algorithm uses different classification variables, organized into four main categories that capture different aspects of the spectral and topographic signature of glacial surfaces.

The first category comprises six spectral bands specifically used for glacier detection, corresponding to average dry-season values. This temporal selection is strategic because cloud interference and snow cover are minimized during this period, maximizing the visibility of glacial surfaces and allowing for better spectral discrimination between glaciers and other land cover types.

The second category includes the use of spectral indices, highlighting the temporal variations of the NDSI (Normalized Difference Snow Index) calculated for the dry season, the wet season, and the minimum value. The use of this index is fundamental for identifying snow and glacial ice surfaces.

The third category incorporates the topographic slope variable derived from the SRTM digital elevation model, which is crucial given that tropical glaciers are typically found in high mountain terrain with steep topography, providing an additional geomorphological criterion for classification.

Finally, variables derived from Spectral Mixture Analysis (SMA) are included. SMA is a processing technique that decomposes the reflectance of each pixel into fractions of different pure materials, providing quantitative information about the sub-pixel composition of each area. The snow fraction is particularly relevant for distinguishing glacial surfaces from other types of land cover in high-mountain environments, while the cloud and shade fractions help identify and filter pixels with compromised observations.

Table 1 *–Bands used for classification*

Type	Name	Formula	Description	Reducer[1]				
				Median	Median dry	Median wet	Mine	Max
Band	blue	B1 (L5 y L7); B2 (L8)	Blue visible spectrum		X			
	green	B2 (L5 y L7); B3 (L8)	Green visible spectrum		X			
	red	B3 (L5 y L7); B4 (L8)	Red visible spectrum		X			

	nir	B4 (L5 y L7); B5 (L8)	Near infrared		X			
	swir1	B5 (L5 y L7); B6 (L8)	Shortwave infrared 1		X			
	swir2	B7 (L5); B8 (L7); B7(L8)	Shortwave infrared 2		X			

Table 2 *–spectral indices used for classification*

Type	Name	Formula	Description	Reducer[1]				
				Median	Median dry	Median Wet	Mine	Max
Indices	NDSI	$(\frac{green-swir1}{green+swir1})$	Normalized Differential Snow Index		X			

Table 3 *–SMAs used for classification*

Type	Name	Formula	Description	Reducer[1]				
				Median	Median dry	Median Wet	Mine	Max
Fractions	Cloud fraction	SMA	Cloud fraction		X			
	Snow fraction	SMA	Fraction of snow		X			

Use:[1]The reducer is based on the NDSI index, using percentiles; the 75th and 25th percentiles for the wet and dry seasons respectively.

3.2. Reference maps

The study area (Figure 5) was defined based on the Randolph glacier inventory(RGI Consortium, 2017, p. 0)It was also inspected and some missing glaciers were added.

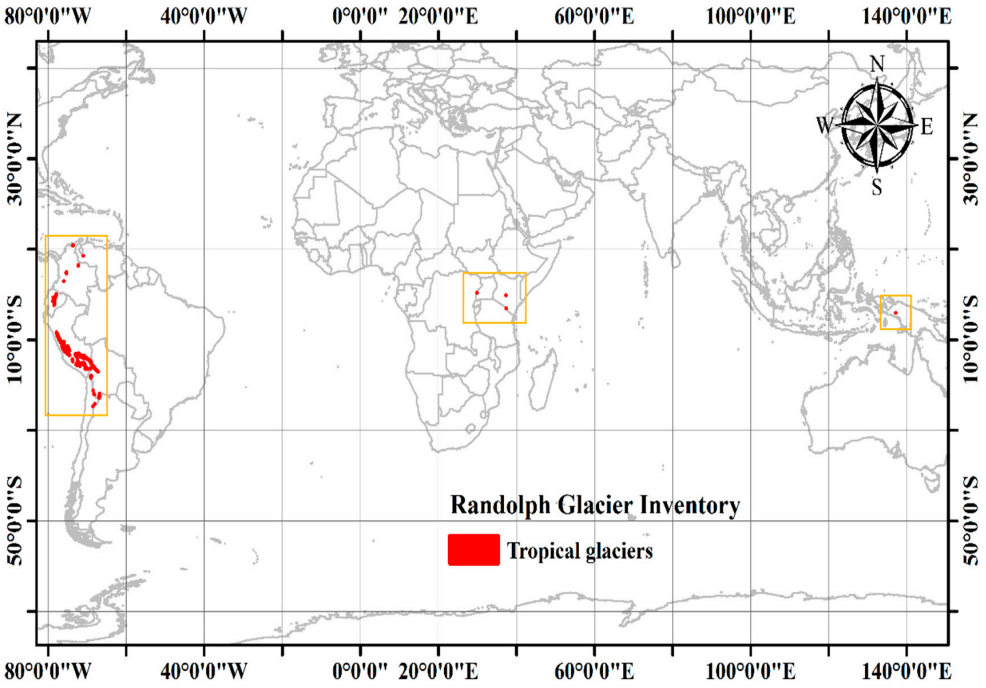


Figure 5 *–Distribution of tropical glaciers; source:(RGI Consortium, 2017, p. 0; Veettil & Kamp, 2019)*

4. Post-classification

Given the pixel-based nature of the classification model and the processing of an extensive time series, a chain of post-classification filters was implemented. This process included the sequential application of fill, temporal, spatial, frequency, and overestimation filters.

4.1.Gap Filling

The filter sequence begins with a post-processing technique called Gap Fill, used to fill gaps in glacier classification data. The main objective of this filter is to correct missing or misclassified pixels in a time series. This process is fundamental in glacier time series analysis, as it addresses one of the main challenges in remote sensing: the presence of pixels with no data or misclassified pixels due to factors such as cloud cover, topographic shadows, or limitations in satellite image capture.

The Gap Fill function implements a bidirectional approach to gap filling. This algorithm operates in two complementary phases that ensure the completeness of the analyzed time series.

During the first phase, the algorithm traverses the time series from the initial evaluation year to the final year. During this forward traversal, when the algorithm encounters a pixel without data in a specific year, it assigns it the value from the immediately preceding year.

The second phase implements a reverse process, working from the final year back to the initial year. Pixels that remained without data

after the first phase receive values from the nearest subsequent year. This bidirectional strategy ensures that data gaps only persist if a pixel lacks valid information across the entire time series.

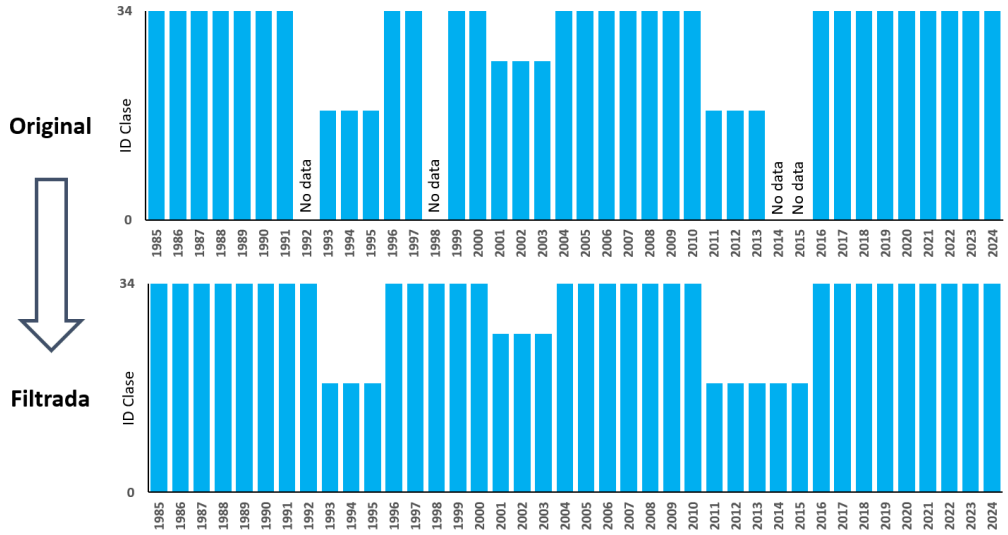


Figure 6 *–Filtro Gap Fill*

4.2. Temporal filter

The second filter implements a process that is critical in refining glacier classifications after the Gap Fill process. This filter focuses on correcting erroneous or inconsistent classifications that may occur in individual pixels throughout the time series, applying a temporal consistency logic based on multi-year analysis windows.

This filter is especially important for removing "noise" from classifications—that is, abrupt and illogical changes in glacial cover that do not correspond to natural processes. For example, a pixel classified as a glacier that suddenly appears as non-glacial for a year and then reverts to glacial cover likely represents a classification error rather than an actual change in cover.

The filter implements three types of time analysis windows: 3, 4, and 5 years. Each window applies specific logic to detect and correct classification inconsistencies.

The 3-year window detects situations where a pixel has the same classification in the preceding and following years, but a different classification in the middle year. This situation indicates a classification error, so the algorithm corrects the middle year to maintain temporal consistency.

The 4-year window extends this logic to detect gaps between two consecutive years. If a pixel has the same rating for year t_1 y $t_4$but different classifications in the t_2 y t_3The algorithm corrects both

intervening years. This is useful for correcting classification errors that persist for two consecutive years.

The 5-year window handles complex cases where there are three consecutive years with incorrect classifications between two years with the same correct classification. This window is particularly useful for correcting prolonged periods of cloud cover or systematic problems in satellite imagery.

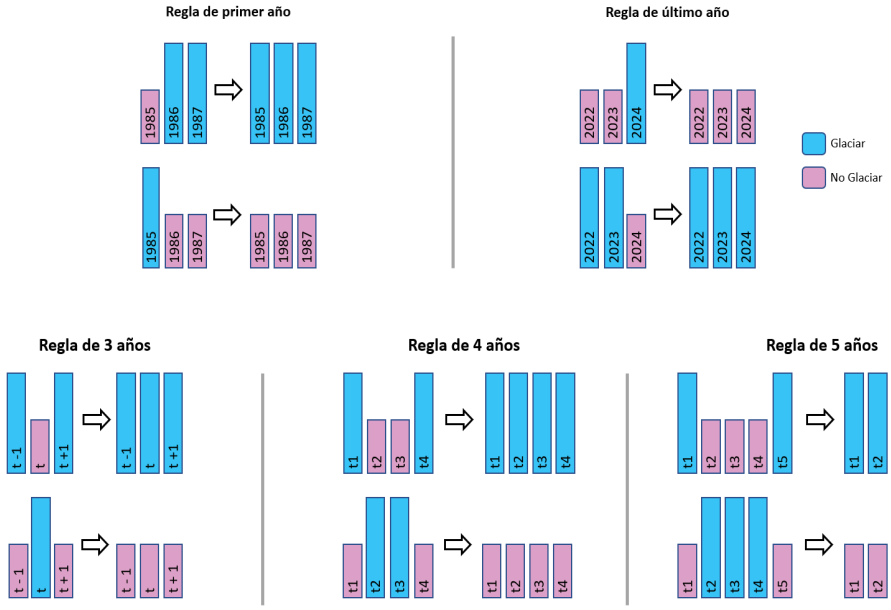


Figure 7 *–Temporal Filter*

The filter also implements specific functions to handle the extreme years of the time series. For the first year, the filter checks if the following two years have the same ranking; if so, it adjusts the ranking of the initial year to maintain consistency. This process is applied similarly to the last year of the time series, verifying its consistency with the two preceding years.

This special treatment is necessary because these years lack a complete temporal context, requiring a tailored filtering logic. This ensures that the entire time series maintains consistency from beginning to end.

The order in which these rules are applied is not arbitrary, as the temporal filter begins with smaller windows, processing the outermost years, allowing the simplest and most obvious errors to be corrected, and then processing the larger windows, leaving the more complex cases for later applications, maximizing the effectiveness of the filtering, while minimizing the risk of overestimation.

4.3. Base year filter

The third filter aims to correct the overestimation and underestimation present in the base years of the glacial time series. Its application is structured in two parts: correction using an auxiliary

persistence layer and time sequence analysis. The first part consists of constructing an auxiliary layer of persistent pixels classified as glaciers. To do this, the three years following the year to be corrected are evaluated: if a pixel is detected as a glacier in at least two of those three years, it is incorporated into the auxiliary layer. This layer serves as a filler for the information gaps present in the evaluated year, based on the premise that if a pixel is consistently detected as a glacier in subsequent years, it is highly probable that it was also detected as such in the previous base year.

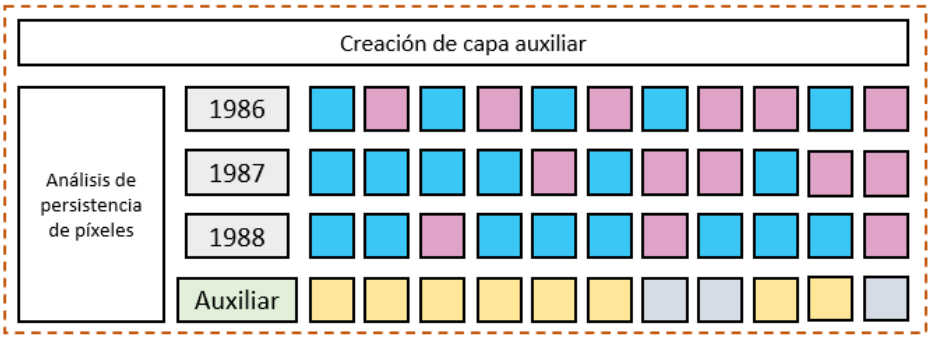


Figure 8 –*Creation of an auxiliary layer*

The correction of the evaluated year is applied considering three possible cases, derived from the combination between the classification of the year to be corrected and the state of the pixel in the auxiliary layer:

- If the pixel for the year to be corrected is classified as non-glacial and the auxiliary layer identifies it as a persistent glacier, the pixel is reclassified as a glacier. The underlying logic is that consistent glacier detection in subsequent years supports its presence in the base year, so mapping it as non-glacial is attributed to an underestimation resulting from the limited availability of images for the base years of our assessment.
- If the pixel for the year to be corrected is classified as a glacier and the auxiliary layer also identifies it as a persistent glacier, the pixel is preserved as a glacier without modification.
- If the pixel for the year to be corrected is classified as a glacier but the auxiliary layer identifies it as non-persistent, the pixel is retained as a glacier. This decision is based on the criterion of respecting the coverage classified for the base year due to insufficient mosaics; its potential overestimation cannot be confirmed solely with the auxiliary layer and requires further analysis, which is addressed in the following section.

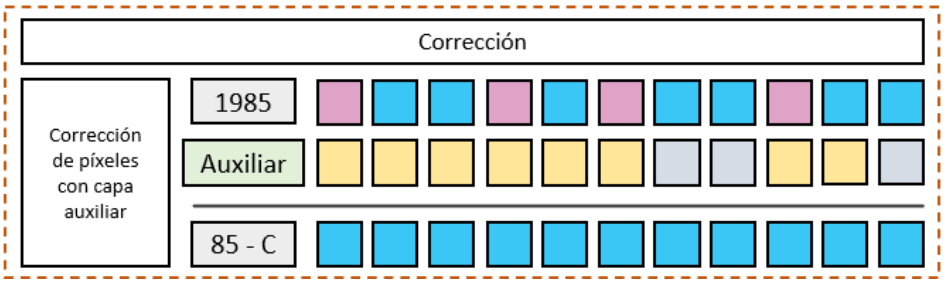


Figure 9 –*Base year correction with auxiliary layer*

The result of this process is the layer called**85-C**, corresponding to the corrected classification of 1985.

The second section specifically addresses the third case from the previous section: pixels classified as glaciers in the base year, but identified as non-persistent in the auxiliary layer. For these pixels, a detection frequency analysis is performed in the years following the evaluated year.

Starting with layer 85-C, each pixel is evaluated to see if it is detected as a glacier in two or more of the subsequent years considered. If it meets this condition, the pixel is retained as a glacier within the time series, confirming its validity as glacial cover. If, on the other hand, the detection frequency as a glacier is less than two years, the pixel is reclassified as non-glacial, since the lack of temporal continuity supports its overestimation. This overestimation originates from the limitations of the base year mosaics: given the limited availability of satellite imagery during that period, the mosaic composition sometimes includes pixels from the wet season, which may exhibit seasonal snow cover that is incorrectly mapped as permanent glacier. As the time series progresses, image availability gradually increases, allowing for better mosaic composition and more precise discrimination between permanent glacier and transient snow cover, thus reducing this type of error in the more recent years of the series.

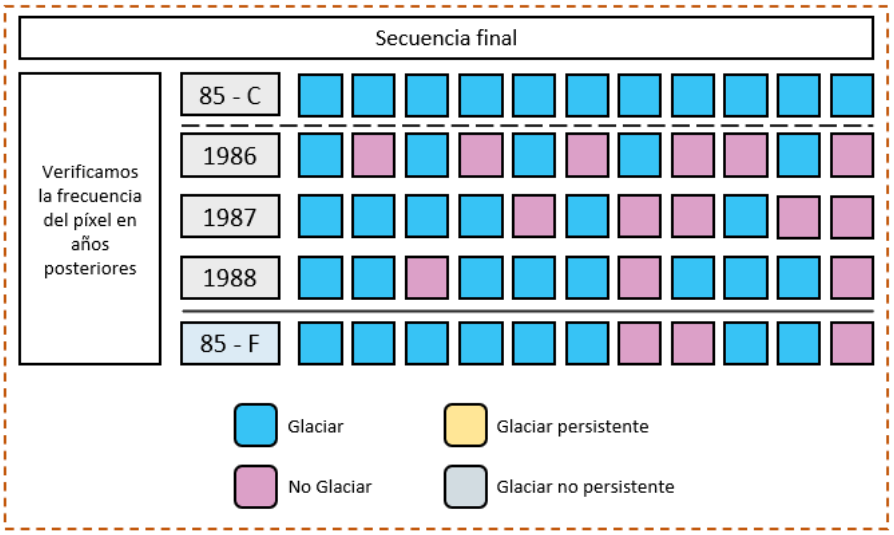


Figure 10 *–Pixel sequentiality analysis*

These results are complemented by the application of subsequent filters.

4.4. Frequency filter

The fourth filter in the processing chain is the frequency filter. This filter operates on the principle that stable land cover, such as glaciers, should maintain a consistent presence over time. Unlike previous filters that work with specific time windows, the frequency filter analyzes the entire time series to identify and consolidate the most persistent classifications.

The filter is effective at distinguishing between real changes in glacial cover and sporadic misclassification errors. By establishing minimum frequency thresholds, the algorithm can determine with high confidence which pixels truly represent permanent glaciers versus those that have been misclassified.

The established threshold (70%) defines the minimum percentage of years in which a pixel must be classified as glacial or non-glacial to be definitively considered as such. In other words, if a pixel has been classified as glacial in more than 70% of the years, it is definitively classified as glacial. The same principle applies to the non-glacial classification.

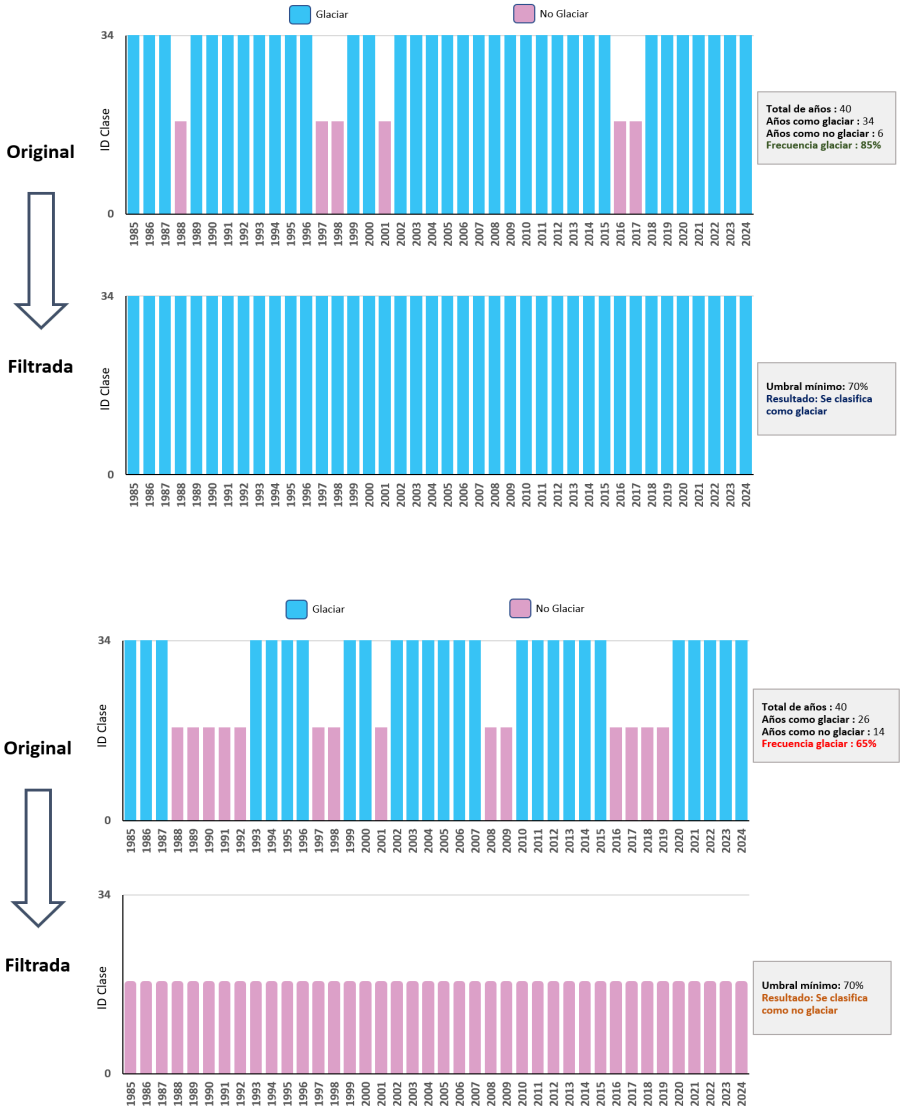


Figure 11 *–Frequency Filter*

4.5. Masking of gaps

The masking process addresses a specific but critical problem in glacier mapping: spectral confusion between glaciers and bodies of water. This masking filter uses data from the MapBiomasWater collection to identify and remove areas where lagoon glaciers have been misclassified as glaciers.

This confusion is common in high mountain environments because both glacial ice and water can exhibit similar spectral responses under certain conditions, especially when glacial sediments are present or when lakes are partially frozen.

The filter processes each year independently, integrating two complementary data sources: the data are the glacier classifications, resulting from all previous filters applied, and the water masks, specifically identified by MapBiomasAgua.

The filter implements a simple exclusion logic: if a pixel is classified as both glacier and water, it is reclassified as water, removing this data from the glacier map; if a pixel is classified only as glacier, it remains as glacier in the final result; and finally, if a pixel is classified only as water, it is no longer considered in the final glacier map. This reduces false positive readings in glacier mapping, eliminating a common source of error.

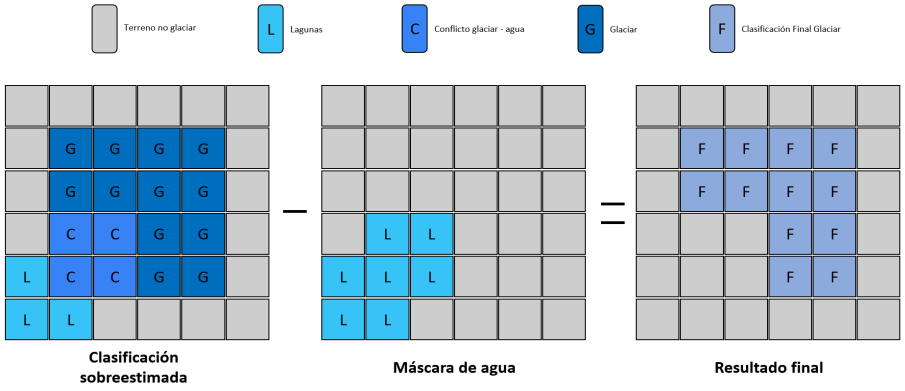


Figure 12 *–Gap masking sequence*

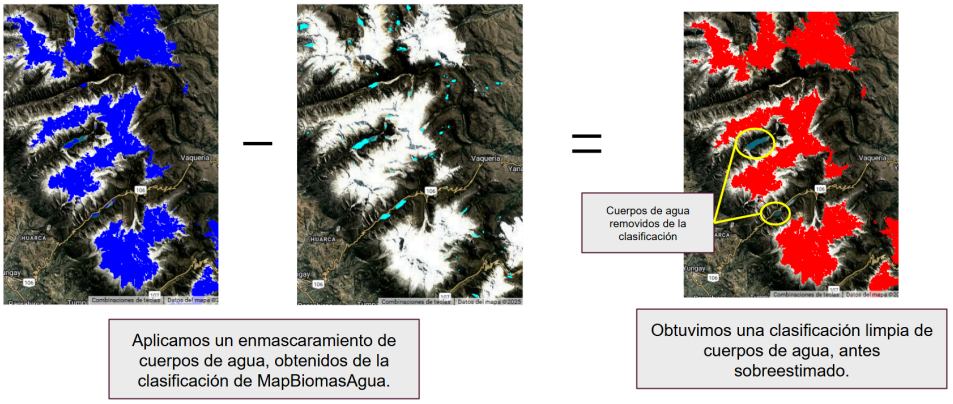


Figure 13 *–Effect of masking gaps*

4.6. Temporary stay filter

The time-duration filter complements the frequency filter. While the frequency filter evaluates the statistical persistence of classifications throughout the entire time series, the time-duration filter implements a logic of irreversibility in glacial cover changes.

This filter operates on the glaciological principle that once an area loses its glacial cover, it is extremely unlikely to form a glacier again in the short term. This premise is based on the understanding that glacier formation requires specific climatic conditions sustained over extended periods, something that does not occur on timescales of decades.(Marangunic, 2016).

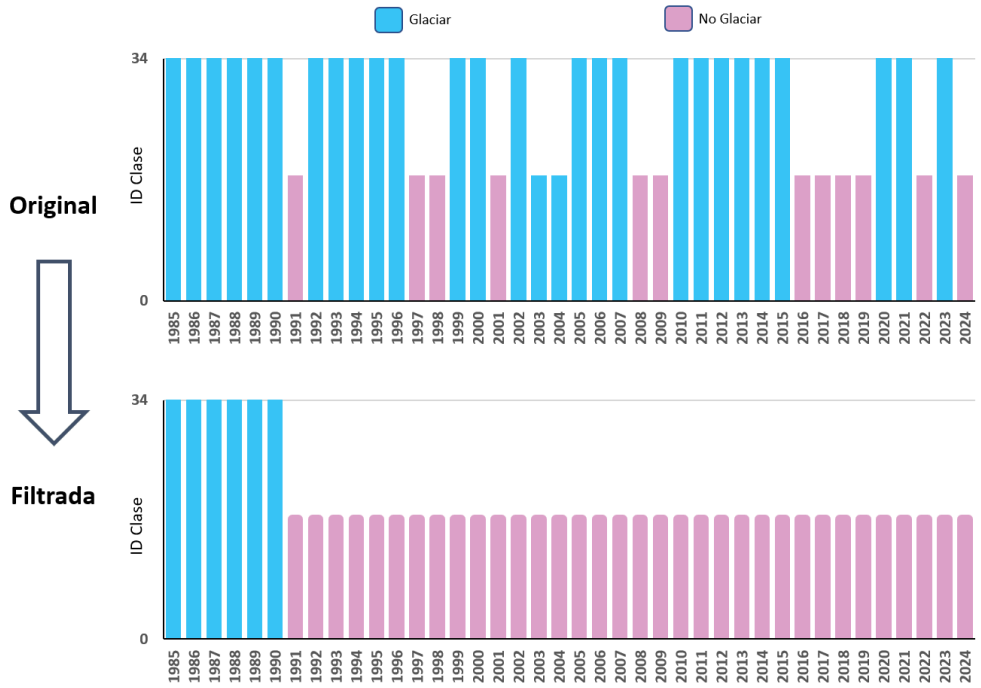


Figure 14 *–Temporary Stay Filter*

4.7. Spatial filter

The spatial filter works exclusively with the spatial dimension of the images, analyzing the coherence and connectivity of the pixels classified in each individual study year.

This filter addresses a common problem in satellite image classification: the presence of isolated pixels or small groups of pixels that are classified differently from their immediate surroundings. These isolated pixels often represent classification errors rather than actual terrain features.

The spatial filtering process is implemented in two main stages, this for each year of the time series.

During the first stage, the filter uses the "connectedPixelCount" function to calculate how many glacier pixels are connected to each other. The second stage of the filter involves grouping pixels; the filter marks all groups with fewer than 5 connected pixels. These small groups are likely to be errors because real glaciers are rarely that small.

For mapping glaciers in the Andes, where glaciers can have complex shapes due to mountainous topography, the spatial filter is valuable, as it helps to eliminate misclassifications caused by topographic shadows, undetected small clouds, or variations in snow reflectance, while preserving the actual structure of the glaciers.

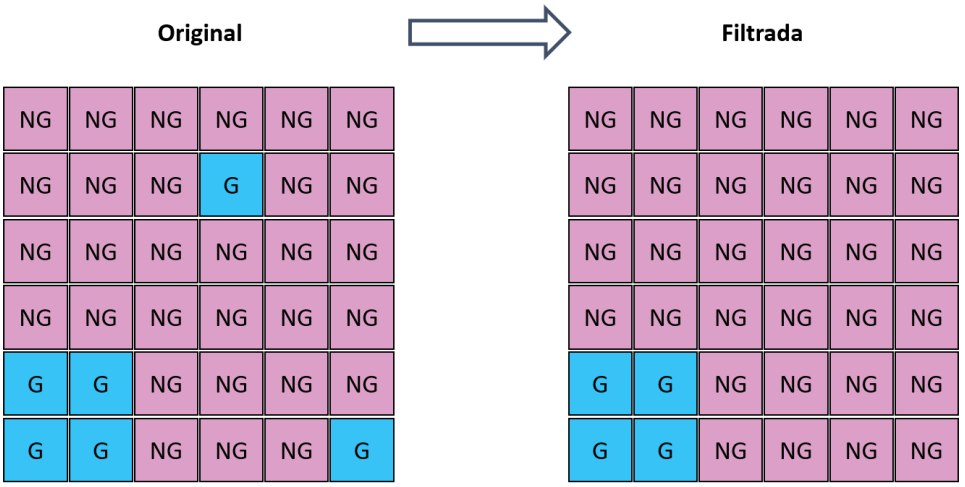


Figure 15 *–Space filter effect*

4.8. Integration with other MapBiomas Amazonia classes

After applying the filter sequence in the Post-classification stage, the cross-cutting themes and general maps for each biome are integrated. This integration depends on a series of specific hierarchical rules that assign a prevalence order to each class. The result of this stage is the annual land cover and land use maps for the entire Amazon region.

5. Final considerations

During the development of Collection 3, an excessive glacial loss and an apparent abrupt retreat in the base years of the time series were identified. Initially, one of the hypotheses considered to explain this anomaly was radiometric calibration errors between the Landsat 5 TM, Landsat 7 ETM+, and Landsat 8 OLI sensors.

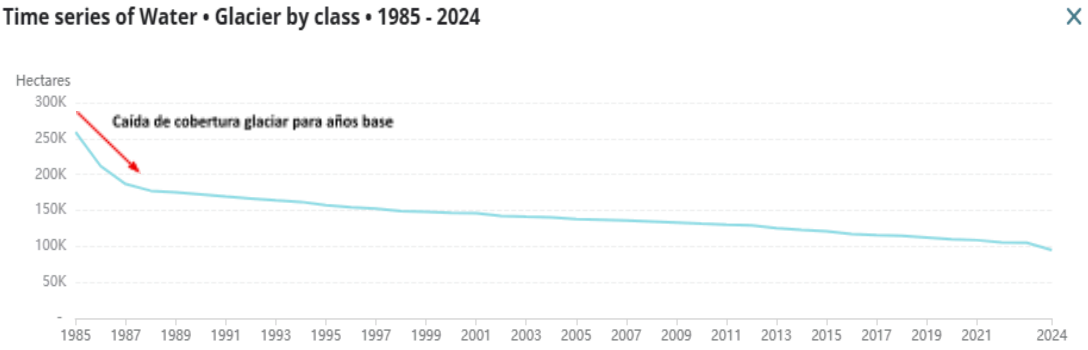


Figure 16 *–C3 statistics of glaciers*

The hypothesis indicated that these errors manifest as an overestimation of radiance and reflectance values, the magnitude of which varies according to the spectral band and gradually attenuates towards the mid-nineties. For glacial and nival surfaces, the most critical bands (1, 2, and 3) show overestimations of 0.72%, 3.87%, and 4.09% respectively, which artificially amplifies their high reflectance in the visible spectrum and

would inflate multiband spectral indices, leading to an overestimation of the glacial area in temporal analyses that include this period (Micijevic et al., 2016).

However, this hypothesis was discarded for Collection 4 as the primary cause of the anomaly observed in Collection 3. The images processed in Google Earth Engine come from the Landsat surface reflectance collections (Collection 2, Level-2), which incorporate radiometric corrections and cross-calibration between sensors, ensuring the consistency and linearity of spectral values over the time series (USGS, 2021; Vermote et al., 2016; Dwyer et al., 2018). Consequently, the calibration errors described by Micijevic et al. (2016) do not constitute a source of bias in the present analysis.

The most significant explanation for the overestimation detected in the base years lies in the quality of the mosaic composition. For the base years of the time series, the limited availability of satellite imagery severely restricts the construction of optimal mosaics. In the absence of sufficient seasonal scenes, the composition algorithms incorporate pixels from the wet season, characterized by high precipitation and seasonal snow cover. These wet-season pixels introduce transient snow cover that is incorrectly classified as permanent glacier, leading to an overestimation of the glacier area in the base years.

This phenomenon is clearly observable in the analysis of the composition of the mosaics.**Figure 17**This problem is illustrated in the Coropuna area of the Ampato mountain range: the pixels marked in red correspond to images from wet months, for a mosaic composed of only 20 images. The areas identified in green and orange mostly correspond to pixels from the dry season, highlighting the heterogeneity in the mosaic's composition and the impact that incorporating wet pixels has on estimating glacial cover.

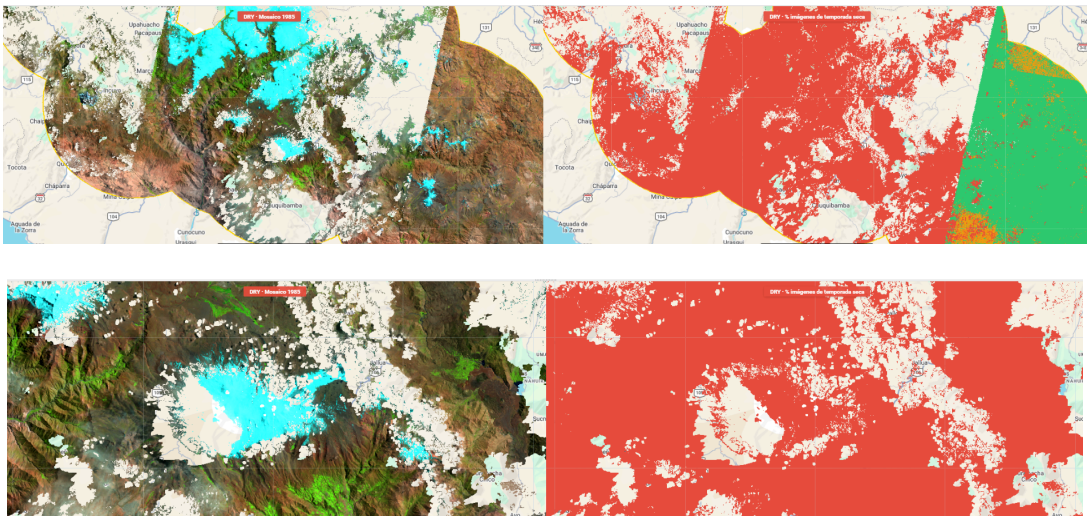


Figure 17 *–Pixel origin analysis for the composition of the 1985 mosaic – Coropuna Zone*

As the time series progresses, the availability of images gradually increases, allowing for the construction of higher-quality mosaics with a greater proportion of dry-season pixels and more precise discrimination between permanent glaciers and seasonal snow.**Figure 18**This shows that even in the 2000 mosaic, pixels from the dry season predominate, reflecting a cleaner glacial cover, free from the influence of wet snow. This progressive improvement in mosaic quality is what generates the apparent sharp drop in glacial cover between the base years and subsequent years observed in Collection 3.

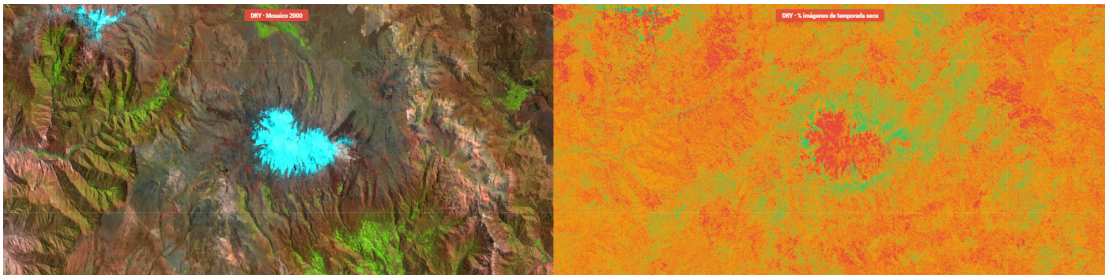


Figure 18 –*Pixel origin analysis for the composition of the 2000 mosaic – Coropuna Zone*

For Collection 4, this problem was addressed by applying the base year correction filter, which aims to estimate glacier cover closer to the actual figures for the base years, ensuring traceability and temporal consistency with subsequent years in the series. As a result, the glacier cover statistics for Collection 4 no longer exhibit the abrupt drop observed in Collection 3, correcting the overestimation associated with the inclusion of wet season pixels in the base year mosaics.

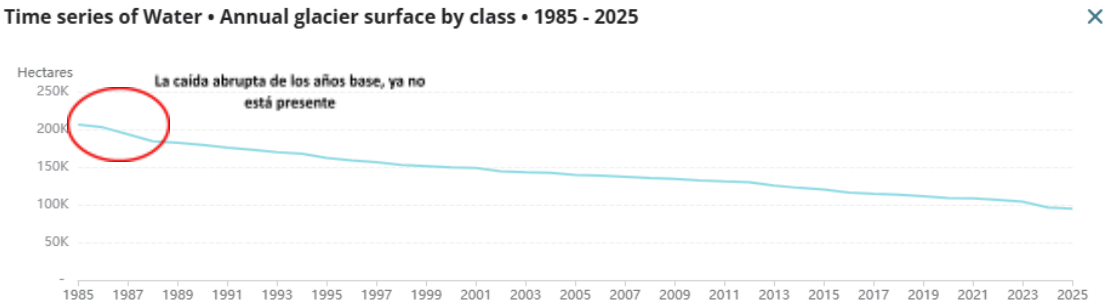


Figure 19 –*C4 statistics of glaciers.*

6. References

Baťka, J., Vilímek, V., Štefanová, E., Cook, S. J., & Emmer, A. (2020). Glacial Lake Outburst Floods (GLOFs) in the Cordillera Huayhuash, Peru: Historic Events and Current Susceptibility.*Water*, *12*(10), 2664. https://doi.org/10.3390/w12102664

Dwyer, J. L., Roy, D. P., Sauer, B., Jenkerson, C. B., Zhang, H. K., & Lymburner, L. (2018). Analysis ready data: Enabling analysis of the Landsat archive. *Remote Sensing*, 10(9), 1363. https://doi.org/10.3390/rs10091363

Huang, L., Li, Z., Zhou, J. M., & Zhang, P. (2021). An automatic method for clean glacier and nonseasonal snow area change estimation in High Mountain Asia from 1990 to 2018. *Remote Sensing of Environment*, *258*, 112376. https://doi.org/10.1016/j.rse.2021.112376

Kaser, G., & Osmaston, H. (2002).*Tropical glaciers*. Cambridge Univ. Press. https://onggem.files.wordpress.com/2011/02/kaser-osmaston-2002-tropical-glaciers1.pdf

Marangunic, C. (2016). Glaciers and mountain ecosystems: The important tasks still to be done.*Journal of Glaciers and Mountain Ecosystems*, *1*. https://doi.org/10.36580/RGEM.I1.11-19

Micijevic, E., Haque, Md. O., & Mishra, N. (2016). Radiometric calibration updates to the Landsat collection. *Earth Observing Systems XXI*, *997299720D*. https://doi.org/10.1117/12.2239426

Philip, G., & Ravindran, K. (1998). Glacial mapping using landsat thematic mapper data: A case study in parts of gangotri glacier, NW himalaya. *Journal of the Indian Society of Remote Sensing*, *26*(1-2), 29-34. https://doi.org/10.1007/BF03007337

Racoviteanu, A. E., Paul, F., Raup, B., Khalsa, S. J. S., & Armstrong, R. (2009). Challenges and recommendations in mapping of glacier parameters from space: Results of the 2008 Global Land Ice Measurements from Space (GLIMS) workshop, Boulder, Colorado,

USA. *Annals of Glaciology*, *50*(53), 53-69. https://doi.org/10.3189/172756410790595804

RGI Consortium. (2017). Randolph Glacier Inventory – A Dataset of Global Glacier Outlines: Version 6.0. En *GLIMS Technical Report*. https://www.glims.org/RGI/00_rgi60_TechnicalNote.pdf

Turpo Cayo , E. Y. , Borja , M. O. , Espinoza-Villar , R. , Moreno , N. , Camargo , R. , Almeida , C. , Hopfgartner , K. , Yarleque , C. , & Souza , C. M. (2022). Mapping Three Decades of Changes in the Tropical Andean Glaciers Using Landsat Data Processed in the Earth Engine. *Remote Sensing*, *14*(9), 1974. https://doi.org/10.3390/rs14091974

U.S. Geological Survey (USGS). (2021). *Landsat Collection 2 Level-2 Science Product Guide*. U.S. Department of the Interior. https://www.usgs.gov/landsat-missions/landsat-collection-2-level-2-science-products

Veettil & Kamp. (2019). Global Disappearance of Tropical Mountain Glaciers: Observations, Causes, and Challenges. *Geosciences*, *9*(5), 196. https://doi.org/10.3390/geosciences9050196