MAPBIOMAS
[AGUA]

PERÚ

Methodological Document

Snow Surface

Collection 4

Version 1

June 2026

Content

1. Introduction

Snow is an important component of the hydrological cycle at both global and regional scales, and understanding the characteristics of the temporal and spatial evolution of the snowpack is of great importance in the fields of climate research and water resource management (Luo et al., 2022).

Snow is one of the most dynamic and influential components of the Earth's cryosphere. As a form of solid precipitation, it acts as a seasonal reservoir of freshwater, storing large volumes during the cold season and gradually releasing them during the thaw, thus regulating river flow and aquifer recharge in mountain regions (Schaefli, 2015).

In the tropical Andes of Peru, snow-capped mountain ranges form the origin of vital river systems that supply millions of people, so understanding the spatio-temporal dynamics of snow cover is important for integrated water resource management (SENAMHI, 2018) SENAMHI. (2018).

Quite apart from its hydrological role, snow exerts a decisive influence on high mountain ecosystems and the global climate system, since its albedo reduces the absorption of solar radiation and directly affects regional temperatures (Brown & Robinson, 2011) Brown, R. D., & Robinson, D. A. (2011).

Given this reality, accurate and systematic mapping of snow surface is a necessity; however, the high spatio-temporal variability of snow, combined with the inaccessibility of high mountain environments, severely limits traditional measurement approaches (Dietz et al., 2012).

In this context, satellite remote sensing has become the most effective tool for the continuous monitoring of snow cover at regional and global scales. Sensors such as MODSI have made it possible to generate time series spanning several decades with near-daily coverage, while higher-resolution platforms such as Sentinel-2 and Landsat offer more refined detection capabilities, essential for mapping the heterogeneity of the snowpack in high mountain areas (Gascoin et al., 2019; Aalstad et al., 2020).

In this document, we present a description of the methodology applied for mapping the snow surface in Peru.

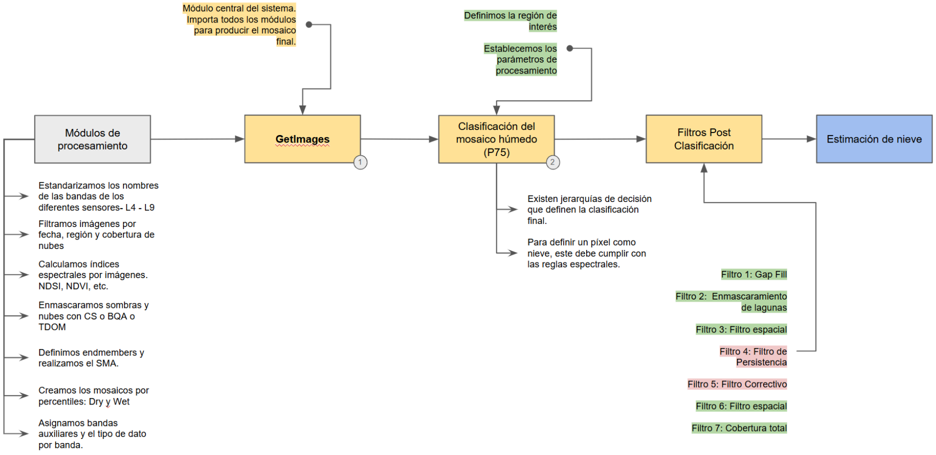


Figure 1– *Methodological flowchart for obtaining snow cover maps*

1.1. Study Area

The study region was initially defined based on the glacier inventory of *Randolph Glacier Inventory* (RGI Consortium, 2017), whose boundaries are suitable for characterizing glacial presence. However, since the objective of this work is to map snow cover, this initial reference framework proved insufficient.(Figure 2).

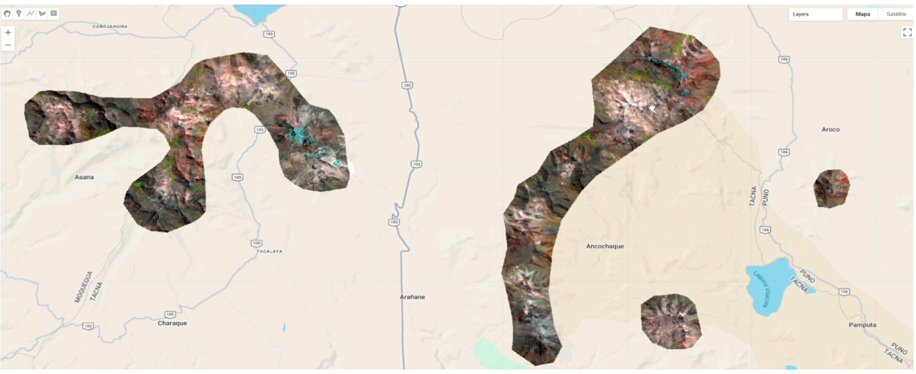


Figure 2:*Glacier cover for the year 2000 in the RGI regions.*

The mapping procedure uses mosaics constructed from the pixels with the highest NDSI value within the time series, aiming to capture the maximum possible snow cover extent for each year. When the RGI regions were directly applied to these mosaics, the snow cover was truncated at its edges (Figure 3), which led to the need to expand the processing regions.

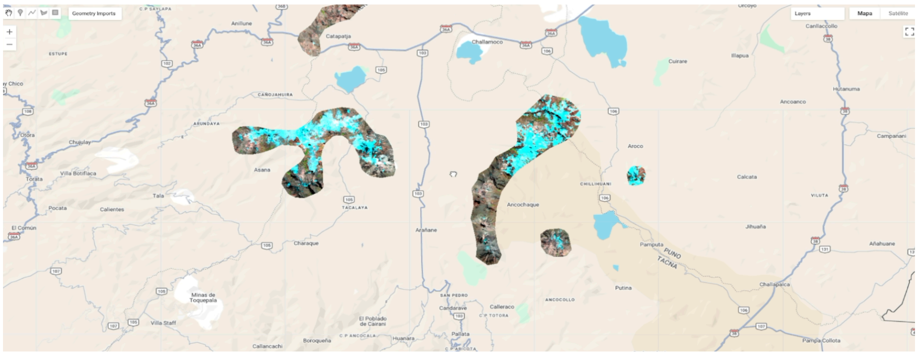


Figure 3:*Snow cover for the year 2000 in the RGI regions.*

To do this, it is necessary to expand our processing regions and accurately define how far the snow extends in the Andes for different years.

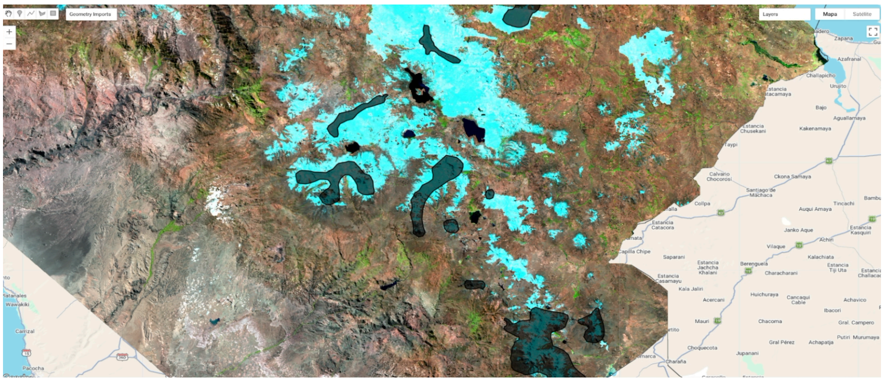


Figure 4:*Snow extent for compared to RGI regions.*

To this end, a buffer of approximately 50 km was applied over the original RGI regions (Figure 5), with the purpose of ensuring that the mosaics covered the entire possible snow extent in different years.

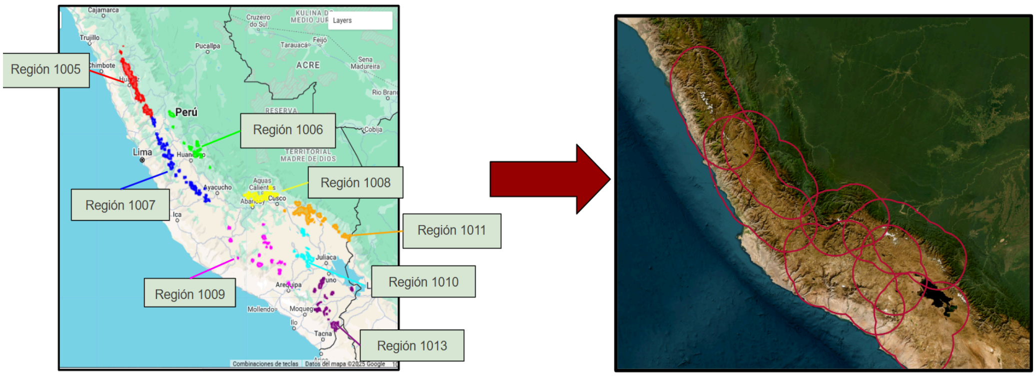


Figure 5:*Expansion of processing areas.*

Additionally, the processing regions were adjusted by integrating two complementary layers: the Andean biome layer and the SENAMHI snow presence altitudinal layer (Bomshoms Calvelo et al., 2018). The latter, derived from the analysis of 35 meteorological stations distributed nationwide, establishes that snowfall in Peru occurs most frequently during the summer months, although it is present throughout the year and above 3400 m a.s.l.

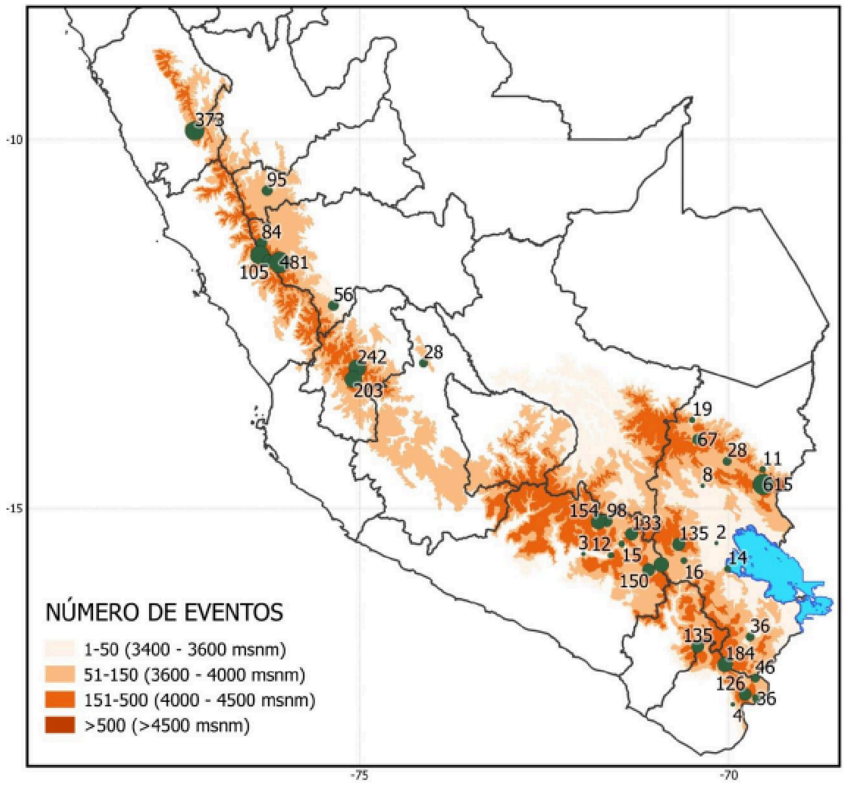


Figure 6:*Number of events at the national level according to SENAMHI and the altitudinal distribution of these, Bomshoms Calvelo et al. (2018).*

The combination of these layers allowed for a more precise delimitation of areas with snow potential. The final working region corresponds to that delimited by the Andean biome, trimmed and complemented by the altitudinal layer of snow presence, and was subdivided into 45 processing subregions to optimize computational performance in the cloud.

Figure 7:Work area for mapping persistent snow in Peru.

2. Image mosaics

The snow classification used Landsat image mosaics specifically generated for snow mapping. These mosaics were constructed from pixels belonging to the wet season. To create a mosaic, the percentile distribution of pixels from the filtered images for the analysis region is used.

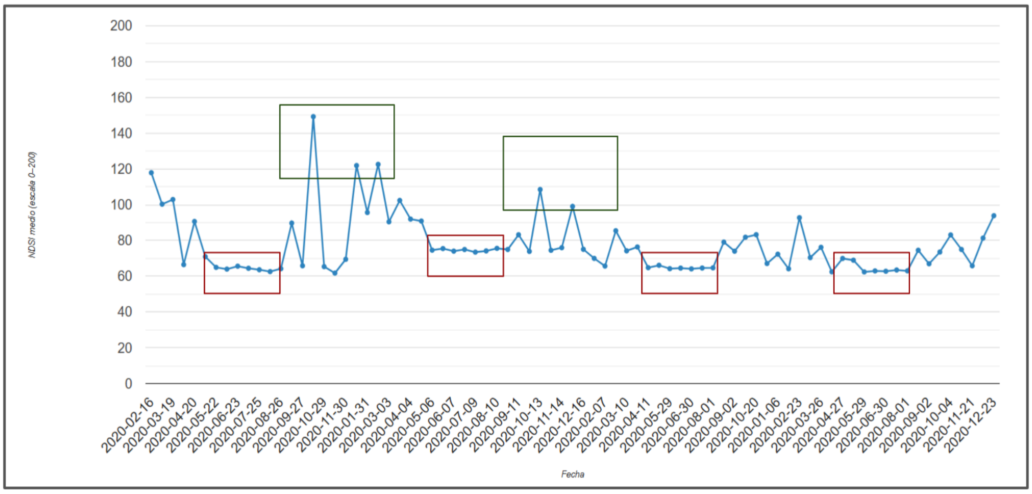


Figure 8:Analysis of NDSI index pixel values for wet season mosaic composition.

This pixel analysis responds to a percentile distribution, where the pixels belonging to the 75th percentile are taken for the composition of the wet season mosaics.

The wet season mosaic captures the conditions of an estimated maximum annual snow cover and is constructed by selecting images where the NDSI value is greater than or equal to the 75th percentile of the annual distribution. The median is then calculated pixel by pixel on these filtered images, generating the "wet" bands. The 75th percentile was chosen to strike a balance between maximizing the captured snow cover and minimizing the noise introduced by seasonal clouds.

The percentile behavior in image selection operates pixel by pixel on the temporal distribution of NDSI values. For a pixel to be included in the 75th percentile mosaic as snow cover, it must have been classified as snow in approximately 25% or more of the valid images for the year, corresponding to the statistical position of the 75th percentile in the ordered distribution of values.

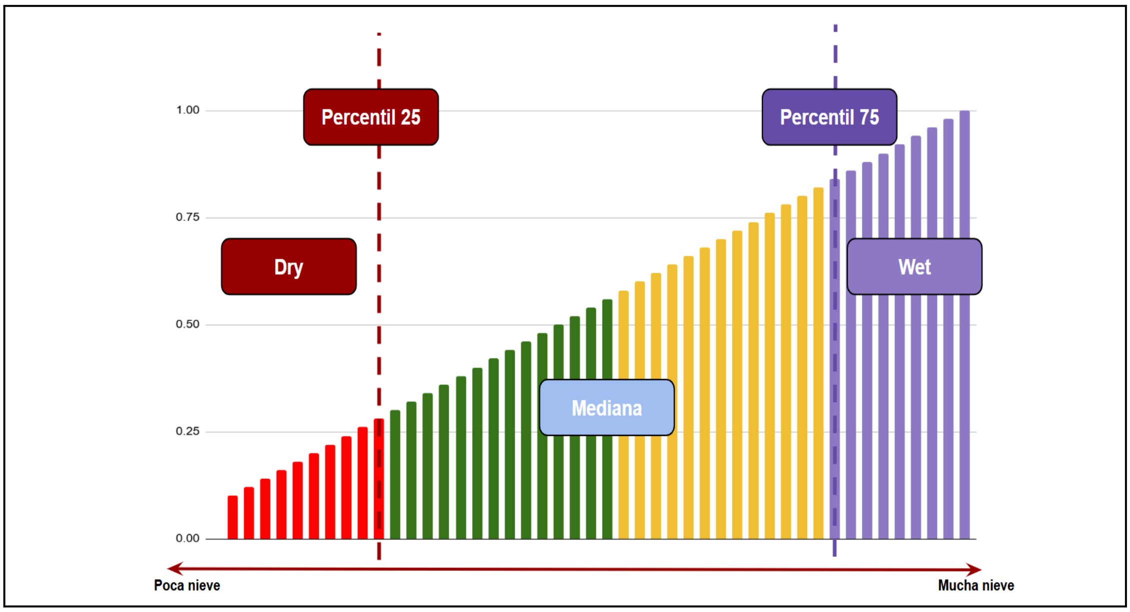


Figure 9:*Construction of wet season mosaics, based on percentile distribution.*

3. Classification

The snow classification process operates on the wet season mosaic (75th percentile of the annual NDSI series), which aims to capture the maximum snow cover of the year. Landsat mosaic classification was performed entirely on the Google Earth Engine platform, based on thresholds (Figure 10). The classification of snow cover from the wet season mosaic does not employ a supervised classifier like Random Forest. Instead, it exclusively uses spectral threshold rules applied directly to the mosaic variables through conditional operations.**where.** This classification process in the methodology responds to the temporal instability of seasonal snow, which prevents the construction of a set of training pixels that are sufficiently stable for the wet period.

A set of sequential conditions is applied to this mosaic, where all pixels start with an initial assignment of "no snow cover," and each threshold rule can overwrite the result of the previous one. The first condition evaluates the reflectance in the near-infrared band of the wet-season mosaic (1-1), discarding as snow cover those pixels that do not exceed the minimum threshold established according to the sensor: values close to 0.2114 for Landsat 5 and Landsat 7, and close to 0.1730 for Landsat 8 and Landsat 9. Similarly, the second condition evaluates the reflectance in the red band of the wet-season mosaic (2-2), applying thresholds of 0.2497 for Landsat 5 and Landsat 7, and 0.2304 for Landsat 8 and Landsat 9, also discarding pixels that do not reach these thresholds.

The definitive identification of snow cover is based on a dual condition that must be met simultaneously (3-2 and 4-2): the minimum annual NDSI value, representing the lowest snow cover recorded for that pixel throughout the year, must be greater than or equal to 100 on the rescaled scale (equivalent to an actual NDSI greater than or equal to 0.0), and the snow fraction derived from Spectral Mixture Analysis must be greater than or equal to 20 percent. Pixels that satisfy both conditions are assigned to the snow presence category.

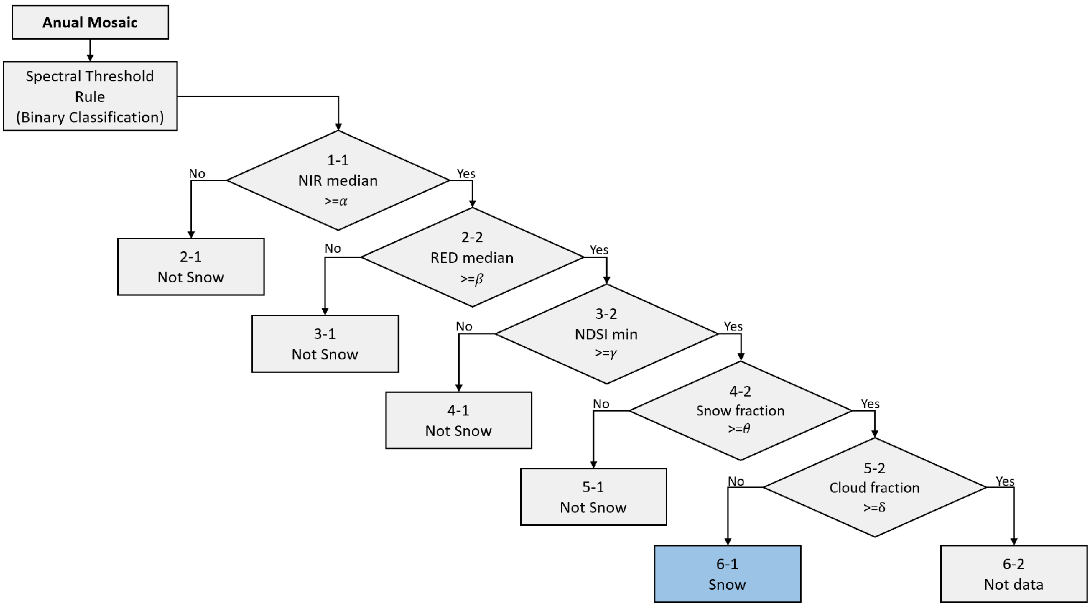


Figure 10 –Empirical tree combined with Random Forest for snow classification.

Finally, as a quality control step, pixels with residual cloud contamination are identified using the cloud fraction derived from the SMA (6-2). Those exceeding the threshold value of 170 on the 0 to 200 scale are classified as “invalid data” and discarded from the final classification, ensuring that only pixels with spectrally reliable observations are included in the snow cover product.

3.1. Classification variables

The classification algorithm uses different classification variables, organized into three main categories that capture different aspects of the spectral signature of snow surfaces.

The first category comprises six spectral bands at their wet season median values (75th percentile of the NDSI).

The second category includes the use of spectral indices, highlighting the temporal variations of the NDSI (Normalized Difference Snow Index) calculated at three different times.

- Ndsi_median_dry: Median dry season NDSI (25th percentile), which captures the spectral condition of the permanent glacier without seasonal snow, or failing that, minimal snow presence.
- Ndsi_median_wet: Median NDSI of wet season (75th percentile), reflecting the maximum snow cover of wet period.
- Ndsi_min: Minimum NDSI value for the entire year. Restrictive threshold relevant for classification.

In the snow cover classification, we did not use topographic slope as a direct discriminating variable, given that the spatial distribution of seasonal snow differs from the morphological pattern of permanent glaciers. Snow can accumulate on low-slope surfaces, which would cause a slope threshold to exclude valid snow-covered areas. Instead, in the post-classification filters, we used the SENAMHI snow presence record, which documents the occurrence of snow events nationwide, from 3400 meters above sea level to higher altitudes. This post-classification layer acts as an altitudinal mask that filters out spectral noise from lower altitudes, mainly in high-altitude rainforest areas, where pixels of dense clouds have spectral similarity to snow and can be misclassified as snow.

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 snow-covered 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.

4. Post-classification

Given the nature of the classification model and the processing of an extensive time series, a chain of post-classification filters was implemented.

4.1. Gap Filling

The post-processing filter sequence begins with Gap Fill, which aims to correct information gaps in the snow cover classification time series. This process is fundamental in remote sensing time series analysis, as it addresses one of the most recurring challenges in the field: the existence of pixels with no data or compromised classification due to cloud cover, topographic shadows, or limitations inherent in satellite imagery.

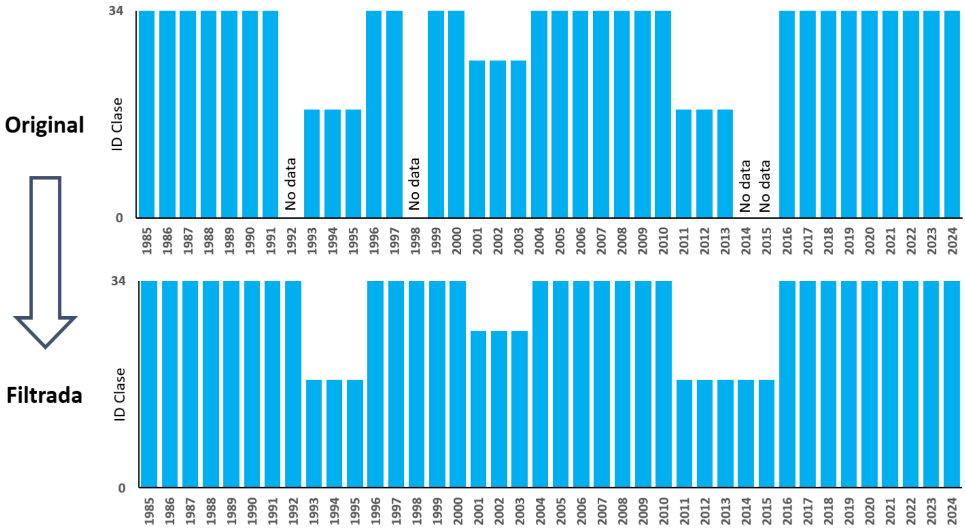
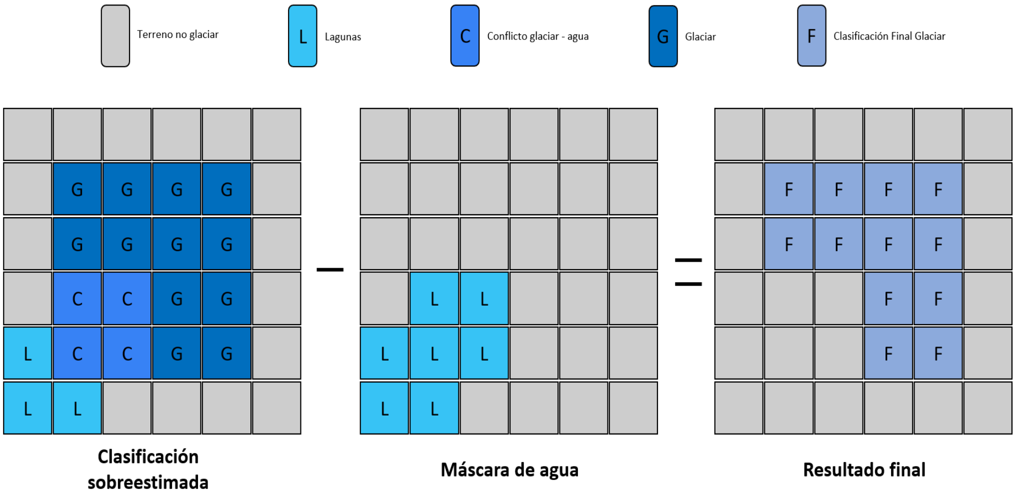


Figure 11 *–Filtro Gap Fill*

The algorithm implements a bidirectional approach that operates in two complementary phases to ensure the completeness of the time series. During the first phase, the algorithm traverses the series from the initial year to the final year. Each time it encounters a pixel without data in a given year, it assigns it the ranked value of the immediately preceding year, propagating the information forward in time. The second phase performs the reverse traversal, from the final year to the initial year, and fills in the pixels that remained without data after the first phase by assigning them the value of the nearest subsequent year with valid information. This bidirectional strategy ensures that gaps only persist in those pixels that lack valid observations across the entire time series, minimizing information gaps without introducing artificial values.

4.2. Masking of gaps

This filter addresses a specific but critical problem in snow cover mapping. Lakes are frequently misclassified as either snow cover or glacier cover because snow and water can exhibit similar spectral responses under certain conditions, such as when lakes are partially frozen. This spectral ambiguity is especially pronounced in high-mountain environments, where both types of cover coexist.



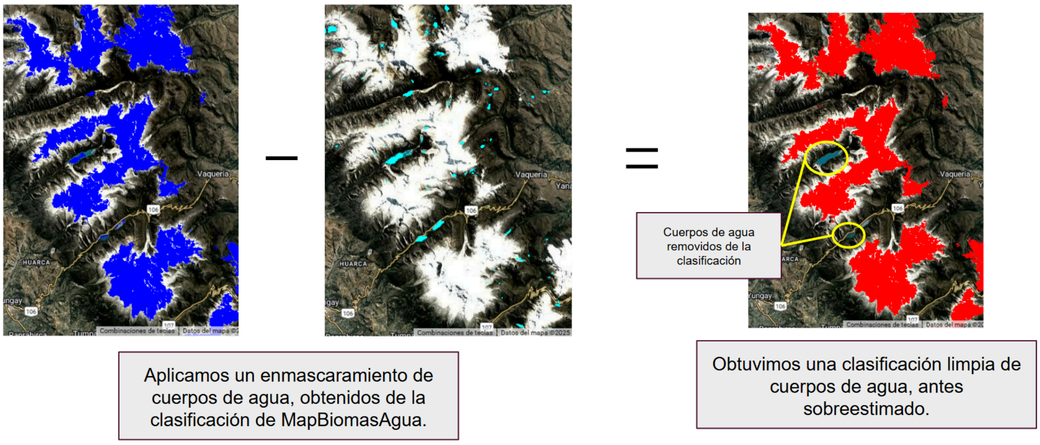
Clasificación sobreestimada

Máscara de agua

Resultado final

__Figure 12__ –Gap masking sequence

To correct this problem, the filter processes data independently each year, integrating two complementary sources: snow cover classifications obtained after applying the previous filters, and annual water body masks derived from the MapBiomas Water collection. The exclusion logic applied is straightforward: pixels classified as snow but actually representing a water body are reclassified as water and excluded from the snow cover map, thus eliminating false positives generated by spectral confusion. Pixels classified as snow and confirmed by mosaic analysis are retained unchanged, and pixels identified exclusively as water are excluded from the final snow cover product. This strategy systematically reduces a recurring source of error in mapping high mountain areas.



Cuerpos de agua removidos de la clasificación

Aplicamos un enmascaramiento de cuerpos de agua, obtenidos de la clasificación de MapBiomasAgua.

Obtuvimos una clasificación limpia de cuerpos de agua, antes sobreestimado.

__Figure 13__ –Effect of masking gaps

4.3. Spatial filter

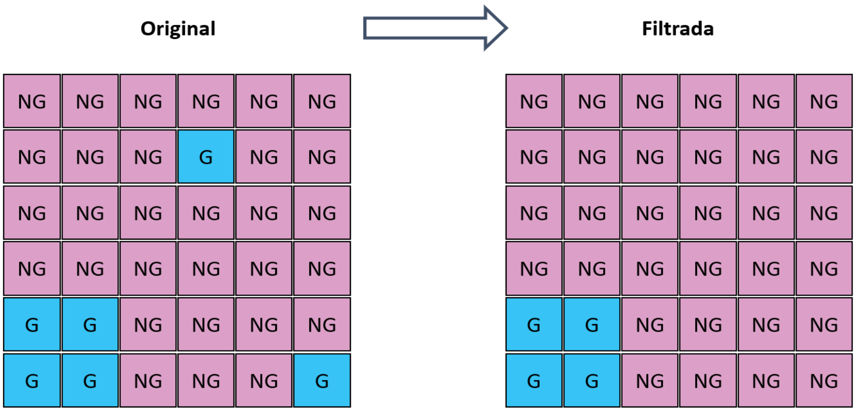
The spatial filter works exclusively with the spatial dimension of the images, analyzing the coherence and connectivity of the pixels classified in each individual study year.

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.

For snow mapping in the Andes, where the snow surface can have complex shapes due to mountainous topography, the spatial filter is valuable, as it helps to eliminate misclassifications caused by topographic shadows and undetected small clouds.



Original

NG	NG	NG	NG	NG	NG
NG	NG	NG	G	NG	NG
NG	NG	NG	NG	NG	NG
NG	NG	NG	NG	NG	NG
G	G	NG	NG	NG	NG
G	G	NG	NG	NG	G

Filtrada

NG	NG	NG	NG	NG	NG
NG	NG	NG	NG	NG	NG
NG	NG	NG	NG	NG	NG
NG	NG	NG	NG	NG	NG
G	G	NG	NG	NG	NG
G	G	NG	NG	NG	NG

Figure 14 –*Space filter effect*

4.4. Persistence filter

The persistence filter evaluates the frequency of each pixel's presence as snow cover across the entire 26-year time series, eliminating pixels classified sporadically that correspond to residual noise or very occasional seasonal snow. For each pixel, the number of years it was classified as snow cover and the total number of years with valid data are counted. The persistence frequency is calculated as the ratio between these two values, expressed as a percentage. In this case, we use a persistence threshold of 35%. This threshold is equivalent to a minimum presence in approximately 9 of the 26 years. The decision rule for application is as follows: If the frequency is greater than 35% (present in more than 9 of the 26 years), the pixel remains as snow cover in the years it was mapped; and if the pixel's presence is less than 35% (less than 9 years), the pixel is reclassified as no snow in all years of the series. The persistence filter does not modify the years in which the pixel was not mapped as snow; we only remove the presence of snow cover in years where the threshold is not exceeded.

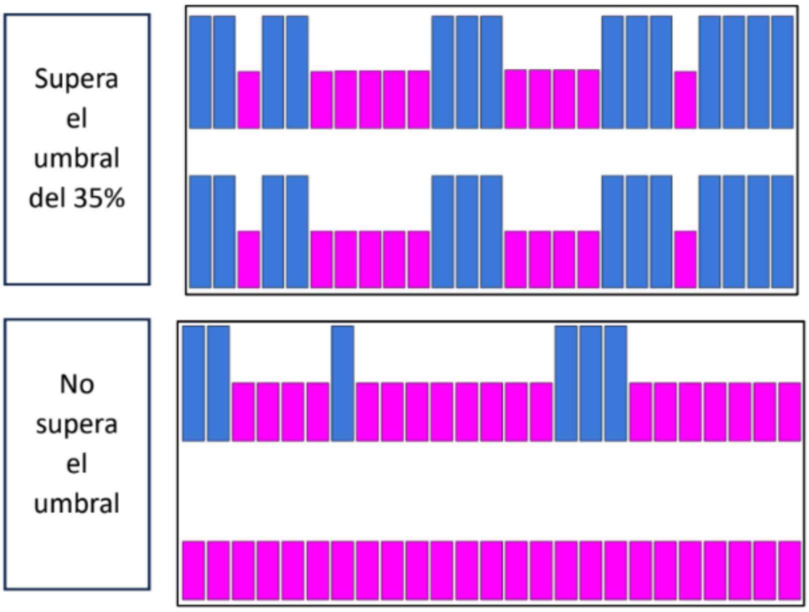


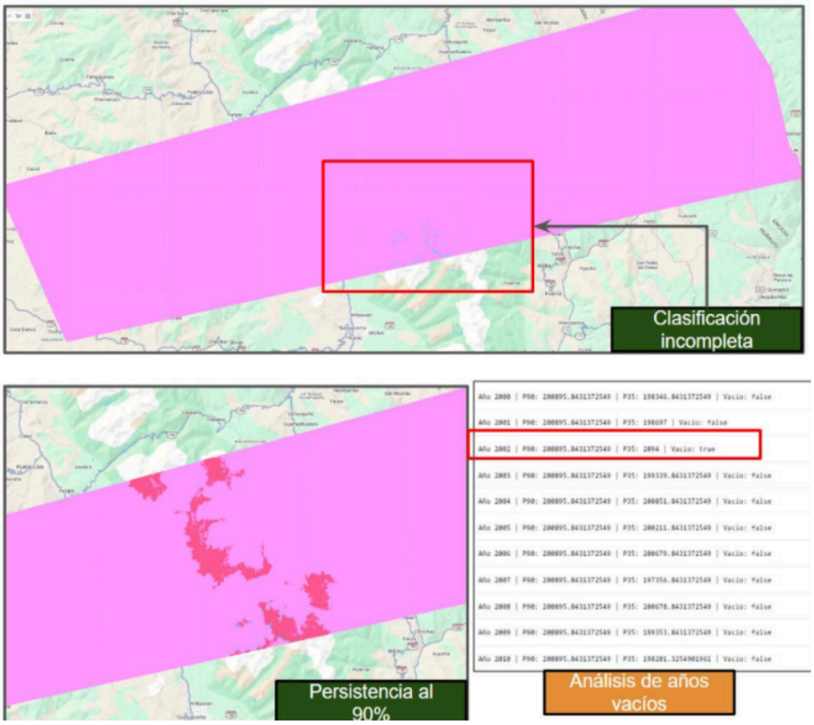
Figure 15 *–Persistence filter*

4.5. Corrective Filter

The presence of years with virtually empty classifications is a result of the lack of valid mosaics due to persistent high cloud cover throughout the wet season. These years show very few or no pixels classified as snow cover, even in areas with confirmed permanent glaciers.

The filter uses as a reference an auxiliary layer of pixels that are 90% persistent over the time series, representing the core of glacial/snow cover present in 23 of the 26 years analyzed. For each year and region, the ratio between the pixels in the 90% persistence layer and the pixels classified as snow cover in that year within the persistence zone is evaluated. If the pixels in the 90% reference layer exceed the pixels classified in the evaluated year by more than 30%, that year is declared as empty.

This condition holds because the pixels in the 90% layer should be contained within the 35% annual classification ($P90 \subseteq P35$). When this containment relationship is reversed—that is, when the permanent core exceeds the annual classification—it indicates the absence of a valid mosaic or poor classification for that year.



```
Año 2000 | P90: 200895.8431372549 | P35: 198346.8431372549 | Vacio: false
Año 2001 | P90: 200895.8431372549 | P35: 198697 | Vacio: false
Año 2002 | P90: 200895.8431372549 | P35: 2094 | Vacio: true
Año 2003 | P90: 200895.8431372549 | P35: 199339.8431372549 | Vacio: false
Año 2004 | P90: 200895.8431372549 | P35: 200051.8431372549 | Vacio: false
Año 2005 | P90: 200895.8431372549 | P35: 200211.8431372549 | Vacio: false
Año 2006 | P90: 200895.8431372549 | P35: 200679.8431372549 | Vacio: false
Año 2007 | P90: 200895.8431372549 | P35: 197356.8431372549 | Vacio: false
Año 2008 | P90: 200895.8431372549 | P35: 200678.8431372549 | Vacio: false
Año 2009 | P90: 200895.8431372549 | P35: 199353.8431372549 | Vacio: false
Año 2010 | P90: 200895.8431372549 | P35: 198201.3254901961 | Vacio: false
```

Figure 16 **–***Corrective filter*

For each year declared as empty, the algorithm searches for the nearest valid years before and after it in the time series. The correction is applied by intersecting (logical AND operation) the classifications of both valid neighboring years: only the pixels identified as snow cover in both years are used to fill the empty year. This conservative criterion ensures that only the minimum glacial/snow cover verified by temporal continuity is filled, avoiding overestimation.

Solo corregimos el año declarado como "vacío" en la serie temporal, el resto de años, se mantiene con su clasificación original, de este modo, mantenemos la consistencia temporal.

Figure 17 **–***Corrective filter action*

5. References

Bomshoms Calvelo, M., Quispe Gutiérrez, N., & Quispe Vega, K. (2018).*Study* *of the frequency of snowfall in Peru*National Meteorology and Hydrology Service of Peru (SENAMHI).

Brown, R. D., & Robinson, D. A. (2011). Northern Hemisphere spring snow cover variability and change over 1922–2010 including an assessment of uncertainty. The Cryosphere, 5(1), 219–229.

Dietz, A. J., Kuenzer, C., Gessner, U., & Dech, S. (2012). Remote sensing of snow – A review of available methods. International Journal of Remote Sensing, 33(13), 4094–4134.

Gascoin, S., Grizonnet, M., Bouchet, M., Salgues, G., & Hagolle, O. (2019). Theia Snow collection: High-resolution operational snow cover maps from Sentinel-2 and Landsat-8 data. Earth System Science Data, 11(2), 493–514.

Luo, Jianfeng, Dong, Chunyu, Lin, Kairong, Chen, Xiaohong, Zhao, Liqiang, & Menzel, Lucas. (2022).*Mapping forest snow cover using optical remote sensing, machine learning, and time-lapse photography*. *Remote Sensing of Environment*. https://doi.org/10.1016/j.rse.2022.113017

SENAMHI. (2018). Study of the frequency of snowfall in Peru. National Meteorology and Hydrology Service of Peru.

Schaefli, B. (2015). Projecting hydropower production under future climates: a guide for decision-makers and modelers based on a review of published studies. WIREs Water, 2(4), 311–340.