# Colombia's Central Eastern Andes Páramo Ecosystem Land Cover Monitoring: A Disturbance Temporal Trajectory Assessment for Protected Areas

## MASTER OF SCIENCE IN GEOINFORMATICS ENGINEERING

## TESI DI LAUREA MAGISTRALE IN INGEGNERIA GEOINFORMATICA

Author: **Daniel Alberto Aguirre Nieto**

Student ID: 10776457
Advisor: Dr. Giovanna Venuti
Co-advisor: Dr. France Gerard
Academic Year: 2022-23

# Abstract (ENG)

The Páramo is a high-elevation ecosystem found only in the tropical Americas. It is characterised by rich biodiversity, plant endemism, and plays a key role in carbon storage and water regulation. Anthropogenic disturbance from agriculture, grazing, and mining have historically threatened this fragile ecosystem. Colombia hosts nearly half of the world's total Páramo surface areas and has delimited 37 different protected Páramo Complex areas since 2011, as a measure to counteract their decline. This decline is significant in the central Eastern Andes, which hosts the majority of the country's population and forms a central hub of economic activity. This Thesis focused on 3 Paramo Complexes (*Guerrero*, *Rabanal y Río Bogotá* and *Altiplano Cundiboyacense*) that were sub-divided into 6 Protected Areas (GUE, RRB, ACB_A, ACB_B, ACB_G, ACB_R). The disturbance due to agriculture was assessed over 47 years using Landsat and Sentinel-2 multispectral images, from a strictly land-cover perspective. The assessment revealed that between 1977 and 2024, Agriculture-Grass areas have increased in most of the protected areas with some of them being historically dominated by this land cover: GUE: **34,1%** to **41,4%**; RRB: **9,9%** to **16,3%**; ACB_A: **78,7%** to **74,2%**; ACB_B: **69,9%** (year 1988) to **79,2%**; ACB_R: **55,4%** to **69,2%**; ACB_G: **27,0%** to **31,9%**. Change analysis detected dynamic agriculture gain-loss behaviour, with a clear deceleration of Agriculture-Grass gain. A spatial-temporal Index to determine agriculture disturbance was proposed and calculated (GUE: **40,52** RRB: **12,85**; ACB_A: **74,82**; ACB_B: **75,02**; ACB_R: **71,84**; ACB_G: **39,96**). The Index should be interpreted as a representative—and optimistic lower bound— persistent Agriculture-Grass land cover percentage of the given area for the 47-year period. The findings of this study should be used to form a multidisciplinary approach to Páramo ecosystem monitoring, in order to fully understand the consequences of agricultural disturbance.


**Key-words:** Páramo, Ecosystem monitoring, Land cover change, Land cover temporal trajectory, Landsat, Sentinel-2, Google Earth Engine.

# Abstract (ITA)

Il Páramo è un ecosistema di alta quota che si trova solo nelle Americhe tropicali. È caratterizzato da una ricca biodiversità, endemismi vegetali e svolge un ruolo chiave nello stoccaggio del carbonio e nella regolazione dell'acqua. I disturbi antropici derivanti da agricoltura, pascolo ed estrazione mineraria hanno storicamente minacciato questo fragile ecosistema. La Colombia ospita quasi la metà delle aree di superficie totale del Páramo del mondo e ha delimitato 37 diverse aree protette del Complesso del Páramo dal 2011, come misura per contrastare il loro declino. Questo declino è significativo nelle Ande centro-orientali, che ospitano la maggior parte della popolazione del paese e costituiscono un fulcro centrale dell'attività economica. Questa tesi si è concentrata su 3 Complessi del Páramo (Guerrero, Rabanal y Río Bogotá e Altiplano Cundiboyacense) che sono stati suddivisi in 6 Aree Protette (GUE, RRB, ACB_A, ACB_B, ACB_G, ACB_R). Il disturbo dovuto all'agricoltura è stato valutato in 47 anni utilizzando immagini multispettrali Landsat e Sentinel-2, da una prospettiva strettamente di copertura del suolo. La valutazione ha rivelato che tra il 1977 e il 2024, le aree agricole-erbacee sono aumentate nella maggior parte delle aree protette, alcune delle quali sono state storicamente dominate da questa copertura del suolo: GUE: dal 34,1% al 41,4%; RRB: dal 9,9% al 16,3%; ACB_A: dal 78,7% al 74,2%; ACB_B: dal 69,9% (anno 1988) al 79,2%; ACB_R: dal 55,4% al 69,2%; ACB_G: dal 27,0% al 31,9%. L'analisi del cambiamento ha rilevato un comportamento dinamico di guadagno-perdita dell'agricoltura, con una chiara decelerazione del guadagno agricolo-erbaceo. È stato proposto e calcolato un indice spazio-temporale per determinare la perturbazione agricola (GUE: 40,52 RRB: 12,85; ACB_A: 74,82; ACB_B: 75,02; ACB_R: 71,84; ACB_G: 39,96). L'indice dovrebbe essere interpretato come un limite inferiore rappresentativo e ottimistico della percentuale di copertura persistente di terreni agricoli e pascoli dell'area data per il periodo di 47 anni. I risultati di questo studio dovrebbero essere utilizzati per formare un approccio multidisciplinare al monitoraggio dell'ecosistema di Páramo, al fine di comprendere appieno le conseguenze della perturbazione agricola.

**Parole chiave:** Páramo, Monitoraggio degli ecosistemi, Cambiamento della copertura del suolo, Traiettoria temporale della copertura del suolo, Landsat, Sentinel-2, Google Earth Engine.

iv

¡Error! Utilice la pestaña Inicio para aplicar Heading 1 al texto que desea que aparezca aquí.

# Contents

¡Error! Utilice la pestaña Inicio para aplicar Heading 1 al texto que desea que aparezca aquí.

viii

# Document Structure

The present document is structured by chapters in the following manner:

**Chapter 1** introduces the different abbreviations used within the document and presents the main objectives of the Thesis.

**Chapter 2** aims to present the reader with a review of the state of the art and the methods, covering all relevant theoretical foundations that have supported the development of the Thesis.

**Sub-Chapter 2.1** presents a summary of what constitutes a Páramo ecosystem focusing on a land cover perspective, its location, characteristics, as well context on its importance. This Sub-Chapter also provides an overview of land cover monitoring efforts in the Colombian context and the role of this Thesis in contributing to that monitoring. Also, it highlights the importance of continuous multidisciplinary monitoring for a more holistic understanding of the health of the Páramo ecosystems, beyond the perspective of landcover.

**Sub-Chapter 2.2** focuses on the remote sensing basics related to satellite-based land cover monitoring. It includes an explanation of the physical principles behind the instruments used to capture the data, their resolution, factors influencing the processing of their acquisitions, and labelling of the available data products. It then presents the different satellite missions with the sensor characteristics chosen for this Thesis.

**Sub-Chapter 2.3** covers the aspects behind the data procurement and characteristics from a historic context and the theory behind the pre-processing methods necessary to obtain a suitable product for analysis. It includes a brief explanation of the different levels at which data is made available for end users, as well as the Google Earth Engine platform and its innovative approach. It also includes the fundamentals for the pre-processing methods used in this Thesis including radiometric corrections, geometric corrections, data visualisation strategies, data masking, and data integration.

**Sub-Chapter 2.4** explains the theory behind digital image classification and its extension to satellite acquisitions, the machine learning basics and the classification algorithms used in this Thesis, the sampling design for robust statistical inference, as

well as examples of different Land Use Land Cover (LULC) products with their class schemas at a global, continental-regional, national and national-regional level.

**Sub-Chapter 2.5** provides an overview of the accuracy assessment and uncertainty estimation for LULC classification products, the importance of the error matrix, the description of the different accuracy metrics and their unbiased estimators based on best practices found in literature.

**Sub-Chapter 2.6** presents the conceptualisation of the change detection, the consistency required for effective ecosystem monitoring, the algorithms employed and the role of statistical testing with the tests used in the present work.

**Sub-Chapter 2.7** provides a general perspective on indices and their use for different applications within Earth Observation.

**Chapter 3** presents the scope and methodology followed to obtain, pre-process, and classify the satellite data, the statistics and metrics used for accuracy assessment and validation, as well as the change detection analysis and disturbance index proposal. A complete methodological flowchart for the Thesis is available in Appendix A.1.

**Sub-Chapter 3.1** presents the Páramo study assessment areas selected for this Thesis in terms of their location, their grouping for analysis, their specific characteristics in terms of climate, history, size, anthropogenic activities, land use pressures and population characterisation.

**Sub-Chapter 3.2** deals with the construction of the satellite image Data Pool and data selection by objective ranking, to define the study epochs around reference years. It presents the datasets accessed and their availability, the ranking principles followed and the resulting 11 epochs defined for the Thesis.

**Sub-Chapter 3.3** explains the pre-processing steps applied to the satellite image data and parameter selection of the transformations applied to the different datasets. It also explains the generation of the epoch Mosaics that comprise the data cube.

**Sub-Chapter 3.4.** explains in detail the selection of the number of classes and schema, the data clustering approach for the oldest datasets, the definition of the sample areas under a random stratified sampling approach, and the sample areas spectral signatures analysis. It also describes the applied classification methods, the hyperparameter tunning and iterative approach, with the respective accuracy assessment and statistics of the classification results.

**Sub-Chapter 3.5** covers the scope methods for change detection employed within the Páramo Complex Protected Areas, the reduced binary schema and the three timesteps defined. It explains the change typologies both for the bi-temporal cases and the change trajectories defined for the multitemporal case.

**Sub-Chapter 3.6** presents the conceptual and mathematical formulation of the agriculture disturbance index for the Páramo areas and explains how to interpret it.

**Chapter 4** summarises the results obtained for land cover and land cover change, their statistics, accuracies and error estimation.

**Sub-Chapter 4.1** presents the classification results for each Páramo Complex of the assessment areas, their statistical characterisation and its corresponding accuracy assessment. It defines the baseline results, presents the LULC products obtained, analyses the high confidence overall trends, and class-specific accuracy behaviours.

**Sub-Chapter 4.2** reveals the change detection and temporal trajectory results of the LULC classes for each Páramo Complex of the assessment areas for three different time steps of the study period. It explains the reduction to the binary Agriculture-Grass/Other schema and the conversion of the Agriculture-Grass to a 'MSI equivalent' area proportion, reviews the statistical significance of the changes, presents the bi-temporal change and temporal trajectory maps, and analyses the time step results.

**Sub-Chapter 4.3** shows the results for the Agriculture Disturbance Index proposed and presents the final Páramo areas ranking.

**Chapter 5** discusses the results considering similar work conducted, mirroring the Sub-Chapters of Chapter 4.

**Chapter 6** condenses the main conclusions, findings and implications of the work conducted. Highlights the limitations enunciated throughout the development and proposes considerations to guide further investigation.

# 1   Introduction

## 1.1.  Abbreviations

The following table compiles all the abbreviations used throughout the document.

Table 1.1: List of Abbreviations.

| Abbreviation | Definition |
| --- | --- |
| ACB | Altiplano Cundiboyacense |
|  | (*Colombia's Páramo Complex Protected Area*) |
| AGRI | Agriculture visualisation composite |
| ANN | Artificial Neural Network |
| AOI | Area of Interest |
| API | Application Programming Interface |
| BOA | Bottom of the Atmosphere |
| CAR | Corporación Autónoma Regional de Cundinamarca |
|  | (*Cundinamarca's Regional Environmental Authority*) |
| CART | Classification and Regression Tree |
| CHI | Chingaza |
|  | (*Colombia's Páramo Complex Protected Area*) |
| CI | Confidence Interval |
| CLC | CORINE Land Cover |
| CORINE | Collaboration for Information on the Environment |
| CVS | Cruz Verde y Sumapaz |
|  | (*Colombia's Páramo Complex Protected Area*) |
| CRS | Coordinate Reference System |
| DEM | Digital Elevation Model |
| DL | Deep Learning |
| DN | Digital Number |
| EMR | Electromagnetic Radiation |
| EO | Earth Observation |
| EPSG | European Petroleum Survey Group |
|  | (*Public Geodetic Parameter Dataset Registry*) |
| ESA | European Space Agency |
| ETM | Enhanced Thematic Mapper |

| | |
|---|---|
| FOV | Field of View |
| GML | Gaussian Maximum Likelihood |
| GCP | Ground Control Point |
| GEE | Google Earth Engine |
| GSD | Ground Sampling Distance |
| GUE | Guerrero |
| | (*Colombia's Páramo Complex Protected Area*) |
| HC | High Confidence |
| IDE | Interactive Development Environment |
| IFOV | Instantaneous Field of View |
| IGBP | International Geosphere-Biosphere Programme |
| IGM | Iguaque y Merchán |
| | (*Colombia's Páramo Complex Protected Area*) |
| IR / VIR | Infrared / Visible Infrared |
| L2-…-L9 | Short for Landsat Mission Number |
| LCCS | Land Cover Classification System |
| LEDAPS | Landsat Ecosystem Disturbance Adaptive Processing System |
| LaSRC | Land Surface Reflectance Code |
| MADS | Ministerio de Ambiente y Desarrollo Sostenible |
| | (*Colombia's Ministry for Environment and Sustainable Development*) |
| MAGNA | Marco Geocéntrico Nacional de Referencia |
| | (*Colombia's National Geocentric Reference Frame*) |
| MIR | Middle Infrared |
| MODIS | Moderate Resolution Imaging Spectroradiometer |
| MSI | Multispectral Imager |
| MSS | Multispectral Scanner System |
| NASA | National Aeronautics and Space Administration |
| NIR | Near Infrared |
| OLI | Operational Land Imager |
| PA | Protected Area |
| QA | Quality Attributes (of a pixel) |
| RE / VRE | Red Edge / Vegetation Red Edge |
| RF | Random Forest |
| RGB | Red Green Blue (*additive colour model*) |
| RMSE | Root Mean Square Error |
| RRB | Rabanal y Río Bogotá |
| | (*Colombia's Páramo Complex Protected Area*) |
| RTM | Radiative Transfer Model |
| S2 | Sentinel-2 |

| | |
|---|---|
| SE | Standard Error |
| SIRGAS | Sistema de Referencia Geocéntrico para las Américas (*Geocentric Reference Frame for the Americas*) |
| SLC | Scan Line Corrector |
| SR | Surface Reflectance |
| SVM | Support Vector Machine |
| SWIR | Shortwave Infrared |
| TIFF | Tag Image File Format |
| TIR | Thermal Infrared |
| TIRS | Thermal Infrared Sensor |
| TM | Thematic Mapper |
| TOA | Top of the Atmosphere |
| UAV | Unmanned Aerial Vehicle |
| URL | Uniform Resource Locator |
| UTM | Universal Transverse Mercator (*plane coordinate grid system*) |
| USGS | United States Geological Survey |
| WRS | Worldwide Reference System |
| YMC | Yellow Magenta Cyan (*subtractive colour model*) |

## 1.2. Objectives

This thesis aimed to employ Satellite Earth Observation (EO) free and open data to assess the level of disturbance of the Páramo ecosystem in the central Eastern Andes of Colombia. The ecosystem and its disturbances were approached strictly from a land cover perspective and within the government-declared boundaries for several of these protected areas, formally addressed as the Páramo Complex. The study developed the following 4 key objectives:

1. **Classify and establish a baseline land cover map for the Páramo Complex**. Using the earliest Satellite EO records, estimate the initial extent and location of the Páramo ecosystem land cover within the Páramo Complexes of Guerrero (GUE), Rabanal y Río Bogotá (RRB), and Altiplano Cundiboyacense (ACB) as a baseline for later comparison.

2. **Quantify the degree of change in Páramo land cover over time**. Derive a temporal trajectory of the ecosystem's land cover from the baseline to the most recent data, estimating the conversion and evolution of the different land cover classes within each Páramo Complex.

3. **Develop a Disturbance Index that encompasses spatial-temporal changes to appropriately determine Páramo quality**. The proposed index intends to summarise the current state and evolution of the ecosystem, to use as a metric for future monitoring.

4. **Identify the most pristine and most disturbed Páramo Complex and Páramo areas**. Temporal trajectory analysis and the disturbance index at different granularities is applied to the Páramo Complex and their corresponding administrative and hydrological sub-divisions, to produce an objective ranking of the best preserved to most disturbed Páramo areas.

All code developed for the Thesis is publicly available on the following repository:

https://github.com/Daenel35/Paramo-Land-Cover-Monitoring

## 1.3. Overview of Results

This Thesis focused on 3 Paramo Complexes (*Guerrero, Rabanal y Río Bogotá* and *Altiplano Cundiboyacense*) that were sub-divided into 6 Protected Areas (GUE, RRB, ACB_A, ACB_B, ACB_G, ACB_R). The disturbance due to agriculture was assessed over 47 years using Landsat and Sentinel-2 multispectral images, from a strictly land-cover perspective.

This work can conclude that from a land-cover perspective, Protected Areas in the central Eastern Andes have undergone an expansion of Agriculture-Grass for the past 47 years. This expansion behaviour is different among the Protected Areas included in the assessment, with some of them having a dominant agricultural vocation at the start of the study period. The change analysis showed that stable Agriculture-Grass areas also steadily increased, with a dynamic gain/loss behaviour, and a clear decelerating trend. However, further research is needed to determine whether this deceleration is correlated with economical or public-policy factors.

The assessment revealed that between 1977 and 2024, Agriculture-Grass areas have increased in most of the protected areas with some of them being historically dominated by this land cover: GUE: **34,1%** to **41,4%**; RRB: **9,9%** to **16,3%**; ACB_A: **78,7%** to **74,2%**; ACB_B: **69,9%** (year 1988) to **79,2%**; ACB_R: **55,4%** to **69,2%**; ACB_G: **27,0%** to **31,9%**. Change analysis detected dynamic agriculture gain-loss behaviour, with a clear deceleration of Agriculture-Grass gain.

A spatial-temporal Index to determine agriculture disturbance was proposed and calculated (GUE: **40,52** RRB: **12,85**; ACB_A: **74,82**; ACB_B: **75,02**; ACB_R: **71,84**; ACB_G: **39,96**). The Index should be interpreted as a representative—and optimistic lower bound— persistent Agriculture-Grass land cover percentage of the given area for the 47-year period.

The main contribution of this work, beyond the results reported, is the methodological approach focused on data and accuracy driven choices, detailed processing steps, open-access resources and replicability. Expansion, correction and re-purposing of the methodology is actively encouraged. The author hopes that this work will offer a key contribution and baseline for further multidisciplinary research on ecosystem monitoring and conservation that is so vital for the future of Páramo.

# 2 State of the Art & Methods

This chapter provides review of the relevant theoretical background required to understand the development of the Thesis and its results.

## 2.1.    The Páramo Ecosystem

The Páramo is a high-altitude ecosystem found in the tropical Americas, from northern Peru and along Ecuador, Colombia, Venezuela and small patches between Costa Rica and Panamá (Figure 2.1). It is located above the High Andean Humid Forest tree line and below the rocky glacial mountain tops, constituting an ecological archipelago along the northern Andes [1]. It has been recognised for its rich biodiversity and plant endemism [2], as well as its role in soil formation, high carbon storage [3], and hydrological regulation [4]. Consequently, it provides critical ecosystem services such as a steady supply of fresh surface water for human consumption and hydropower generation [5], as well as a unique landscape for biological research and nature tourism. Alongside this, it also provides an element of identity for its inhabitants, and is even considered  sacred  within some indigenous cultures [6].
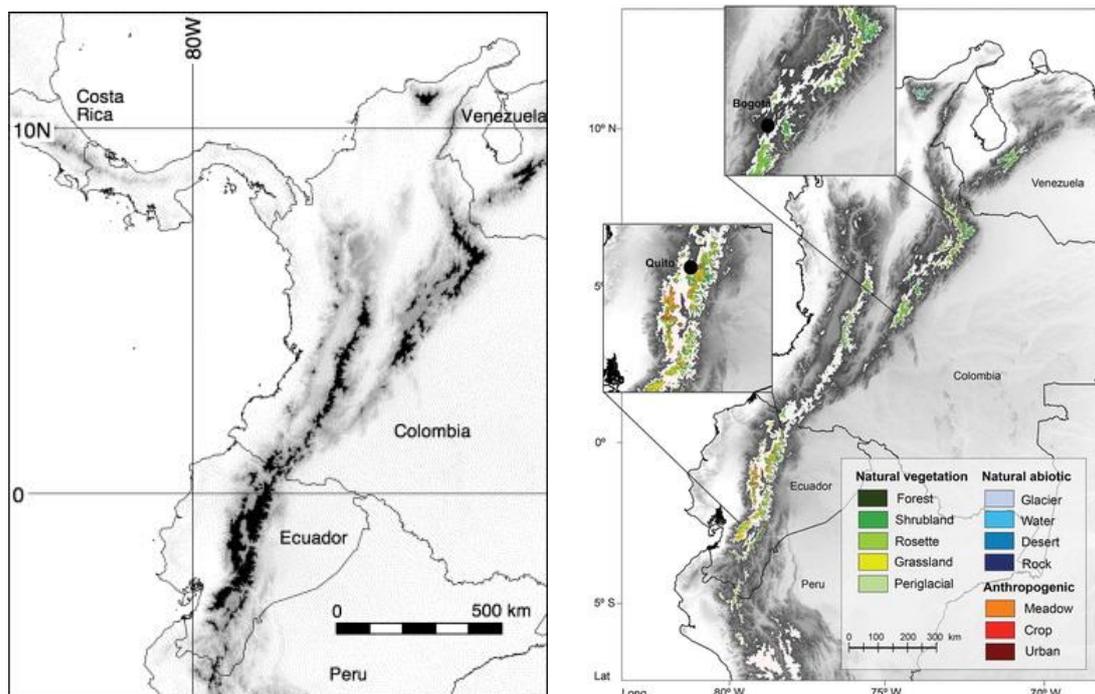


Figure 2.1: Left: Approximate extension (in black) of the Páramo in the Northern Andes [5]. Right: Landcover mapping of the Northern Andes [7]

Much work has been conducted on the biological and ecological characterisation of the Páramo and other Andean mountain ecosystems in Colombia [8], [9], [10], while more recent work has focused strictly on establishing the land cover and estimating the extent using satellite image classification [7]. This latter land cover perspective will be the focus of this Thesis.

It is not possible to define a unique altitudinal range for the Páramo due to the confluence of local factors such as climate, soil, vegetation and fauna and even the aspect of the mountain range (East / West) that determine the transition from the High Andean Humid Forest to the proper Páramo [11]. Nevertheless, Figure 2.2 provides a valuable insight into the classifications for the Central Eastern Andes Area of Interest (AOI). A lower bound of 2.500 m was chosen for this study (green box in Figure 2.2) to also incorporate the urban and agricultural Land Use/Land Cover (LULC) of the high plateau of the AOI, commonly referred as *Altiplano Cundiboyacense* due to its location between the Cundinamarca and Boyacá regions. The upper bound for the AOI is determined by the elevation of the highest peaks, that do not reach Nival heights.
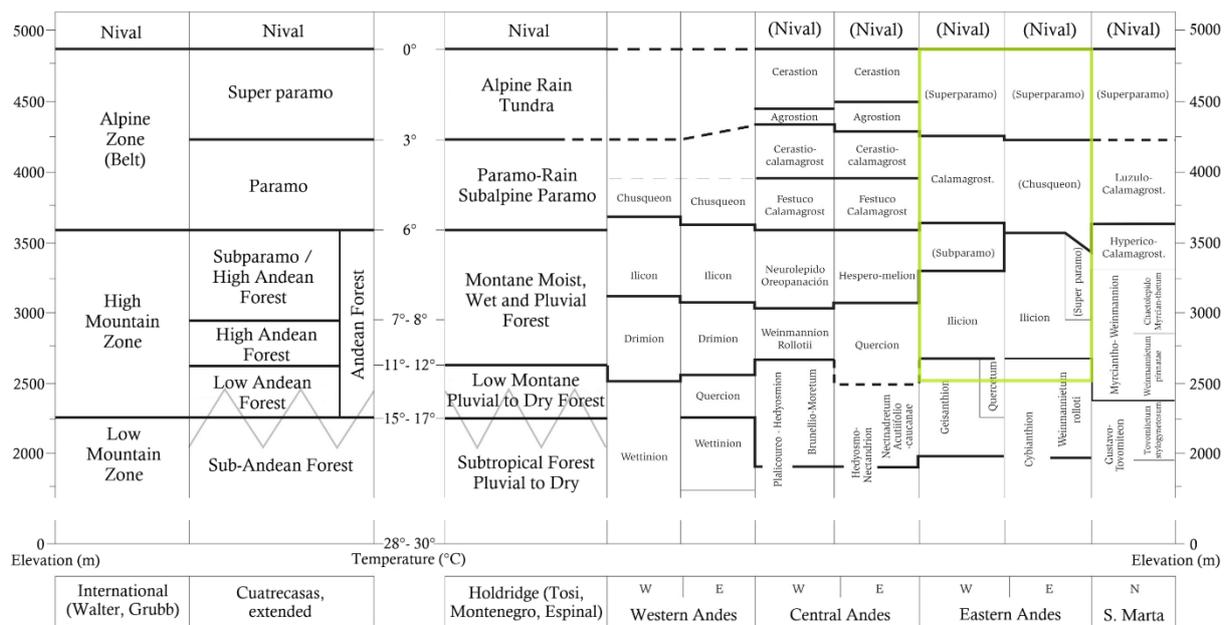


Figure 2.2: Altitudinal Zones and vegetation alliances/associations (Ecosystems) in the three Andes mountain ranges and the Santa Marta Massif Biomes from the Ecoandes Project, translated and adapted from [11, Tbl. 2]; The green box highlights the associations included within this study's area of interest.

Páramo landscape is characterised by short slow-growing riparian trees (*Polylepis*, *Chusquea*), evergreen bushes (*Hypericum*), pale-green and ochre mosaics of vegetation including giant rosette plants like Frailejones (*Espeletion*), tussock grasslands (*Calamagrostis*) along with a soil matrix of cushion plants, lichens and moss [1]. This vegetation and soil matrix actively intercepts air humidity and possesses a high water-holding capacity and low transpiration, favouring water retention and regulating release into permanent and temporary water bodies and streams [5], [12]. While

retaining water, this ecosystem slowly stabilises and accumulates organic matter in the soil, making it particularly fragile and vulnerable to anthropic action with decade long restoration horizons and enhanced climate change vulnerability due to average temperature increase [11], [13], [14], [15].

From a LULC perspective, the Páramo has all the above-mentioned elements of short vegetation, along with elements such as water bodies -lakes and wetlands-, exposed rock formations, riparian and transitional forest. Figure 2.3 gives the reader a glimpse of the Páramo landscape from ground level in three different Protected Areas. In addition to these natural elements, infrastructure related to human economic activities and sustenance like dams, water reservoirs, power line towers, communication towers, surface gravel mining, sub-surface coal mining, their related facilities, roads, paths, houses, schools, farms and agricultural fields, among others, can also be present in different proportions.



(a) Chingaza Complex (CHI), municipality of Guasca (30/04/2017)

(b) Guerrero Complex (GUE), municipality of Carmen de Carupa (27/12/2017)

(c) Rabanal y Río Bogotá Complex (RRB), municipality of Turmequé (06/07/2019)

(d) Chingaza Complex (CHI), municipality of Guasca (12/03/2020)

Figure 2.3: Author's photographs of sample ecosystem landscapes for different Páramos.

Colombia hosts nearly half of the world's total Páramo surface [16] due to its geographic location and the split of the Andes mountain range within its boundaries, which causes further diversification of its landscapes as evidenced in Figure 2.2. Since 2011 the Colombian Ministry of Environment and Sustainable Development (MADS) has officially delimited 36 different Páramo Complexes [17] –a terminology aimed to

highlight that some of them may be represented by more than one mapped polygon. The National System of Protected Areas, accounts for a total of 19.323 Km$^2$ of land delimited as Páramo [18]. The relatively recent inclusion of the Páramo to the System of Protected Areas highlights an increasing awareness of its importance and aims to counteract the accelerated degradation due to, among others, human intervention from mining, cropping and livestock grazing [19]. The location of the delimited Páramo Complexes in Colombia and the ones in the study's Central Eastern Andes AOI are shown in Figure 2.4.
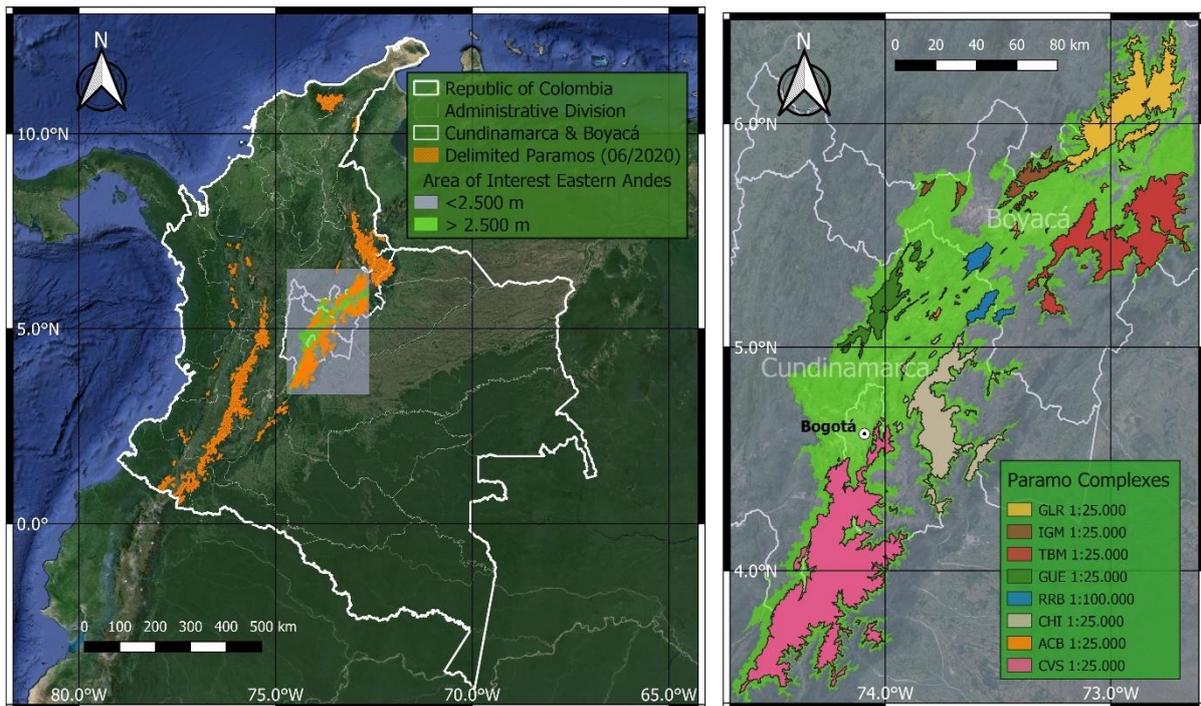


Figure 2.4: Government Delimited Páramo Complexes in Colombia and in the Central Eastern Andes AOI.

Colombian regulations state full prohibition of extractive and agricultural activities on Páramo areas that were undisturbed prior to June 2011 within the delimitation [20]. Regulations provided further guidelines for the substitution and conversion of ongoing agricultural activities for areas that were disturbed prior to June 2011, according to the land-use defined by regional environmental authorities [21], [22]. These regulations are issued per Páramo Complex. This date-based approach results in a recognition of the fact that some of these delimited areas had never been granted any protection status before June 2011 and economic activities —whether for profit or for sustenance— were traditionally and lawfully conducted there. By applying the substitution and conversion policy, consideration could be given to the Páramo inhabitants and their roots. However, the transitional nature of this substitution and conversion of agricultural activities requires a monitoring system at a scale that Remote Sensing can address.

Remotely sensed monitoring efforts for the High Mountain ecosystems—including Páramo— have been developed in Colombia at different temporal scales and granularities. A large scale (1:100.000[1]) cartographic analysis based on USGS Landsat Satellite imagery estimated that between 1970 and 1990 the country's share of High Mountain ecosystems areas –above 2.744 m in elevation– converted to agriculture reached 24,9% (242.373 Ha) and the central region of Cundinamarca underwent a 211% increase in its agricultural areas for this period [12]. A multitemporal analysis spanning 2005 to 2017 on 5 Páramo Complex Areas belonging to the AOI (ACB, CHI, GUE, RRB & IGM), found that 4 out of 5 exhibit high fragmentation and a negative balance in the gain-loss of their natural coverage [24]. Agricultural LULC in GUE Páramo Complex has been estimated to be 31% for the 2014-2016 period within the Ubaté Region [25], and 37% for January 2020 within the municipality of Tausa [26]. In this same municipality, the agriculture trend was estimated at a 6% annual increase between 2007 to 2016 [27]. A very recent multitemporal study estimated that within 1984 and 2021 the Guerrero and Rabanal Páramos have respectively lost 47.96% and 59.96% of their native vegetation, replaced primarily by crops, pastures, and planted forests [28].

These research examples set the stage and portray how Páramo ecosystem, as well as many other natural ecosystems, have suffered from the strain of anthropogenic pressure. They also demonstrate how by monitoring the LULC of these areas, their state and evolution in time can be measured and quantified. This is of course, a single perspective on a multidisciplinary issue, as ecosystems might be maintaining their extent while still suffering from disturbances impacting their species composition and/or diversity, or being exposed to chemical, water, or air pollution that impact their overall health (Please refer to Sub-Chapter 3.1 for a description of the disturbances of the assessment areas). Measuring the extent of Paramo landcover will be the metric used to assess Páramo ecosystem wellbeing under the scope of this Thesis. A holistic multidisciplinary approach to Páramo ecosystem monitoring is necessary, and this thesis provides one strand of that multidisciplinary approach with a focus on land cover.

---

[1] Note of contextualisation on scale concept: A scale represents the ratio between the physical map size and the real size of the elements depicted on it, 1:100.000 represents that 1 m on the map is equivalent to 100 Km in reality, or 1 cm equivalent to 1 Km. The use of an embedded scale bar that shows this relationship on maps is usually preferred. A more informative approach to scale is considering the smallest element to be represented in a paper or screen map —as a dot— is about 0.2 mm, therefore the smallest element on the 1:100.000 scale map is the product of the nominal scale by this fixed length: $100.000 \times 0,2mm \times 10^{-3}m/mm =$ 20 m, this length is referred to as the ground error. As the scale is usually constant on the horizontal and vertical axis, and to represent an object you need at least 2 dots on each direction it is evident anything smaller than 40mx40m will not be represented on the map at this scale. In fact, for something to have any area some space is needed between the lines formed by these dots or for it to be about 0.5mm in size, so following the same logic as before, for a 1:100.000 map the detectable size of an object is 50m, and for a 1:25.000 map is 5m [23].

## 2.2. Remote Sensing and Satellite EO

The work undertaken in this Thesis will focus on free and open data derived from passive multispectral Satellite EO missions. The present Sub-Chapter will deal with the background needed to understand this statement, and the relevant factors influencing data acquisition and mission characteristics.

Remote Sensing refers to making observations of an object using a device that is physically distant from it. Collecting data about the Earth and its atmosphere recorded by airborne and satellite-borne sensors through remote sensing is referred to as Earth Observation (EO). As the sensor is not in direct contact with the measured object, the energy detected from the object is assumed to be indicative of its properties and the relationship between the measured energy and the object properties must be established [29]. For Satellite sensors, energy in the form of magnetism, gravity, radioactivity and electromagnetic radiation can be collected. The Electromagnetic Radiation (EMR) is one of the most relevant and commonly used forms of energy for Satellite EO. It is also the one that is most familiar to us due to the sensor's parallelism with human vision, as we rely on an object's reflection of visible light —a small portion of the spectrum—to detect it. A logarithmic scale with the common names assigned to different portions of the spectrum, highlighting the visible part and our colour interpretation of it, is displayed in Figure 2.5.



Figure 2.5: Electromagnetic spectrum [30, Fig. 2].

It is important to note that all materials both reflect EMR or emit EMR as a function of their temperature, and that the incoming EMR energy to an object is either reflected, absorbed or transmitted, following the energy conservation principle. Consequently, both the Earth's surface and its atmosphere reflect, absorb and transmit the incoming EMR irradiated by the Sun—the biggest energy source among others—, while also emitting EMR themselves (function of their temperature and energy absorbed and re-

emitted usually at higher wavelengths). The amount of the Sun's energy reflected depends on two material properties: first, the roughness of the objects surface, as smooth surfaces will reflect incoming energy in a single direction like a mirror, while rough ones will scatter the light in many directions, reflecting less in the direction of the sensor. Second, the materials albedo, which is the proportion of incident energy reflected by a material that provides a metric (%) on how well an object reflects light instead of absorbing it. Snow and clouds, for example, have a very high albedo, reflecting up to 90% of incoming solar radiation. The ocean, water bodies, and dark wet soil on the other hand, reflect less than 15% of incoming solar radiation and absorb the rest [31]. Figure 2.6 shows an informative albedo chart and table of different surfaces, including different kinds of atmospheric clouds (*Cirrus*, *Stratus*, *Altostratus*, *Cumulus*).



| Surface Material | Sample Albedo Values |
|---|---|
| Perfect absorber (blackbody) | 0 |
| Ocean | 0.03–0.10 |
| Forest | 0.05–0.20 |
| Bare Soil | 0.05–0.40 |
| Grass and Crops | 0.1–0.25 |
| Granite | 0.30–0.35 |
| Sand | 0.15–0.45 |
| Ice | 0.20–0.50 |
| Clouds | 0.30–0.90 |
| Snow | 0.40–0.90 |
| Perfect reflector (absolutely white surface) | 1 |

Figure 2.6: Albedo Values[29, Fig. 5.15] and Sample Albedo Values
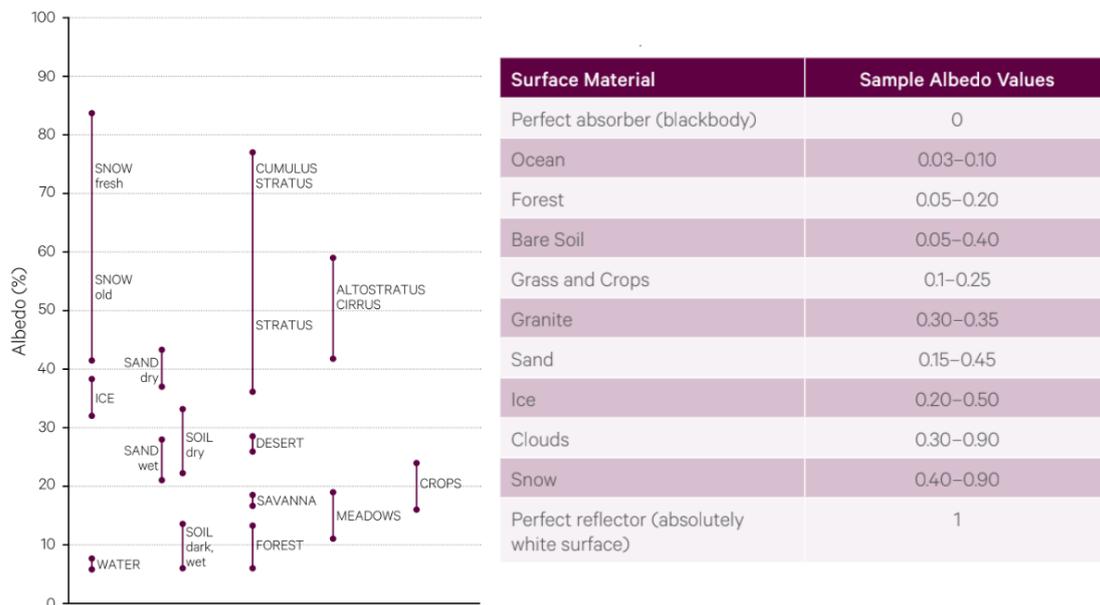
Passive remote sensing sensors measure the EMR energy reflected and emitted from the Earth's surface, while active remote sensing sensors beam EMR towards the surface –like radio waves— and capture the reflected response (Figure 2.7).
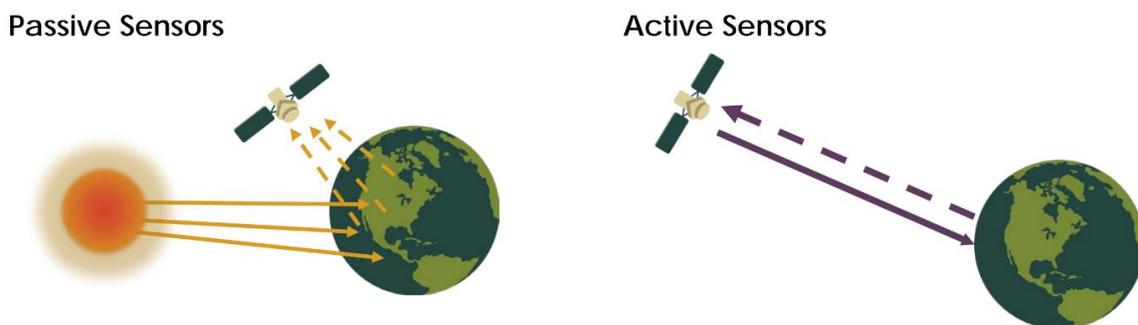


Figure 2.7: Diagram of a passive sensor versus an active sensor. Credit: NASA Applied Sciences Remote Sensing Training Program

Earth's surface reflected solar energy is detected by passive remote sensing devices in the Visible, Near Infrared (NIR), Short Wave Infrared (SWIR) and Middle Infrared (MIR) regions, while the Earth's own emitted energy may be detected in the Middle Infrared (MIR), Thermal Infrared (TIR) and microwave wavelengths [29]. Figure 2.8 presents the spectral regions of interest, including the subdivision of the Infrared spectrum, as well as the name given to the sensors used to detect them.

| Spectral region | Wavelength | EO sensor | |
|---|---|---|---|
| | | Passive | Active |
| Visible | 0.38 to 0.7 μm | Passive radiometers Passive spectroradiometers Passive imaging spectrometers Polarimeters | Lidar |
| Near Infrared (NIR) | 0.7 to 1.1 μm | | |
| Short wave Infrared (SWIR) | 1.1 to 3.0 μm | | |
| Middle Infrared (MIR) | 3.0 to 8.0 μm | Passive radiometers Polarimeters | |
| Thermal Infrared (TIR) | 8.0 to 15.0 μm | | |
| Microwave | 1 mm to 1 m | Microwave radiometers Microwave spectrometers | Radar Polarimeters Scatterometers |
| Radio | >10 cm | Passive radiometers Sounders | Imaging radar Altimeters Sounders |

Figure 2.8: Spectral regions observed by EO sensors[29, Tbl. 13.1]

Satellite EMR sensors collect energy readings from different wavelength ranges or portions of the spectrum, commonly referred to as bands. Our eyes and optical devices like cameras work like passive sensors, collecting energy readings in three visible bands, Blue, Green and Red, corresponding to the peak-sensitivity wavelengths through which humans perceive colour.

Different land covers have varying, and often characteristic, reflectance in the visible and non-visible wavelengths (SWIR, MIR, TIR). Consequently, engineered sensors capable of detecting and quantifying both visible and non-visible waves, enable us to identify an object with more information than what we perceive as colour. Surface features or land covers can have characteristic spectral reflectance curves showing a 'typical' shape over a particular range of wavelengths [29]. This typical behaviour is often referred to as a spectral signature, for which several examples are shown in Figure 2.9. It is important to note that spectral signatures are the idealised line that 'connects' the dots of each sensed band reading (as not all wavelengths are or can be sampled). Energy is commonly reported in dimensionless units of relative energy like reflectance, obtained by dividing the reflected energy by the incident energy.
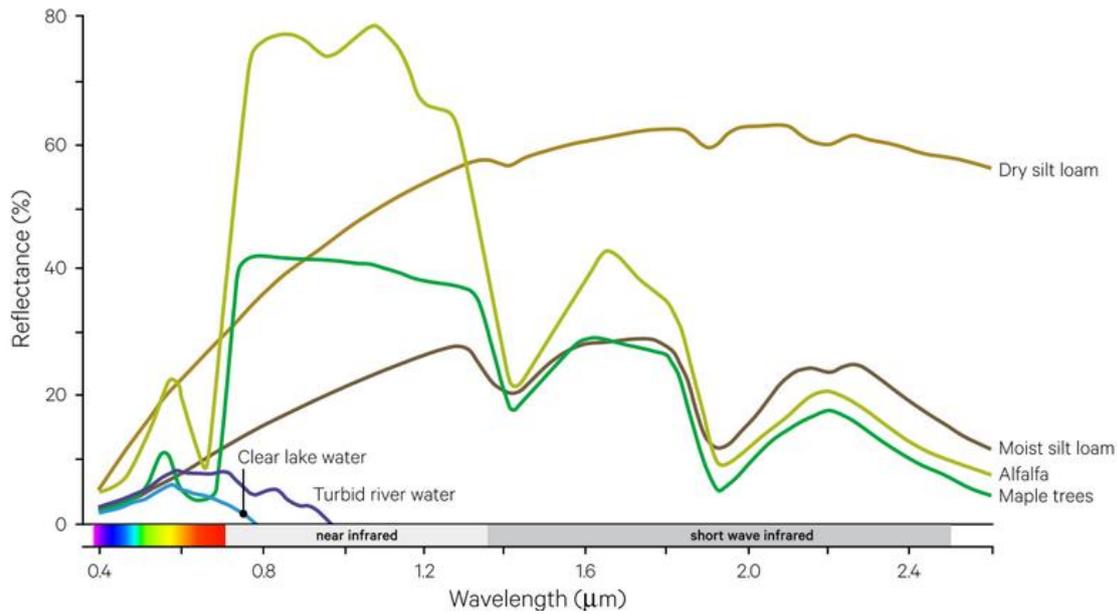
Figure 2.9: Idealised Spectral signatures[30, Fig. 6]

Much like a camera, a satellite sensor has a limited Field of View (FOV), representing the angular extent of observations acquired, and an Instantaneous Field of View (IFOV) represented by the angle delimited by a single detector element of an optical system [32]. The projected FOV or ground FOV is also known as the Swath width, while its instantaneous counterpart ground IFOV is known as the Ground Sample Distance (GSD) or picture element—referred to as a pixel from now on— size. Figure 2.10 provides a schematic of these concepts for two sensors with different acquisition modes: cross track or whiskbroom, that collects observations while scanning side to side along the orthogonal direction of the flight, and along track or push broom, collecting observations simultaneously for the entire FOV.
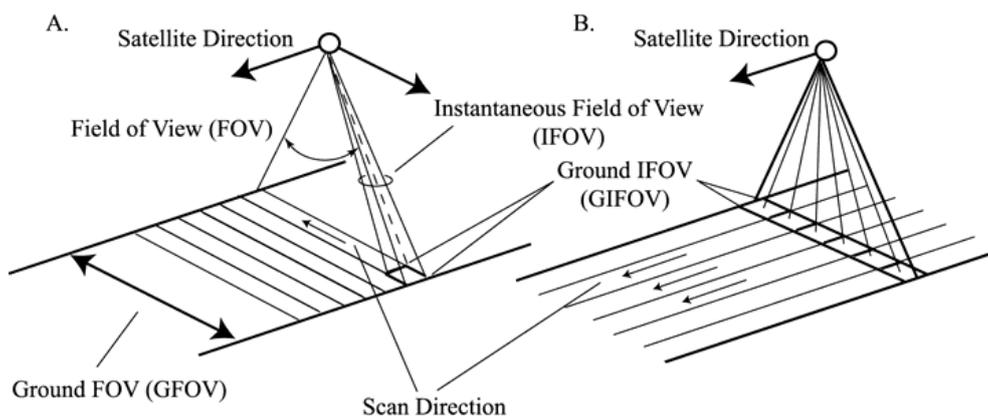


Figure 2.10: Cross track (A) and along track (B) scanners [32, Fig. 5]

The data collected by satellite EO is characterised by four types of resolution: radiometric, spatial, spectral, and temporal[31] as graphically presented in Figure 2.11. This resolution is important to understand what sets the different existent satellite missions and their sensors apart.
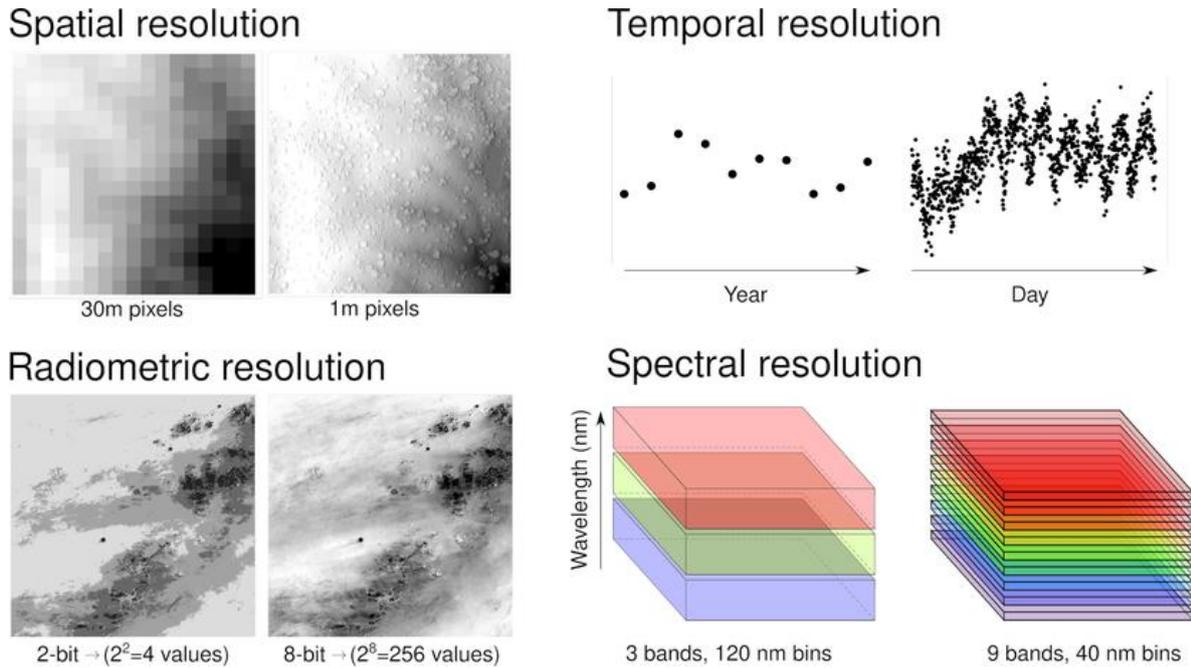
Figure 2.11: Examples of different kinds of resolution [33, Fig. 1]

**Radiometric resolution** refers to the levels of quantisation of the energy of a single detector element or pixel, that is, the number of bits representing the energy recorded. Energy values are encoded in bits, or 0/1 values. For two bits, 4 values are possible (00, 01, 10 and 11), and the number of possible values is 2 to the power of the number of bits. For an *8-bit* resolution, the sensor has $2^8 = 256$ potential digital values (0 - 255) to store information, these values are often referred to as Digital Number (DN) and represent the raw data recorded from the sensor. A set of DNs is recorded for each band the device can sense. Thus, the higher the resolution, the more values are available to store information, and better discrimination between the slightest differences in energy is provided [31]. The calibration data for the sensor is what allows conversion of the DN readings to actual energy magnitudes —like radiance [W/m$^2$-sr] in a process called *radiometric correction*.

**Spatial resolution** depicts the size of the individual reading or pixel at ground level (also referred to as ground IFOV or GSD) [32]. This pixel size can vary from kilometres to centimetres depending on the sensor, as resolution varies between global, continental, national, regional, and local applications. Very fine resolution Satellite EO data (30cm – 9m) belongs to the realm of commercial data, while coarser resolution (≥10m) is usually free or open-access for research and non-commercial purposes. The span of the ground FOV or Swath width is proportional to the number of pixels and will vary in magnitude between metres and hundreds of kilometres, and due to data size constraints, the observed strip will be broken down in scenes as depicted in Figure 2.12.
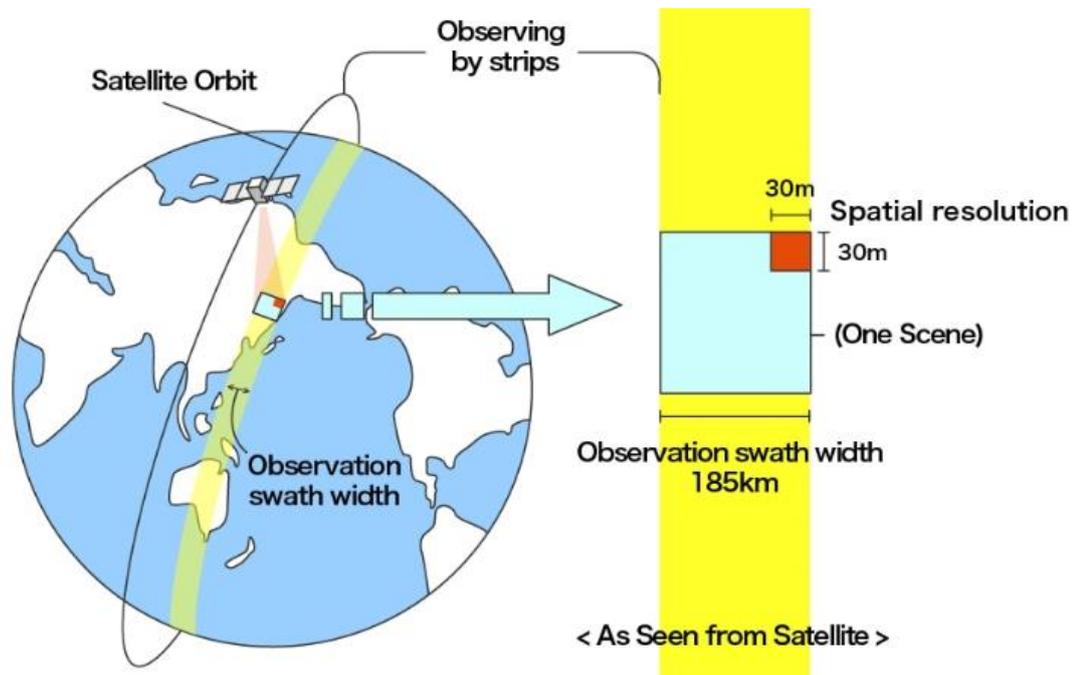
Figure 2.12: Strips and Spatial resolution[34]

**Spectral resolution** refers to the ability of a sensor to discern finer EMR wavelengths, that is, having more and narrower bands. The narrower the range of wavelengths for a given band, the finer the spectral resolution [31]. Sensors are broadly grouped in *multispectral*, having 3 to 10 bands, while others with tens to hundreds are named *hyperspectral*, to even thousands of bands that are considered *ultra spectral*. Although names for different portions of the EMR spectrum are the same, not every sensor has the same range for the band that sense it. Bands in the NIR, SWIR and even Red, Green, and Blue bands are not always referring to the same wavelength ranges.

**Temporal resolution** is related both to the sensor and to the satellite's orbit, as it is defined as the time taken for the next observation of the same area. In other words, it is the time it takes for the sensor to revisit the same observation area [30]. The better the spatial resolution, the narrower the swath, and the more orbits needed to complete the full sweep around the earth to arrive to the same area. This means that there is a trade-off between spatial and temporal resolution, as both cannot be maximised together. EO Satellites typically follow a near polar and near circular orbit, meaning they pass over a given point on the Earth's surface at the same local time on each overpass, which minimises surface illumination and shadowing differences [29].

As mentioned at the beginning of the Sub-Chapter, this work will focus on free and open data derived from passive multispectral Satellite EO missions. Before reviewing the mission's characteristics, it is important to further describe some factors directly related to the corrections and data manipulation prior to the data usage, in what is known as *Pre-processing*.

Among all the many physics principles involved in remote sensing, it is important to note that observations of the Earth surface from satellites are subject to several factors related to **1.** The pixel's geometric correspondence and georeferencing, **2.** The atmosphere composition and state, and **3.** The constraints inherent to the sensor resolution. All three of them affect the data collected and the comparison between different datasets, so an overview of the factors is presented below.

**Factor 1:** When remote sensing the Earth's surface, geometric constraints need to be considered and corrected to assign the readings to a specific location, that is, to correctly georeference each pixel.

Many idealisations on the geometries of the earth, the satellite orbit and the sensor are often far from the reality: The Earth is neither a sphere or an ellipse, nor is its surface evenly curved or flat —or even static—, its gravitational pull is not homogeneous in all directions, making satellite orbits not totally Keplerian, and signals are not homogeneously and perpendicularly collected by the sensor surface. A good model for the geometry of the data acquisition incorporates the error induced by all these factors and allows for what is denominated as *geometric correction*. This correction allows the collected data to be rectified and registered, that is, to accurately represent the geometric features of the Earth's surface over a regular grid and assign a coordinate to the pixels of a given acquisition. Geometric distortions in satellite borne sensors are well-defined and routinely corrected by data suppliers before distribution of data products [35]. Figure 2.13 illustrates how pixel size increases away from the centre of an image —the nadir—, a distortion further exaggerated by the Earth's curvature depending on the sensor's swath.
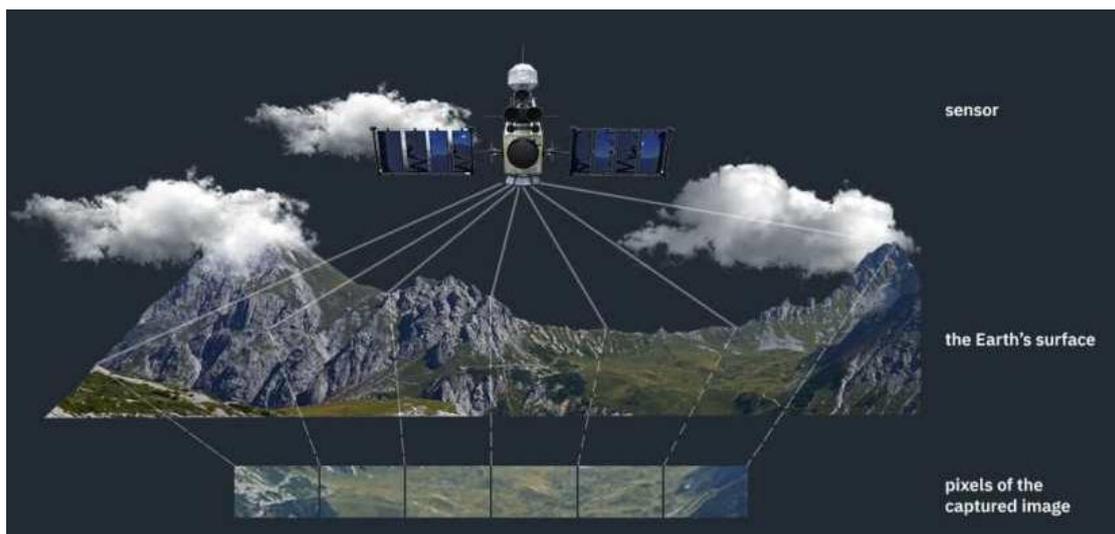


Figure 2.13: Example of sensor geometric distortion [36]

Geometric corrections require the precise location and attitude of the satellite obtained through satellite ephemerides and geoid models, sensor distortion correction and corrections due to the earth shape and topography are performed through photogrammetric analysis of Ground Control Points (GCPs) and Digital Terrain Models (DTMs). Once geometrically corrected data is made available, there is still the possibility for further transformations to be made due to changes in the Coordinate Reference System (CRS) or the projection —to match an existing map like in Figure 2.14—, image co-registering —to match different spatial datasets—, or sub-pixel level adjustments. These transformations are performed by *resampling*.



Figure 2.14: EO scene geometry and resampling [35].

The resampling of EO data refers to the problem of assigning a new pixel value to a pixel that is a different size and shape from the original optical pixel of the EO image. This is achieved by different strategies that use the values of the neighbouring pixels, either by selecting the nearest pixel value —known as nearest neighbour— or by interpolating these values. The latter approach implicitly assumes that there is a continuous function that has been sampled to derive the discrete values of image pixels and derive new values from the original image pixels [37].

There are trade-offs when considering using the resampling methods, as sometimes it is preferred to preserve the original pixel values with the nearest neighbour, so as not to attenuate or smoothen the data with the interpolation. Sometimes, however, interpolation can reduce the local image variance, and the smoothening becomes desirable to reduce the noise, or it is necessary due to a significant change in pixel size or orientation. An illustration of the three most used resampling methods is presented in Figure 2.15.

Figure 2.15: Resampling methods and implementation. a) Nearest Neighbour, b) Bilinear Interpolation, and c) Cubic Convolution, also named Bi-cubic interpolation. Adapted from [37, Fig. 5.4,5.14,5.15]

**Factors 2:** When remote sensing the Earth's surface, atmosphere is a media that stands in between the sensor and the objective surface, and its composition and state affects the energy readings as it too absorbs, reflects and emits EMR.

On one hand, atmospheric gas composition is roughly constant and various gases that make up the atmosphere (particularly $O_3$, $CO_2$, and $H_2O$), completely absorb certain portions of the EMR spectrum [29]. The opacity or absorptance of these regions (ratio of absorbed over incident energy) is 100% or close to 100%, and it is evident that wavelengths in these regions cannot be used for remote sensing of the Earth's surface features. Spectral regions with low opacity or absorptance have high transmittance or transmission (ratio of transmitted over incident energy) and are called *atmospheric windows*, as illustrated by the opacity dips in Figure 2.16.

Figure 2.16: Earth's atmospheric EMR opacity[29, Fig. 5.14]

On the other hand, within these atmospheric windows the atmosphere state or atmospheric conditions —such as haze, fog, clouds and aerosols[2]— can still affect the EMR readings through different types of scattering [39].

These atmospheric interactions need to be considered after the sensor raw DN readings are translated to energy units, also referred to as at-sensor radiance or Top of the Atmosphere (TOA) radiance through the sensor calibration data. The incorporation of these interactions is known as *atmospheric correction* and allows the at-sensor radiance or TOA radiance to be converted to Bottom of the Atmosphere (BOA) reflectance or Surface Reflectance (SR)[40].

Through the conceptualisation of an atmospheric RTM or Radiative Transfer Model (see Figure 2.17 as an example) it can be stated that the radiance picked up by the sensor has two main components: the energy reflected by the target surface and transmitted by the atmosphere to the sensor ($L_T$) and the diffuse energy from the atmosphere whose additive path reaches the sensor ($L_p$). The intercorrelations between the atmospheric interactions can be described at different levels of complexity, as the atmosphere itself can be further considered as different layers with different properties, or other factors related to atmospheric conditions can be considered, which generates an array of RTMs [41]. In essence, this means that regardless of the model, meteorological data from the moment of the observation acquisition is needed to perform atmospheric correction.

---

[2]Note on Aerosols: Aerosols or airborne particles is a broad term for suspended solid, liquid or mixed particles in the atmosphere, such as dust, smoke, ash, water and salt. Their small size, mostly in the range 0.1 to 1 μm, results in scattering interactions with the EMR but also in interactions with atmospheric water gas, acting as condensation nuclei and thus, a powerful climate driver [38].

26



Figure 2.17: Simple radiative transfer model [39]

**Factors 3:** When comparing data from different mission sensors, special attention should be put on their differences in resolution and corrections.

As noted by the previous two factors and the different resolution types described before, it is evident that direct comparison between EO data from different missions is to be handled with care, as they can be very heterogeneous even among passive multispectral sensors. Even when co-registering is performed among datasets and their pixels are aligned, direct spectral comparison when spectral bands are not matching, or atmospheric correction was performed under different atmospheric RTMs, can be an issue. These limitations were considered when developing the methodology for this Thesis and are reflected in the data analysis decision-making. The efforts to harness a vast amount of remotely sensed data of different kinds is referred to as data fusion and is a field of active research [42].

Data made available by satellite EO missions are often classified in levels —whose name labels can vary— that reflect whether it is raw sensor data, or if it has been subject to a certain level of processing based on the geometric and atmospheric corrections described. Conventionally, data is classified according to a hierarchy of increasingly refined levels of processing from raw instrument data (Level 0), data calibrated and geolocated (Level 1), data converted into geophysical parameters (Level 2), and gridded, and sometimes temporally composited, into an Earth-based coordinate system (Level 3)[43], data products are model output or results from analyses of lower level data, e.g. variables derived from multiple measurements (Level 4) [44]. The data levels used in the development of this Thesis and a general schema of the processing levels is shown in Figure 2.18.

Figure 2.18: General schema for data products of EO observations, highlighting the levels of interest for the Thesis.

The United States Geological Survey (USGS) Landsat mission products, including the Thematic Mapper (TM), Enhanced Thematic Mapper (ETM+), and Operational Land Imager (OLI) sensors, have ideal sampling characteristics for monitoring diverse land cover types at a regional and historic level [45] and given its long-standing record of EO acquisitions, was the obvious choice for the Páramo ecosystem land cover analysis for this Thesis. Older Landsat mission products from the Multispectral Scanner System —Commonly referred to as Multi Spectral Scanner—(MSS) were also considered, particularly because they represent the earliest records and relate to the baseline state of the LULC in the AOI. Figure 2.19 presents the Landsat's different instrument bands and their wavelength ranges, atop the atmospheric windows of the EMR spectrum.



Figure 2.19: Spectral Band passes for all Landsat Sensors [46].

Research, scientific knowledge and policymaking has been fuelled by over 5 decades of Remote Sensing: from the launch of Landsat 1 satellite in 1972 and the beginning of the era of land Remote Sensing, following the 2008 release of the freely accessible USGS Landsat data archive, to the launch of Landsat 9 in 2021. The current understanding of the status and dynamics of global ecosystems is widely informed by Landsat data, with scientific, management and relevant policy insights [47]. In fact, Landsat has been pivotal in observing and documenting long-term ecological changes, with implications spanning regional, national, and international levels. The data archives and the developed analytical methods in EO have shaped and continue to inform polices aimed at improving the management of natural resources at a global scale and play a key role for the Sustainable Developing Goals Targets [48], particularly on the monitoring of terrestrial ecosystems, as is the case for this work.

Following the success of the Landsat missions, The European Space Agency (ESA) launched the Copernicus programme and developed a family of Sentinel Satellites to carry on the legacy of accurate, timely and easily accessible information to improve the management of the environment, understand and mitigate the effects of climate change and ensure civil security [49]. The Sentinel-2A and Sentinel-2B twin satellites —Often referred to as simply Sentinel-2—, launched in 2015 and 2017 respectively, are designed to deliver high-resolution optical images for land services[50]. The mission's spectral bands are presented in Figure 2.20.



Figure 2.20: Spectral Band passes for the Copernicus Sentinel-2

Due to their suitable characteristics and free availability, Landsat and Sentinel-2 data were selected to perform the temporal trajectory analysis of the Páramo in the AOI. A compendium of all the missions and bands used in the present work, and their characteristics can be found in

Table 2.1.

Table 2.1: Landsat satellites & corresponding bands specifications used in this work.

| Satellite Mission | Temporal Resolution (days) | Spectral Resolution (bits) | Sensor | Spectral Bands | Visible | | | Infrared | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Blue | Green | Red | RE1 | RE2 | RE3 | NIR | | Narrow NIR (RE4) | SWIR1 | SWIR2 |
| Landsat 2 | 18 | 6 | MSS | Band | - | B4 | B5 | - | - | - | B6 | B7 | - | - | - |
| | | | | ⁀λ [μm] | - | 0,5-0,6 | 0,6-0,7 | - | - | - | 0,7-0,8 | 0,8-1,1 | - | - | - |
| | | | | λ [μm] | - | ~0,55 | ~0,65 | - | - | - | ~0,75 | ~0,90 | - | - | - |
| | | | | Res. [m] | - | 60* | 60* | - | - | - | 60* | 60* | - | - | - |
| Landsat 4 | 16 | 6 | MSS | Band | - | B1 | B2 | - | - | - | B3 | B4 | - | - | - |
| | | | | ⁀λ [μm] | - | 0,5-0,6 | 0,6-0,7 | - | - | - | 0,7-0,8 | 0,8-1,1 | - | - | - |
| | | | | λ [μm] | - | ~0,55 | ~0,65 | - | - | - | ~0,75 | ~0,90 | - | - | - |
| | | | | Res. [m] | - | 60* | 60* | - | - | - | 60* | 60* | - | - | - |
| | | 8 | TM | Band | B1 | B2 | B3 | - | - | - | B4 | | - | B5 | B7 |
| | | | | ⁀λ [μm] | 0,45-0,52 | 0,52-0,60 | 0,63-0,69 | - | - | - | 0,77-0,90 | | - | 1,55-1,75 | 2,08-2,35 |
| | | | | λ [μm] | ~0,485 | ~0,56 | ~0,66 | - | - | - | ~0,83 | | - | ~1,65 | ~2,22 |
| | | | | Res. [m] | 30 | 30 | 30 | - | - | - | 30 | | - | 30 | 30 |
| Landsat 5 | 16 | 8 | TM | Band | B1 | B2 | B3 | - | - | - | B4 | | - | B5 | B7 |
| | | | | ⁀λ [μm] | 0,45-0,52 | 0,52-0,60 | 0,63-0,69 | - | - | - | 0,77-0,90 | | - | 1,55-1,75 | 2,08-2,35 |
| | | | | λ [μm] | ~0,485 | ~0,56 | ~0,66 | - | - | - | ~0,83 | | - | ~1,65 | ~2,22 |
| | | | | Res. [m] | 30 | 30 | 30 | - | - | - | 30 | | - | 30 | 30 |
| Landsat 7 | 16 | 8 | ETM+ | Band | B1 | B2 | B3 | - | - | - | B4 | | - | B5 | B7 |
| | | | | ⁀λ [μm] | 0,441-0,514 | 0,519-0,601 | 0,631-0,692 | - | - | - | 0,772-0,898 | | - | 1,547-1,749 | 2,064-2,345 |
| | | | | λ [μm] | 0,485 | 0,56 | 0,66 | - | - | - | 0,835 | | - | 1,65 | 2,215 |
| | | | | Res. [m] | 30 | 30 | 30 | - | - | - | 30 | | - | 30 | 30 |
| Landsat 8 | 16 | 12 | OLI | Band | B2 | B3 | B4 | - | - | - | B5 | | - | B6 | B7 |
| | | | | ⁀λ [μm] | 0,452-0,512 | 0,533-0,590 | 0,636-0,673 | - | - | - | 0,851-0,879 | | - | 1,566-1,651 | 2,107-2,294 |
| | | | | λ [μm] | 0,482 | 0,561 | 0,655 | - | - | - | 0,864 | | - | 1,614 | 2,202 |
| | | | | Res. [m] | 30 | 30 | 30 | - | - | - | 30 | | - | 30 | 30 |

For abbreviations, refer to Table 1.1. Sentinel-2 values in brackets refer to the joint mission values.

*Original GSD sample is 68m x 83m, commonly resampled to 57 m or 60 m

Table 2.2: Sentinel-2 satellites & corresponding bands specifications used in this work.

| Satellite Mission | Temporal Resolution (days) | Spectral Resolution (bits) | Sensor | Spectral Bands | Visible | | | Infrared | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Blue | Green | Red | RE1 | RE2 | RE3 | NIR | Narrow NIR (RE4) | SWIR1 | SWIR2 |
| Sentinel 2-A | | | MSI | Band | B2 | B3 | B4 | B5 | B6 | B7 | B8 | B8A | B11 | B12 |
| | | | | $\widehat{\lambda}$ [μm] | 0,451 -0,539 | 0,538- 0,585 | 0,641- 0,689 | 0,695- 0,715 | 0,731- 0,749 | 0,769- 0,797 | 0,784- 0,900 | 0,855- 0,875 | 1,565- 1,655 | 2,100- 2,280 |
| | | | | λ [μm] | (0,49) 0,4966 | (0,56) 0,560 | (0,665) 0,6645 | (0,705) 0,7039 | (0,74) 0,7402 | (0,783) 0,7825 | (0,842) 0,8351 | (0,865) 0,8648 | (1,61) 1,6137 | (2,19) 2,2024 |
| | 10 (5) | 12 | | Res. [m] | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 |
| Sentinel 2-B | | | MSI | Band | B2 | B3 | B4 | B5 | B6 | B7 | B8 | B8A | B11 | B12 |
| | | | | $\widehat{\lambda}$ [μm] | 0,451 -0,539 | 0,538- 0,585 | 0,641- 0,689 | 0,695- 0,715 | 0,731- 0,749 | 0,769- 0,797 | 0,784- 0,900 | 0,855- 0,875 | 1,565- 1,655 | 2,100- 2,280 |
| | | | | λ [μm] | (0,49) 0,4921 | (0,56) 0,559 | (0,665) 0,665 | (0,705) 0,7038 | (0,74) 0,7391 | (0,783) 0,7797 | (0,842) 0,833 | (0,865) 0,864 | (1,61) 1,6104 | (2,19) 2,1857 |
| | | | | Res. [m] | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 |

For abbreviations, refer to Table 1.1. Sentinel-2 values in brackets refer to the joint mission values.

## 2.3. Data Access, Selection and Pre-processing

EO Satellite missions rely on a ground segment, a term used to describe the infrastructure and staff related to the spacecraft control centre, and data management & distribution to users. EO data has shifted from its original centralised and restricted approach for proprietary software processing to a cloud-based service that is free for anyone to access, and efforts have been made to introduce new tools that allow access and processing online on web-based platforms, further widening the community of users.

Following the United States 2008 Consolidated Appropriations Act directing the USGS to make Landsat data freely available to the public, the demand for these datasets surged in the following years and efforts were made to consolidate the Landsat archive as an open provision of standard products, following transparent methods, and focusing on inter compatibility with other missions [51].

In 2010 the USGS began the Landsat Global Archive Consolidation to centralise all the raw imagery, and in 2016 initiated the first collection-based processing of the Landsat 1-8 Level 0 and all newly acquired data—known as Landsat Collection 1— to generate Level 1 data. The success of this Collection and several factors such as improved geolocation using Landsat 8 harmonised with ESA data, novel opportunities from cloud computing, and continuation of the programme through Landsat 9, lead to the decision to process the Landsat data archive again as Landsat Collection 2 [52]. The immense dataset —now in the order of petabytes ($10^{15}$ bytes)— is available for users on web platforms to explore and download[3]. Figure 2.21 shows the dramatic rise in the number of downloads from the Landsat scenes following 2008.



Figure 2.21: Landsat Level 1 scenes downloads from the USGS EROS Archive [53].

---

[3] At https://earthexplorer.usgs.gov or https://registry.opendata.aws/usgs-landsat/

The Landsat Collection 2 is composed by scenes belonging to the Worldwide Reference System (WRS), a global system that catalogues Landsat data by Path and Row numbers. Landsat satellites 1, 2 and 3 followed WRS-1, and Landsat satellites 4,5,7, 8, and 9 follow WRS-2. The WRS-2 grid has 233 paths and 123 rows [54] shown in Figure 2.22.



Figure 2.22: WRS-2 Paths and Rows and UTM zones (Public domain)

This open access era starting in 2008 allowed for the research and EO data users community and researchers to feedback and thrive, thus focusing the EO missions as a global effort of international cooperation. In accordance with this, ESA envisaged a similar open access approach to the data for its missions[4] and for Sentionel-2 products, where it excludes access to raw data (Level 0), but releases the Level-1C (TOA reflectance images) and Level-2A (atmospherically corrected SR) data products to users [55]. Having less than a decade of collected data, but higher spatial resolution than Landsat, the Sentinel-2 collection accounts for a larger size than the Landsat Collection 2, at tens of petabytes.

Considering the exponential growth of the EO data and with it the need to process large geospatial datasets, Google Earth Engine (GEE) platform was released as a cloud-based platform that makes use of Google's high-performance computing resources.

---

[4] At https://dataspace.copernicus.eu/

GEE is accessible through an application programming interface (API) and an associated web-based interactive development environment (IDE), along with a curated catalogue of the most frequently used open-access geospatial datasets [56]. It is a tool designed to tackle society's biggest challenges by simplifying the monitoring, tracking, and management of Earth's environment and resources through easy access to vast data, powerful computing, and advanced analysis tools at the hands of researchers, policy makers, NGOs, field workers, and the public in general.

The methodological approach chosen for the Thesis followed an open data and open-source approach and was therefore based on GEE JavaScript scripts and using the web-based IDE. GEE is available to users registering with a Gmail account.

The methodology used in this Thesis is described in detail in Chapter 3, including the reference to the scripts used for each step. However, some overall background on **Radiometric Corrections**, **Geometric Corrections**, **Data visualisation**, **Data Masking and Data Integration** which comprise the methods related to the Pre-processing pipeline as well as detail relating to **Tier Classification,** are required to provide further context within this current Sub-Chapter of the Thesis. It should be noted that no atmospheric corrections were performed in the present work.

**Tier Classification**: Initial data selection was mostly performed through metadata filtering of suitable datasets from the GEE catalogue[5]. In the catalogue, the Landsat Collection 2 archive is available as directories depending on the mission, the instrument, the level of processing and the Tier (1 or 2). Landsat scenes placed into Tier 1 are considered suitable for time-series processing analysis and include Level-1 Precision Terrain processed data that have well-characterised radiometry and are inter-calibrated across the different Landsat sensors. Scenes not meeting Tier 1 criteria during processing (radial Root Mean Square Error RMSE ≤ 12m with respect to GCPs) are assigned to Tier 2 due to significant cloud cover, insufficient ground control, and other factors [57]. For the case of Sentinel-2, only Level 2-A—Atmospherically corrected— data was used, which is not split in Tiers.

**Radiometric Corrections**: For Landsat MSS Level 0 Tier 2 datasets, radiometric corrections were performed through custom functions developed using a *linear transformation* with calibration coefficients present on each scene metadata — previously computed using [58]— using the general Equations (2.1) (2.1)and (2.2) for each band used.

---

[5] Refer to https://developers.google.com/earth-engine/datasets/catalog/landsat

$$S_{rad} = DN_{raw} * m_{rad} + y_{rad} \qquad (2.1)$$

Where $S_{rad}$ is the radiance at sensor geometry of a given band, $DN_{raw}$ is the raw DN from the dataset, $m_{rad}$ and $y_{rad}$ are the radiance multiplying and additive factor for the band, respectively. These factors are in the scene properties as 'RADIANCE_MULT_BAND_#' and 'RADIANCE_ADD_BAND_#'.

$$TOA_{ref} = S_{rad} * m_{ref} + y_{ref} \qquad (2.2)$$

Where $TOA_{ref}$ is the TOA reflectance of a given band, $S_{rad}$ is the calculated radiance at sensor geometry of a given band, $m_{ref}$ and $y_{ref}$ are the reflectance multiplying and additive factor for the band, respectively. These factors are in the scene properties as 'REFLECTANCE_MULT_BAND_#' and 'REFLECTANCE _ADD_BAND_#'.


**Geometric corrections**: For Landsat MSS Level 0 Tier 2 datasets, when required, shifting and co-registering strategies were implemented. A unique shift to align scenes from the same Path, as a geolocation anomaly was encountered: scenes from the same Path are consecutive acquisitions and should be aligned and partially overlapped. The shift applied can be considered as a vector with horizontal and vertical components that is added to the coordinates of the pixels, translating the grid raster without changing the DN, Radiance or Reflectance values. Equation (2.3) presents the mathematical representation for the shift of coordinates, while Figure 2.23 provides the visual representation of such operation.

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} A \\ K \end{bmatrix} \qquad (2.3)$$

Where $x'$ and $y$ are the new coordinate vector, $I$ is the design matrix (the identity in this case), x and $y$ are the old coordinate vector, and $A$ and $K$ are the components of the shift vector.
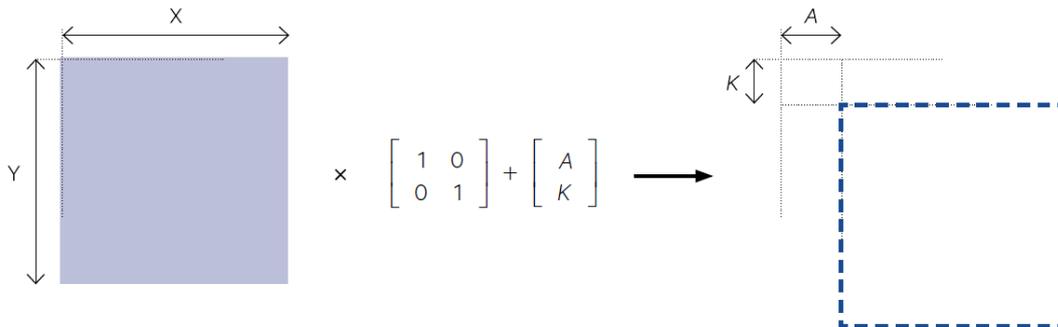


Figure 2.23: Shift operation representation [37, Fig. 3.5]

To calculate a unique shift for the whole misaligned scene, a set of control points are established on the overlapping area of the aligned scene, and coordinates for the points are retrieved in both scenes. A least squares solution for the shift vector is obtained to guarantee the best possible adjustment between the scenes, yielding Equation (2.4) which corresponds to the average difference between the corresponding coordinates for the control points in the two scenes.

$$\begin{bmatrix} \hat{A} \\ \hat{K} \end{bmatrix} = \frac{1}{n} \begin{bmatrix} \sum_{i=1}^{n} (x_i' - x_i) \\ \sum_{i=1}^{n} (y_i' - y_i) \end{bmatrix} \tag{2.4}$$

Where $\hat{A}$ and $\hat{K}$ represent the estimated x and y coordinates for the shift, respectively, $n$ represents the number of control points, $(x_i', y_i')$ represent the coordinates in the misaligned image and $(x_i, y_i)$ the coordinates in the reference image. The GEE function performing the shift is ee.Image.translate[6].

Regarding co-registering, a warping algorithm was applied to shifted images to match the overlap features. This registration algorithm uses a "rubber-sheet" technique, using pixel matching by correlation on each band and applying a local displacement —per pixel, analogous to Equation (2.3)— to match the reference image, and finally performing a resampling method to align to the reference pixel grid. The GEE function performing the shift is ee.Image.register[7] and the one performing the resampling is ee.Image.resample[8].

**Data visualisation**: One way to exploit the non-visible bands of the EMR spectrum is to use them in the colour mixing system 'as if' they were one of the primary colours used in it. The widest used system for colour on screens is the RGB (Red, Green and Blue) additive model, and its YMC (Yellow, Magenta, and Cyan) subtractive counterpart —for printed media—. These models allow any visible colour to be represented by the addition —or subtraction— of the primaries, usually in a relative 8-bit (0-255 colour values) scale, creating a visually uniform three-dimensional colour space. Reflectance values of the bands are scaled in this space on the three primary channels and create a colour composite that responds to our own perception of colour. Figure 2.24 provides a visual representation of how the RGB additive colour system works.

---

[6] https://developers.google.com/earth-engine/apidocs/ee-image-translate
[7] https://developers.google.com/earth-engine/guides/register
[8] https://developers.google.com/earth-engine/apidocs/ee-image-resample

a. The additive primaries (red, green and blue) are added to a black background to give other colours.

b. The subtractive primaries (yellow, magenta and cyan) subtract from a white background to produce other colours.

c. Orientation of RGB colour cube

d. Colour planes on surface of RGB colour cube

Figure 2.24: RGB & YMC Colour System and cube, adapted from [35, Fig. 5.7,5.9]

Any other band assignation to the blue, green and red channel primaries is called a *false colour composite* since features in the resultant colour image do not have their true colours, often involving a previous contrast enhancement before scaling the reflectance values by a *scaling factor* [35]. These false colour composites are useful for visually enhancing the contrast of water, soil, urban areas, forests, crops, among other land cover, between them and within themselves. Some false colour composites are used across sensors using comparable, but not equal bands. For the present work, apart from the true colour RGB composite, the NIR and Agriculture composite AGRI were used to enhance the contrast between landcover classes.

Table 2.3: List of Bands for used colour composites.

| Mission & Instrument | Colour Composite | | | | | |
| | RGB [R,G,B] | Scale Factor | NIR [NIR,R,G] | Scale Factor | AGRI [SWIR1,NIR,B] | Scale Factor |
| --- | --- | --- | --- | --- | --- | --- |
| Landsat 2 MSS* | - | - | [B6-B5-B4] | 0,4 | - | - |
| Landsat 4 MSS* | - | - | [B3-B2-B1] | 0,4 | - | - |
| Landsat 4/5 TM, 7 ETM+ | [B3,B2,B1] | 0,3 | [B4-B3-B2] | 0,4 | [B5-B4-B1] | 0,4 |
| Landsat 8 OLI | [B4,B3,B2] | 0,3 | [B5-B4-B3] | 0,4 | [B6-B5-B2] | 0,4 |
| Sentinel-2 MSI | [B4,B3,B2] | 0,3 | [B8-B4-B3] | 0,4 | [B11-B8-B2] | 0,3 |

*MSS instrument does not have a sensed Blue Band, so images cannot be interpreted in real colour composite.

Figure 2.25 provides a good example of a set of landscapes with a dominant type of landcover and how the different composites affect the visualisation.



Figure 2.25: RGB spectral reflectance composites [59, Fig. 4]

**Data Masking**: One of the key issues for image processing (e.g. classification) is ensuring there are no clouds, cloud shadows, haze, smoke, dust or other anomalies like distorted scene edges, high altitude objects, etc. Masking refers to the process of flagging of pixels affected by these unfavourable conditions and generating a mask, or a parallel grid to the pixels with a 1/valid or 0/invalid value. Applying this mask to the image excludes the flagged invalid pixels from the analysis and visualisation. To give some context to this work AOI —for a tropical high altitude mountainous area—cloud-free imagery is a wish rarely, if ever, granted.

Many algorithms for cloud detection are available in the literature and many are operational for Landsat products with very good removal accuracies [60], as well as for Sentinel-2 [61]. Most rely on bands and/or processing information not available for the MSS sensor, so cloud masking for these older datasets is particularly challenging. For the Landsat MSS Level 0 Tier 2 datasets, after applying the radiometric and geometric corrections, a custom cloud masking function was developed by thresholding, buffering mask pixels and translating the buffer to mask shadows.

*Thresholding* is applied by setting a high reflectance value beyond which the pixel is considered a cloud pixel. There are trade-offs in this simple algorithm regarding the threshold value, as some surfaces, like built-up areas (gravel roads, farms, greenhouses) can have high reflectance too and might be flagged as cloud, thus loosing valid data. The selection of a threshold number is then an iterative process of selecting reflectance value and visually inspecting the output to assess whether this is targeting troublesome cloudy areas. To balance this process, *Buffering* allows to expand an initial allocation of cloud pixels and flag the surrounding ones as cloud as well, as clouds are usually surrounded by a gradient of haze. Finally, to deal with *cloud shadow*, the pixel mask was translated using the shadow azimuth angle —Defined as 90° minus the solar azimuth angle— and a preset fixed distance, also iteratively set. Figure 2.26 presents a conceptual schema of the cloud masking algorithm and it's three steps.



Figure 2.26: Custom cloud masking algorithm developed on GEE.

For later datasets, namely TM and ETM+, level 2 data products already incorporate a Quality band (QA_PIXEL) representing a bit flag for Dark Dense Vegetation, Cloud, Cloud Shadow, Adjacent to Cloud, Snow or Water. Landsat OLI level 2 data products incorporate an extended version of this Quality band with the above-mentioned bit flags with Dilated Clouds instead of Adjacent to Cloud, plus Cirrus clouds and levels of confidence (None, Low, Medium and High) for these two along with Cloud Shadow and Snow/Ice. Sentinel-2 Level 2-A data products also incorporate a Quality band (QA60) with bit flags for Clouds and Cirrus at 60m resolution. In addition to this, the S2 Cloudless dataset made available by Copernicus contains the pixel cloud

probability at 10m resolution [62], for most of the Sentinel-2 scenes, and a threshold probability can be set to mask out clouds.

While increasing in complexity, none of the Cloud Masking algorithms is foul proof and they are largely dependent on the context-specific occurrence, shape and extension of clouds and haze. The selection of the cloud masking can significantly impact classification outputs particularly in cloud-prone areas [61]. It is safe to note that although great care was taken regarding cloud removal, haze affected pixels could not be removed entirely and, unlike some recent data, no validation for this process could be implemented for the custom or available algorithms removal efficiency results.

**Data integration**: After performing all the required corrections and masking for a given dataset scene, the extension of the AOI requires the integration of several of them, making it necessary to choose an approach to perform this operation. Homologous (i.e. same sensor) data integration can be performed through compositing or mosaicking. In a broad sense, *compositing* refers to the process of combining spatially overlapping images into a single image based on an aggregation *function*. *Mosaicking* refers to the process of spatially assembling image datasets to produce a spatially *continuous* image[9]. Essentially, compositing is using a function to define the value of the overlapping pixels—A median, minimum or maximum value—while mosaicking is defining the order in which the tiles are stacked: the last tile to be stacked will define the value for the pixel, therefore the value for the observation remains unchanged.

Compositing is preferred when a large amount of data is available and aggregation functions like the median or minimum can indirectly remove cloudy or hazy pixels—as reflectance outliers—, but it is evident that the resulting image is not a single true observed value, but a statistical value derived from a set of observations. Mosaicking on the other hand, comes in handy when only a handful of observations are available and the final image is constructed 'stitching' the best available imagery and stacking it in an objective ranking order, like the number of pixels removed from cloud masking. The latter was the preferred approach for the present work, combined with compositing only when necessary. However, for the mosaicking approach to be implemented effectively, acquisition from different dates must be subject to and additional colour correction.

Considering all the principles introduced in Sub-Chapter 2.2 and the present, even if the landcover of the Earth's surface and the instrument remain the same, slight differences in scene lighting due to the day of the year, atmospheric conditions or differences in plant phenology and soil humidity (due to rain), among other factors

---

[9] https://developers.google.com/earth-engine/guides/ic_composite_mosaic

can have an impact on the resulting reflectance measured for a given pixel even at the smallest temporal resolution. This effect is even present when a pixel is acquired on two consecutive swaths or Paths as they are always partially overlapped. One common correction for this issue is to perform a histogram matching, taking one acquisition as a reference and others as targets, stretching or compressing their reflectance histograms to match the reference applying a colour transformation to the targets.

The algorithm requires computing both reference and target images histogram and cumulative distribution function for a given band, using the probabilities to map the values from the reference to the targets. This correction is also performed when dealing with images acquired from different sensors. Figure 2.27 presents an example of colour correction by histogram matching of high resolution ortho mosaics (3cm/pixel) collected by Unmanned Aerial Vehicle (UAV) around Oakland, CA area.
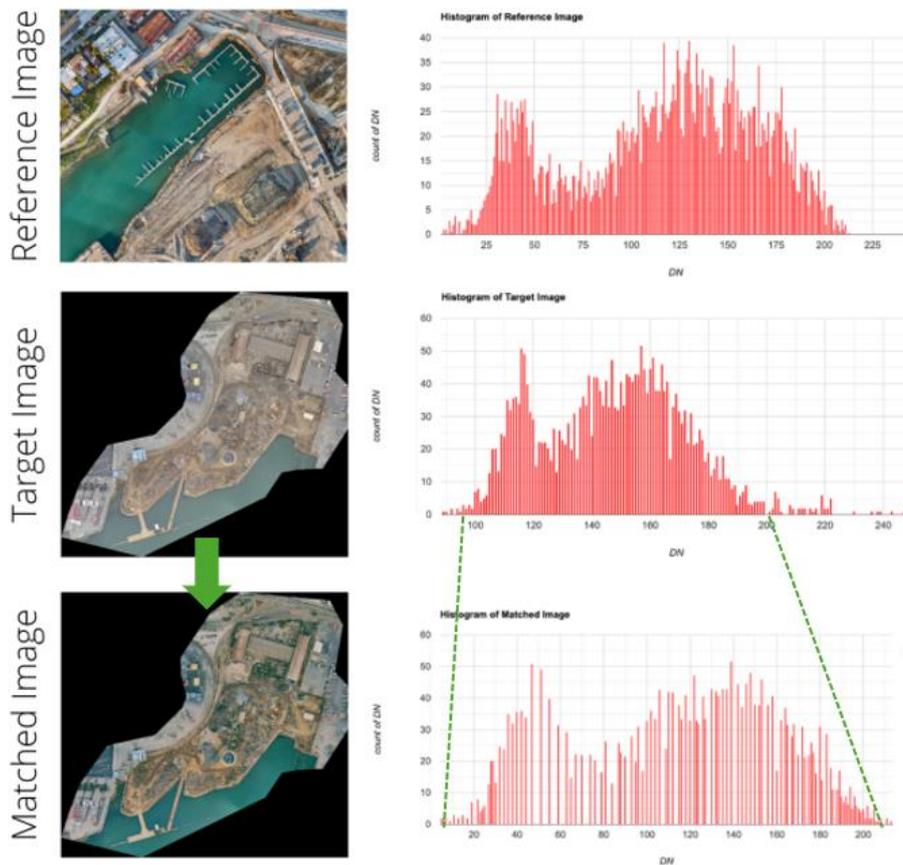


Figure 2.27: Histogram matching example, adapted from[63]

## 2.4.   Land Cover Classification

Classification is a general technique for predicting the class or category of an observation based on its features or attributes. Image classification is commonly used for feature or pattern recognition and usually involves converting an image into a *thematic map* using a *classification algorithm*. The suitability of an image classification algorithm for a particular interpretation task will depend on the nature of the application and the resolution of the features to be identified in the image. Land cover mapping of remotely sensed imagery is frequently based on image classification since all the available bands in an image can help to distinguish between different land cover types [35], by means of their spectral signature differences, as seen in Sub-Chapter 2.2. This type of classification is often referred to as *spectral pattern recognition*.

This Thesis employed an initial unsupervised data clustering *K-means* algorithm for the Landsat MSS data and a supervised *Random Forest* (RF) classifier for all datasets with a *stratified sample* design. The present Sub-Chapter will provide an overview of **image classification methods**, describe the theory behind **machine learning**, the classification **algorithms used**, the **sampling design** and some **examples of Land use/Land cover (LULC) classification**.

**Image classification** can be pixel-based or object-based. When pixel-based this means that it can only consider the individual spectral or band values for a pixel. Whereas if it is object-based it can consider both the values and the values correlation with neighbouring pixels, adding the context of the pixel's surroundings to group pixels in objects with similar geometrical, spatial and spectral properties. Classification can be supervised or unsupervised, meaning that the sorting algorithm can be based on a training sample (a sample dataset of pixels with their corresponding classes, like field data or interpreted imagery) or it can be based merely on the spectral values with no other input, meaning the algorithm will group pixels by statistical methods following the minimisation of a given function, like the Euclidean distance between the points in the bands hyperspace. Supervised classification usually involves setting aside a percentage of the sample data —not involved in the training— to test the model's accuracy on unseen data, this is known as an accuracy assessment and is explained in more detail in the next Sub-Chapter.

**Machine Learning**: The subfield of Artificial Intelligence (AI) that deals with the source of knowledge, automatically extracting relevant information from data and using it to analyse new data. Learning is achieved when performance for a given task improves with experience, allowing the program to automate itself, making informed decisions on new, unseen data. Machine learning differs from traditional

programming as the output is known, and the objective is not the output itself but rather the program or function that data needs to be transformed or processed to obtain this outcome, as shown in Figure 2.28. For this specific case, the data is the pixels with all its spectral band values and the output is the pixel classification label, that is, the pixel LULC class.



Figure 2.28: Traditional programming approach vs. Machine Learning [64].

**K-means**: This thesis employed an unsupervised machine learning algorithm, also known as hard partitioning around the mean method, to cluster the image pixels from Landsat MSS Level 0 Tier 2 datasets to determine which kinds of land cover the algorithm groups together 'naturally', based on their spectral signatures. The only parameter to set for the algorithm is the number of classes or *clusters* ($K$). This iterative algorithm is based on minimising an objective function defined as the Euclidean distance between the samples of a cluster and their respective centroid [65], in the form of Equation (2.5).

$$L = \sum_{i=1}^{m} \sum_{k=1}^{K} \omega_{ik} \|x_i - \mu_k\|^2 \tag{2.5}$$

Where $L$ is the distance to the centroid (minimising function), $m$ is the number of bands, $K$ is the number of classes as well as centroids, $x_i$ = pixel value on band $m$, $\mu_k$= centroid value on band $m$, and $\omega_{ik}$ is the membership defines as in Equation (2.6).

$$\omega_{ik} = \text{membership} \begin{cases} 1 \ if \ k = argmin_j \|x_i - \mu_k\|^2 \\ 0 \ otherwise \end{cases} \tag{2.6}$$

On the first step of the algorithm, the $K$ centroids are randomly chosen, and on the second, pixel data points are assigned to the respective class minimising the distance $L$. For the third step, centroids are recomputed according to Equation (2.7).

$$\mu_k = \frac{\sum_{i=1}^{m} \omega_{ik} x_i}{\sum_{i=1}^{m} \omega_{ik}} \tag{2.7}$$

The iteration of the second and third step continues until the convergence of centroids values is reached, when the previous and recomputed values do not differ meaningfully. The final solution is stable and independent from the initial centroids. The GEE function performing the K-means algorithm is ee.Clusterer.wekaKMeans[10]. Figure 2.29 represents a 2-dimensional (e.g. 2 band values) example of a K-means algorithm implementation for two classes or clusters (Blue and Red).



Figure 2.29: K-means algorithm example. Cluster centroids are shown as crosses. (a) Original dataset. (b) Random initial cluster centroids. (c-f) Two iterations of K-means [66].

**Random Forest (RF)**: A supervised machine learning algorithm, widely used because of its flexibility, composed of an ensemble of *decision trees* trained with a *bagging* method, that classifies the image by *majority voting*.

A *decision tree* is a set of rules or conditions that split the data according to their feature values, in this case the spectral band values of each labelled pixel. The selection process can be described as a sequence of binary selections[67], each node of the tree splits the data according to whether they meet the rule or not, following different paths growing different branches that lead to a classification on the last node, referred to as a leaf node. These sets of rules are the product of the training by the Classification and Regression Tree (CART) algorithm. While decision trees are common supervised learning algorithms, they can be prone to problems such as bias and overfitting.

Ensemble methods, also known as committee methods, are made up of a set of individual classifier algorithms—such as decision trees—, where their predictions are aggregated to identify the most popular, that is, the *majority voting* result. One of the most recognised ensemble methods is *bagging*, also referred to as bootstrap aggregation, where a random sample of data in a training set is selected with replacement [68]. This means that the individual data points can be chosen more than once, like randomly drawing out a certain number of data points from a bag and then

---

[10] https://developers.google.com/earth-engine/apidocs/ee-clusterer-wekakmeans

putting them back, to repeat the process again. After several data samples are generated, the individual models are then trained independently, and classification results from the majority vote of those predictions, yielding a more accurate estimate and reducing variance, particularly when the individual classifiers are uncorrelated with each other [69].

RF algorithm is an extension of the bagging method as it employs both bagging and *feature randomness* to create an uncorrelated forest of decision trees [70]. Feature randomness is achieved by extending the bagging concept to the features—in this case the spectral band values—, and randomly selecting a subset of bands to perform the classification for each tree, ensuring low correlation among them. From the training sample, one-third of it is set aside as validation data, known as the out-of-bag sample, which is used for cross-validation once the majority vote is casted. By accounting for all the potential variability in the data, RF can reduce the risk of overfitting, bias, and overall variance, resulting in more precise predictions. Figure 2.30 shows an example of an RF classifier.
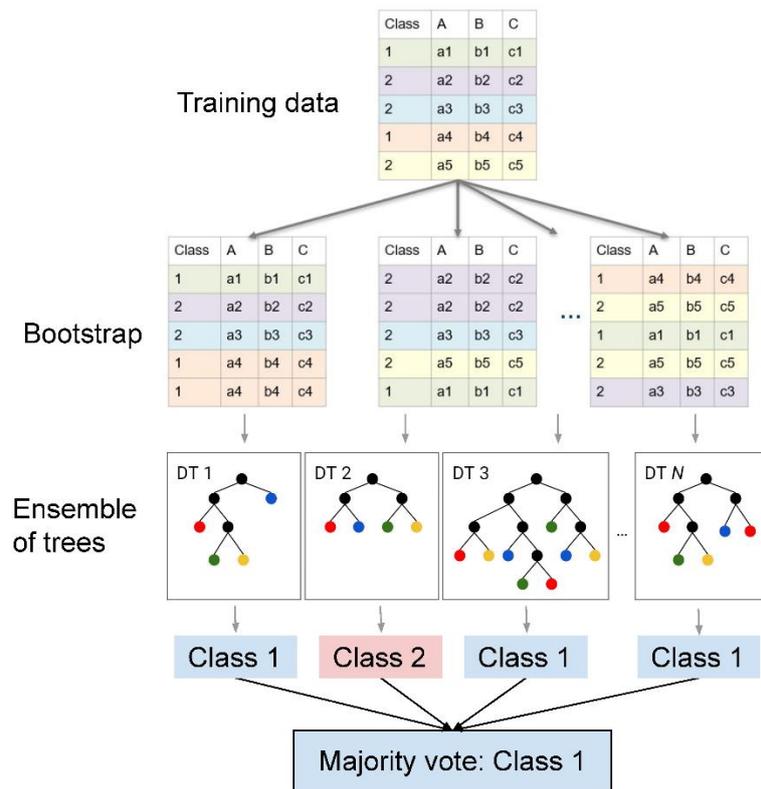


Figure 2.30: RF algorithm diagram [71].

RF algorithms have three main hyperparameters—the parameters of the algorithm itself—, which need to be set before training: the number of trees, the number of nodes, and the number of features sampled. The GEE function that creates an RF classifier is

ee.Classifier.smileRandomForest[11]. Other widely used algorithms for satellite image classification are Gaussian Maximum Likelihood (GML), Support Vector Machines (SVMs), Artificial Neural Networks (ANNs), and Deep Learning (DL). An overview on ML algorithms with their strengths and weaknesses is presented in Figure 2.31.

| Algorithm | Strengths/characteristics | Weaknesses |
|---|---|---|
| *Artificial Neural Networks* Non-parametric | • Manage well large feature space <br>• Indicate strength of class membership <br>• Generally high classification accuracy <br>• Resistant to training data deficiencies—requires less training data than DT | • Needs parameters for network design <br>• Tends to overfit data <br>• Black box (rules are unknown) <br>• Computationally intense <br>• Slow training |
| *Clustering* (partitioning) | • Do not need previous knowledge <br>• Do not need samples | • Cluster-class correspondence not assured <br>• Complex identification of classes <br>• Computationally intense |
| *Decision trees* Non-parametric | • No need of any kind of parameter <br>• Easy to apply and interpret <br>• Handle missing data <br>• Handle data of different types (e.g. continuous, categorical) and scales <br>• Handle non-linear relationships <br>• Insensitive to noise | • Sensitive to noise <br>• Tend to overfit <br>• Not as good as others in large feature spaces <br>• Large training sample needed |
| *Gaussian Maximum likelihood* Parametric | • Simple application <br>• Easy to understand and interpret <br>• Predicts class membership probability | • Parametric <br>• Assumes normal distribution of data <br>• Large training sample needed |
| *Support Vector Machines* Non-parametric | • Manages well large feature space <br>• Insensitive to Hughes effect <br>• Works well with small training dataset <br>• Does not overfit | • Needs parameters: regularization and kernel <br>• Poor performance with small feature space <br>• Computationally intense <br>• Designed as binary, although variations exist |
| *Random Forests* Non-parametric | • Capacity to determine variable importance <br>• Robust to data reduction <br>• Does not over-fit <br>• Produces unbiased accuracy estimate <br>• Higher accuracy than DT | • Decision rules unknown (black box) <br>• Computationally intense <br>• Needs input parameters (#trees and #variables per node) |
| *Bagging* | • Provides measures of classification confidence <br>• Does not overfit | • Complex incomprehensible classifiers |
| *Boosting* | • Provides measures of classification confidence <br>• Does not overfit <br>• Robust to noise | • Stops if a classifier achieves zero training set error <br>• Complex incomprehensible classifiers <br>• Ineffective if excessive error in training sample |

Figure 2.31: Characteristics of algorithms used for large-area land cover characterization with time-series optical data [72, Tbl. 4].

**Sampling design**: As conceptualised before, any ML algorithm must be trained on outputs, in this case, on labelled or already classified pixels. The true power of applying these algorithms resides on their ability to generalise, that is to correctly classify large amount of data from a *sample* dataset. This concept has its parallel in statistics, where drawing general conclusion about a population based on a sample is called statistical inference [73]. A rigorous approach from a statistical framework has surged as a collection of best practices regarding sampling methodologies and their relationship with the generalization of the accuracy assessments [74].

The practical implications of adopting this design-based sampling are that there needs to be enough data points to be able to extend the sample results to the population, and more importantly, all the pixel classes should be adequately represented in terms of their total area fraction in the mapped area. This means that if there is a LULC dominant class that represents half the mapped area and another that represents only one percent of the total, the sampled pixels must be enough to be statistically significant for the total population, and more sampled pixels of the dominant class are

---

[11] https://developers.google.com/earth-engine/apidocs/ee-classifier-smilerandomforest

required. In statistics, the subdivision of the population in brackets is referred to as strata, which in the present case corresponds to pixel LULC classes.

A *stratified sample* means a sample of pixels which proportions mimic the class distribution in the mapping area, as the exact proportions are unknown, a good priori estimate should be provided. This proportion estimate will act as a weight for each class, and along the desired variance to be attained, the total sample size is determined by Equation (2.8)[75, Eq. 5.25]

$$n \approx \left( \frac{\sum_i^M W_i S_i}{S(\hat{O})} \right)^2 \tag{2.8}$$

Where $n$ is the minimum sample size, $M$ is the number of LULC classes, $W_i$ is class $i$ estimated proportion of mapped area, $S(\hat{O})$ the standard deviation of target overall accuracy, and $S_i$ is the standard deviation of the stratum or class, defined as in Equation (2.9).

$$S_i = \sqrt{U_i (1 - U_i)} \tag{2.9}$$

Where $U_i$ is class $i$ **target user accuracy**, defined a priori, not to be mistaken with the estimated one $\hat{U}_i$. All the accuracy metrics will be explained in detail in the next Sub-Chapter.

**Land Use Land Cover Maps (LULC)**: Remotely sensed images have allowed for the successful mapping of the Earth's surface at different granularities, both in terms of spatial resolution and in classification complexity, and at different temporal scales. From the manual approach based on surveying or visual interpretation of aerial images, to the automated, by means of supervised algorithms that are fed by satellite image interpretation and/or field surveys: the mapping of landcover types by Machine Learning is now operational. As presented in Sub-Chapter 2.2, given the open-access and provision of robustly calibrated science-grade data from Landsat and Sentinel-2, it is possible to produce global information products and services—such as general or thematic LULC maps— that are needed for the management of terrestrial resources and ecosystems, as well as to document and understand ecosystem change [47], [76].

Global coverage LULC products have been released like the GlobLand30 (2010), ESA Copernicus Global Land Service Land Cover 100 m dataset (2015–2019)[77], ESA's 10 m World Cover 2020 [78]/2021[79], and Esri's 10 m 2020 Land Cover[80]. All these products harness ML algorithms trained on hand-labelled pixels to achieve the classification. A novel approach of near real-time global mapping is Google's Dynamic World, providing a 10 m LULC map that updates with every new available Sentinel-2 scene (every 5 days) and incorporates a likelihood approach based on previous covers [81].

Systematic evaluation of the latter three global 10 m resolution LULC maps for the year 2020 showed that they have near-equal area estimates for water, built area, trees and crop. Relative to one another, World Cover is biased towards over-estimating grass, Esri towards shrub and scrub, and Dynamic World towards snow and ice cover [82]. With their Overall Accuracies (OAs) ranging from 63% to 75% —see Figure 2.32— when compared to ground truth datasets the authors recommend to critically evaluate each LULC product with reference to the application purpose.



Figure 2.32: Accuracy of global 10 m LULC land maps across classes and continents [82].

A recent review of LULC mapping products found 56 global products, 16 continental and 35 national products, with the spatial resolution ranging from 1 m to 100 km, and with significant differences in classification schemes; this being one of the primary factors leading to inconsistency and uncertainty [83]. The review also found most global products adopt the Land Cover Classification System (LCCS) [84], followed by the International Geosphere-Biosphere Programme (IGBP) [85], while most regional and national ones adopt the USGS classification system—also known as the Anderson classification system [86]—, or the European Community Collaboration for Information on the Environment (CORINE) Land Cover Classes (CLC) [87], or local variations. Figure 2.33 shows the differences in the classification schemas and number of classes between global LULC products.

These reviews highlight one key aspect of LULC mapping, that no matter the scale, coverage or resolution, there is not a one-size-fits-all-purposes map, and therefore class schema definition and individual accuracies should be reviewed when extracting information from LULC products.

Figure 2.33: Global general LULC product classification system and number of classes within the specified general category [83].

In Colombia there is a national LULC product based on a modified CLC schema [88] for the years 2000-2002, 2005-2009, 2010-2012 [89] and 2018 [90], all of which are based on interpreted satellite imagery. Other national thematic LULC include ecosystem mapping, available for the years 2007 [91] and 2017 [92], among other class-focused products like deforestation (forest/no forest maps), glacier extent, etc. Recent LULC products exploiting ML algorithms and cloud computing are being published by researchers with a special focus on robust and replicable methodologies and accuracy assessments. This is the case for the 30 m map for the Colombian Andes and Amazon regions with Landsat-8 imagery centred in the year 2018, based on RF and SVM classifiers, all performed on GEE with open-access to training and validation areas [93]. Figure 2.34 presents the SVM map results, which are accompanied by their respective 95% Confidence Interval (CI); The Andean Shrub and/or Herbaceous class —where Páramo Ecosystem would be included for this region— is reported to be 1'751.050,69 ± 245.318 Ha.

Figure 2.34: SVM-classified product for the Colombian Amazon (Left) and Andes (Right) regions for the target year 2018 [93].

For comparison, the area for Páramo ecosystem in Colombia reported in the article from Figure 2.1(Right) based also on Landsat-8 data is about 1'044.943 Ha [7], which is significantly less than one would expect. Even though one article refers to the whole country while the other article includes just the Andean region (excluding the Sierra of Santa Marta, northernmost Páramo areas in Figure 2.1 and Figure 2.4), discrepancies in area estimation can always be expected. According to the methodological pipeline followed, the same initial dataset can lead to different classification results. Hence the need for a careful definition of the classification scheme, the datasets used, the pre-processing methods employed, the sampling design and dataset on which the ML algorithm is trained and tested, and the accuracy assessment metrics per class reported.

## 2.5.  Accuracy Assessment

Classification is always the result of applying a *model* to sort *observations* in different groups or *classes*, that exhibit some common pattern. From the human interpreters to a hard rule-based model, to a complex ML algorithm, the classification will always be subject to real-world error sources and constraints: the acquisition signal vs. the noise, the resolution, the corrections (radiometric, geometric, atmospheric), the interpretation—both in terms of interpreter bias and class scheme definition—, the outliers, the pixels at the edge between the classes (mixed pixels), etc. Therefore, any classification result is subject to errors both from the data and from the model itself.

ML theory states that for any algorithm that performs very well on a given task, there is another task where the same algorithm will perform poorly [64]. Therefore, there is no perfect learner algorithm for all possible tasks. Conversely, a very good human interpreter for vegetation land cover can perform poorly on coral-reef or marine ecosystem identification, or at control point collimation from a photogrammetric survey. Although high accuracy for a given classification task is always desirable, perfection is highly discouraged, as it represents overfitting to training data and most likely a lack of *generalisation* from the model.

A fundamental step of any classification process is to evaluate errors with respect to an alternative dataset that has been classified, giving the latter the status of reality or *ground truth* (True class) and the classification results as an estimate (Predicted class), with the objective of quantifying the difference between the two. Given the limitations in availability of ground truth datasets ideal for a given classification task, strategies like hold-out data from the training dataset are implemented [94]. This means that once a manually labelled (e.g. classified by an interpreter) dataset is available, a fraction of the data is set aside from the training, and it is only used later as a ground truth to determine the model's accuracy. The hold-out dataset is referred to as the *test* dataset, although sometimes it is also referred to as the *validation* dataset. The confusion arises from the tuning of the ML model hyperparameters, that usually requires a fraction of the training dataset for cross-*validation*, but it is usually performed by the algorithm itself, or it is used for early model selection when there is no information about which algorithm has proven effective at a given task. From here on, the set dedicated to evaluating the model's performance will be referred to as the test set.

For a LULC thematic map to be operational, the products quality must be assessed by the means of its accuracy under different metrics, all of which can be extracted from the *error matrix*, also referred to as the confusion matrix [95]. The error matrix is a simple cross-tabulation of the class labels allocated by the classification of the remotely sensed data against the reference data for the sample sites [96]. In other words, it is simply a class-to-class comparison in the number of pixels from the classified test

dataset vs. the model's resulting class on the same dataset, commonly labelled Predicted class as rows vs. True class as columns. As mentioned before with respect to the sampling process, the test dataset must be of adequate size and with an adequate proportion of the pixels belonging to each class, if significant inference is to be drawn from it [97]. Because of its ease of interpretation and valuable descriptive information, the error matrix remains a cornerstone of the analysis protocol [74]. Best practices derived from literature recommend a series of transformations from the pixel count matrix to an area-weighted one, from where confidence intervals for the area of each class on the map can be extracted [96]. Figure 2.35 presents the error matrix as sample counts or as area proportions obtained by Equation (2.10).

Error matrix of sample counts, $n_{ij}$.

| Class | 1 | 2 | ... | q | Total |
|---|---|---|---|---|---|
| 1 | $n_{11}$ | $n_{12}$ | ... | $n_{1q}$ | $n_{1\cdot}$ |
| 2 | $n_{21}$ | $n_{22}$ | ... | $n_{2q}$ | $n_{2\cdot}$ |
| ⋮ | ⋮ | ⋮ | ⋱ | ⋮ | ⋮ |
| q | $n_{q1}$ | $n_{q2}$ | ... | $n_{qq}$ | $n_{q\cdot}$ |
| Total | $n_{\cdot 1}$ | $n_{\cdot 2}$ | ... | $n_{\cdot q}$ | $n$ |

Error matrix of estimated area proportions, $\hat{p}_{ij}$

| Class | 1 | 2 | ... | q | Total |
|---|---|---|---|---|---|
| 1 | $\hat{p}_{11}$ | $\hat{p}_{12}$ | ... | $\hat{p}_{1q}$ | $\hat{p}_{1\cdot}$ |
| 2 | $\hat{p}_{21}$ | $\hat{p}_{22}$ | ... | $\hat{p}_{2q}$ | $\hat{p}_{2\cdot}$ |
| ⋮ | ⋮ | ⋮ | ⋱ | ⋮ | ⋮ |
| q | $\hat{p}_{q1}$ | $\hat{p}_{q2}$ | ... | $\hat{p}_{qq}$ | $\hat{p}_{q\cdot}$ |
| Total | $\hat{p}_{\cdot 1}$ | $\hat{p}_{\cdot 2}$ | ... | $\hat{p}_{\cdot q}$ | 1 |

Map categories are the rows while the reference categories are the columns.

Figure 2.35: Error matrix configuration [98, Tbl. 1,2]

$$\hat{p}_{ij} = W_i \frac{n_{ij}}{n_{i\cdot}} \tag{2.10}$$

Where $\hat{p}_{ij}$ is the unbiased estimator of the proportion of area in cell $i$, $j$ of the error matrix, $n_{ij}$ is the number of pixels in cell $i$, $j$, $n_{i\cdot}$ is the total number of pixels predicted for class $i$ (row total), and $W_i$ is the proportion of area mapped as class $i$ defined by Equation (2.11).

$$W_i = \frac{A_{m,i}}{A_{tot}} \tag{2.11}$$

Where $A_{m,i}$ is the mapped area of class $i$ and $A_{tot}$ is the total area of the map. The mapped area of a given class is also referred to as the **pixel-count area**, for it is the result of the number of pixels classified in the map by the pixel area.

These two configurations of the error matrix allow for the computation of the accuracy metrics and their *area-weighted* versions, the latter being an unbiased estimator for the accuracies that allow for variance and confidence intervals to be computed.

The most common accuracy metrics used in LULC classification are [98]:

- Overall Accuracy ($O$): Proportion of the area mapped correctly. It provides the user of the map with the probability that a randomly selected location on the map is correctly classified. It results from the summation of the diagonal elements over the total pixels, as presented in Equation (2.12) for the sample counts, and in Equation (2.13) for the area-weighted overall accuracy $\hat{O}$ from the area proportion matrix.

$$O = \sum_{j=1}^{q} n_{jj} \tag{2.12}$$

$$\hat{O} = \sum_{j=1}^{q} \hat{p}_{jj} \tag{2.13}$$

Where $q$ is the number of classes.

- User's accuracy: Proportion of the area mapped as a particular class that matched the true class. User's accuracy is the complement of the probability of *commission error*, that is the error of mistaking a mapped class as another class. For this metric, both the sample counts and area proportions matrix yield the same result as shown in Equation (2.14).

$$\hat{U}_i = \frac{n_{ii}}{n_{i\cdot}} = \frac{p_{ii}}{p_{i\cdot}} \tag{2.14}$$

- Producer's accuracy: Proportion of the area that is a particular true class that is also mapped as that category. Producer's accuracy is the complement of the probability of *omission error*, that is the error of mapping a true class as another class.

$$P_j = \frac{n_{jj}}{n_{\cdot j}} \tag{2.15}$$

$$\hat{P}_j = \frac{\hat{p}_{jj}}{\hat{p}_{\cdot j}} \tag{2.16}$$

Where $n_{\cdot j}$ and $\hat{p}_{\cdot j}$ are the total number of true pixels or true pixel proportions for class $j$ (column total), respectively.

The estimated variances $\hat{V}$ for the unbiased or area-weighted metrics are defined by Equations (2.17),(2.18) and (2.19), respectively [98].

$$\hat{V}(\hat{O}) = \sum_{i=1}^{q} W_i^2 \hat{U}_i (1 - \hat{U}_i)/(n_{i\cdot} - 1) \tag{2.17}$$

$$\hat{V}(\hat{U}_i) = \sum_{i=1}^{q} \hat{U}_i (1 - \hat{U}_i)/(n_{i\cdot} - 1) \tag{2.18}$$

$$\hat{V}(\hat{P}_j) = \frac{1}{\hat{N}_{.j}^2}\left[\frac{N_{j.}^2(1-\hat{P}_j)^2\,\hat{U}_j(1-\hat{U}_j)}{n_{j.}-1} + \hat{P}_j^{\,2}\sum_{i\neq j}^{q} N_{i.}^2\frac{n_{ij}}{n_{i.}}\left(1-\frac{n_{ij}}{n_{i.}}\right)/(n_{i.}-1)\right] \quad (2.19)$$

Where $\hat{V}(\hat{O})$ is the overall accuracy's estimated variance, $\hat{V}(\hat{U}_i)$ the user's accuracy estimated variance of class $i$, and $\hat{V}(\hat{P}_j)$ producer's accuracy estimated variance of class $j$. The variable $q$ is the number of classes, $W_i$ is the proportion of area mapped as class $i$ defined by Equation (2.11), while $n_{i.}$ is the total number of pixels predicted for class $i$ (row total). Finally, $\hat{N}_{.j} = \sum_{i=1}^{q}\frac{N_i}{n_{i.}}n_{ij}$ corresponds to the marginal total number of estimated reference class $j$'s pixels with $N_{j.}$ and $N_{i.}$ is map class j and $i$ marginal total, respectively, and $n_{j.}$ map class $j$ total number of sample units (column total). As $N_{j.}$ is the number of class j pixels in the entire map, it coincides with the definition of $W_i$, the same happens to $N_{i.}$, therefore $N_{j.} = W_j$ and $N_{i.} = W_i$. Considering $\hat{N}_{.j}$ can be expressed as $\hat{N}_{.j} = \sum_{j=1}^{q}\hat{p}_{ij} = \hat{p}_{.j}$ Equation (2.19) can be rewritten as Equation (2.20).

$$\hat{V}(\hat{P}_j) = \frac{1}{\hat{p}_{.j}^{\,2}}\left[\frac{W_j^2(1-\hat{P}_j)^2\,\hat{U}_j(1-\hat{U}_j)}{n_{j.}-1} + \hat{P}_j^{\,2}\sum_{i\neq j}^{q} W_i^2\frac{n_{ij}}{n_{i.}}\left(1-\frac{n_{ij}}{n_{i.}}\right)/(n_{i.}-1)\right] \quad (2.20)$$

Confidence intervals can be built around the accuracy estimators as a mean from a normally distributed variable with z=1,96 for CI=95%, multiplying the Standard Error (SE) as shown by Equations (2.21), (2.22), and(2.23):

$$CI_{95}(\hat{O}) = \hat{O} \pm 1{,}96 * SE_{\hat{O}} = \hat{O} \pm 1{,}96\sqrt{\hat{V}(\hat{O})} \quad (2.21)$$

$$CI_{95}(\hat{U}_i) = \hat{U}_i \pm 1{,}96 * SE_{\hat{O}} = \hat{U}_i \pm 1{,}96\sqrt{\hat{V}(\hat{U}_i)} \quad (2.22)$$

$$CI_{95}(\hat{P}_j) = \hat{P}_j \pm 1{,}96 * SE_{\hat{O}} = \hat{P}_j \pm 1{,}96\sqrt{\hat{V}(\hat{P}_j)} \quad (2.23)$$

Finally, $\hat{A}_j$ the unbiased estimator for the area of class $j$, from now on referred to as **Accuracy derived area,** can be expressed as the total area multiplied by the unbiased estimator of the area proportion $\hat{p}_{.j}$ as shown on Equation (2.24), the estimated SE for $\hat{p}_{.j}$ is defined on Equation (2.25) [74, Eq. 20]

$$\hat{A}_j = A_{tot} * \sum_{i=1}^{q} W_i\frac{n_{ij}}{n_{i.}} = A_{tot} * \hat{p}_{.j} \quad (2.24)$$

$$\hat{V}(\hat{p}_{.j}) = \sum_{i=1}^{q}\frac{W_i^2\hat{p}_i(1-\hat{p}_i)}{n_{i.}} \quad (2.25)$$

Where $\hat{p}_i = \hat{p}_{ij}/\hat{p}_{i.} = \hat{p}_{ij}/W_i$, replacing these values in Equation (2.25) results in Equation (2.26), and considering the variance for $\hat{A}_j$ is the total area by the area

proportion variance, the SE for $\hat{A}_j$ is described by Equation (2.27). The confidence interval for the Accuracy derived area is presented by Equation (2.28).

$$\hat{V}(\hat{p}_{\cdot j}) = \sum_{i=1}^{q} \frac{W_i \hat{p}_{ij} - \hat{p}_{ij}^2}{n_{i \cdot}} \tag{2.26}$$

$$SE_{\hat{A}_j} = A_{tot} * SE_{\hat{p}_{\cdot j}} = A_{tot} * \sqrt{\hat{V}(\hat{p}_{\cdot j})} \tag{2.27}$$

$$CI_{95}(\hat{A}_j) = \hat{A}_j \pm 1{,}96 * SE_{\hat{A}_j} \tag{2.28}$$

Where $A_{tot}$ is the total area of the map, $q$ is the number of classes, $W_i$ is the proportion of area mapped as class $i$ defined by Equation (2.11), $\hat{p}_{ij}$ is the unbiased estimator of the proportion of area in cell $i, j$ of the error matrix defined in Equation (2.10), $n_{i \cdot}$ is the total number of pixels predicted for class $i$ (row total).

## 2.6.   Change Detection & Temporal Trajectories

The Earth's surface is in a state of continuous change and so are its ecosystems, both by natural and anthropogenic effects. These changes can be subtle, gradual, seasonal or abrupt, and as observed by the satellites they can be considered as continuous spectral changes or categorical—LULC class—changes. Previous Sub-Chapters have dealt with how a classification of remotely sensed data is performed, one snapshot at a time and in a statistically rigorous way, but now this Sub-Chapter will review how change is monitored.

This Thesis employed a *bi-temporal post-classification* comparison for the Landsat MSS and Sentinel-2 MSI datasets, and a *post-classification temporal trajectory* for the Landsat TM, ETM+ and OLI datasets, all of which were reduced to a binary class schema focused on agricultural areas vs. other areas. The present Sub-Chapter will provide an overview on **change detection**, the **algorithms employed** and the **statistical comparison** between the epochs analysed in this Thesis.

**Change detection**: The collection of classified images from a given area from different dates (epochs) can give valuable information about the overall behaviour of the area proportion of the classes. However, as they result from statistical inference they provide no pixel-to-pixel change information. For instance, the area proportion for a given class can be the same for two consecutive epochs, but that does not allow one to conclude that no change has happened, as it can be that both changes into the class and from the class have occurred simultaneously, with no net change in proportion.

Change detection aims to quantify change either from a spectral perspective or from a pixel class perspective, it explains how the pixels are changing along different epochs. The general approach taken was the latter, defined as a *post-classification* comparison. As summarised by [99], this involves independently produced spectral classification results from each step of the time interval of interest, followed by a pixel-by-pixel comparison to detect changes in LULC type. By adequately coding the classification results, a complete matrix of change—from one class to another class— is obtained and change classes can be defined. LULC change detection is an overarching method to analyse uniform conversions impacting landscapes within protected zones [100].

When undertaking an analysis of LULC changes, it is recommended to prioritise the utilisation of the same product derived from consistent data sources and data processing methodologies [83]. In the present case, due to the utilisation of multiple distinct products and the lack of additional validation data or other ancillary data, change analysis was constrained to three distinct time steps for the change detection based on the dataset's spatial resolution. Namely, Landsat MSS at 60 m, Landsat TM, ETM+ and OLI at 30 m and Sentinel-2 MSI at 10 m spatial resolution.

Disturbance was monitored by focusing on class change, particularly from and to agricultural areas. Depending on the number of epochs available, change detection was performed on a bi-temporal basis or a multitemporal one, formerly addressed as a change trajectory.

**Bi-temporal** binary analysis is a direct comparison between two consecutive epochs, a pixel-to-pixel comparison that leads to an explicit change matrix, similar in composition to the error matrix. Instead of sample pixel counts it includes population counts and instead of Predicted class vs. True class, it is labelled as From class vs. To class. This configuration presents all 4 possible changes, given classes A and B, a change matrix would yield AA, AB, BB, and BA, as possible outcomes —see Figure 2.36 Left—. Where the notation AB would be considered from class A to class B conversion. Alternatively, depending on the focus, AB notation could be labelled as class B gain or class A loss.

**Temporal trajectory** binary analysis is a pixel-to-pixel comparison between multiple consecutive epochs, this leads to a $2^E$ number of outcomes or trajectories, where $E$ is the number of epochs. Pixel class trajectories for a 4-epoch analysis, yielding 16 possible outcomes, can be further summarised on groups based on the stability of the class behaviour. Following the binary A call and B class example, a temporal trajectory such as AAAA or BBBB can be considered stable class A or stable class B, respectively. A trajectory with only one clear change, such as AABB—see Figure 2.36 Right—, could be considered as a stepped change towards class B, therefore could be labelled as class A stepped loss or class B stepped gain, depending on the focus class. An example of labelling trajectories from stability maps that this Thesis adapted can be found in [101].



Figure 2.36: Change detection bi-temporal (Left) and temporal trajectory (Right) examples.

**Statistical testing**: Considering all the variability involved in the data pre-processing and the variability of the classification model of each epoch, it is of paramount importance to establish whether the difference between two different epochs is statistically significant. This is because even if the data is obtained by the same sensor, or intercalibrated between sensors of similar characteristics, the difference in spectral signatures that grouped the pixels in different classes could be explained by signal noise due to different illumination, atmospheric conditions, etc. Even slight mismatches in image registering can lead to false-positive change detection [99].

To reduce possible sources of variability, or better, to ensure similar or comparable variabilities, strategies like using same season acquisitions, using the same pre-processing pipeline, using the same sample data (when consistent within the time step analysed), using the same classification model with a unique class scheme, and even using the model with different random initialisations, can be implemented. These practices can contribute to the comparability of the two classifications in term of class area or class proportion beyond the merely statistical aspect.

One of the main elements to determine are the variables to compare and their distributions, in order to apply a suitable statistical test. For this case, the estimated class areas are the variables, but their distributions are unknown. In most cases, and from the statistical inference, they are assumed to be normally distributed. However, the class areas being a result of a classification model which may include mixed pixels that are assigned to classes in different ways, the normality assumption must be tested. The statistical tests used for assessing normal distribution of a variable are consolidated in Table 2.4.

Table 2.4: List of statistical tests for normality

| Test | Hypothesis | Type | Python library |
|---|---|---|---|
| Shapiro-Wilk [102] | $H_0$= The sample comes from a normal distribution<br>$H_A$=The sample is not coming from a normal distribution | Parametric | scipy.stats. shapiro() |
| Anderson-Darling [103] | $H_0$= The sample comes from a given (normal) distribution<br>$H_A$=The sample is not coming from a given (Normal) distribution | Goodness-of-fit Non-Parametric | scipy.stats. anderson() |
| Kolmogorov-Smirnoff [104] | $H_0$= The sample cumulative distribution function $Fs(x)$ is equal to a theoretical $Ft(x)$ (Normal) for all x from $-\infty$ to $\infty$<br>$H_A$=The sample cumulative distribution function $Fs(x)$ is not equal to a theoretical $Ft(x)$ (Normal) for at least one x | Non-Parametric | scipy.stats. kstest() |
| D'Agostino $K^2$ [105], [106] | $H_0$= The sample comes from a normal distribution<br>$H_A$=The sample is not coming from a normal distribution | Goodness-of-fit Parametric | scipy.stats. normaltest() |
| Jarque-Bera [107] | $H_0$= The sample comes from a normal distribution<br>$H_A$=The sample is not coming from a normal distribution | Parametric | scipy.stats. jarque_bera() |
| Lilliefors [108] | $H_0$= The sample cumulative distribution function $Fs(x)$ is equal to a theoretical $Ft(x)$ (Normal) with sample estimated mean and variance<br>$H_A$=The sample cumulative distribution function $Fs(x)$ is not equal to a theoretical $Ft(x)$ (Normal) with sample estimated mean and variance | Non-Parametric | statsmodels.stats. diagnostic.lilliefors() |

Once determined whether a variable is normally distributed or not, a test for that variable at different epochs is conducted. First, a test to establish whether there is a variable mean with a statistically significant difference from all the others, then a pair-wise test to find which variable pairs have statistically significant difference. Table 2.5 presents the respective tests both when normality assumptions are met and when they are not.

Table 2.5: List of statistical tests for variable at different epochs

| Test | Hypothesis | Type | Data Assumptions | Python library |
|------|-----------|------|------------------|----------------|
| Kruskal-Wallis [109] | $H_0$= The data samples distributions and medians are equal $H_A$=At least one of the data samples distributions and medians are not equal | Non-Parametric All Data | Independent Ordinal & Continuous Non-Normal Homogeneity: Similar Distribution | scipy.stats. kruskal() |
| Dunn [110] | $H_0$= There is no difference in the ranks of the two Data Samples being compared $H_A$=There is a difference in the ranks of the two Data Samples being compared | Non-Parametric Paired Data | Independent Ordinal & Continuous Non-Normal Equal Variance | scikit_posthocs. posthoc_dunn() |
| ANOVA [111] | $H_0$= The data sample means are equal $H_A$=At least one of the data sample means is different from the others | Parametric All Data | Independent Normal Homogeneity: Similar Variance | scipy.stats. f_oneway() |
| Tukey-HSD [112] | $H_0$= The data sample means compared are equal $H_A$= The data sample means compared are not equal | Parametric Paired Data | Independent & Equal Size Normal Homogeneity: Similar Variance | statsmodels.stats. multicomp. pairwise_tukeyhsd() |

## 2.7. Indices & Applications

An index is a sign or measure of something, usually compared—normalised—with respect to a reference value. The accuracy metrics from Sub-Chapter 2.5 can be considered indices of accuracy, as they are normalised as percentages with respect to the total—also interpretable as probabilities—. During the development of this Thesis, a spatial-temporal index was formulated for the agricultural areas within the Páramo Complexes. Indices are formulated for different applications and this final Sub-Chapter will provide an overview of this matter.

EO research has developed a series of spectral indices for different land cover monitoring or domain applications, resulting from the non-linear combination of spectral bands. In the realm of classification by ML algorithms, linear combination of bands does not add any additional information, although they can be useful for visualisation purposes. Non-linear combinations on the contrary, have proven to be effective when wanting to reduce unwanted or confounding effects when studying a certain land cover or phenomenon, leading to the development of domain-specific indices.

Recent efforts to consolidate a catalogue account for over 200 different spectral indices, most of them related to vegetation (127), followed by urban (19), burn/post-fire analysis (19), among others [113]. One of the most recognised and applied index for vegetation is the Normalised Difference Vegetation Index (NDVI) [114], computed as the normalised difference between NIR reflectance and red reflectance $NDVI = (\rho_{NIR} - \rho_{red}) - (\rho_{NIR} + \rho_{red})$, used to assess vegetation greenness in space and time. Figure 2.37 presents an open-source index catalogue and a multitemporal spectral index cube for Sentinel-2 data where different indices variability over time can be seen.

Other indices might be assembled beyond spectral response and focused on spatial and/or temporal characteristics, sometimes combining remotely sensed data with spatial data derived from authoritative sources (population, infrastructure, hydrology, etc.). Efforts combining this different spatial data and crafting an index for quantifying the human effects on natural ecosystems have been of special research interest.

Figure 2.37: Left: Awesome Spectral Indices catalogue (v 0.4.0) per application domain. Right: Spectral indices mini cubes from Sentinel-2 displaying a 2.56 km radius site between 2018-07-09 (front) and 2021-09-26 (back) [113, Figs 2, 6]

Global-scale efforts like the human disturbance index [115], or the human footprint [116], have tried to provide a scale in which to assess the pressure human activity exerts on the Earth's ecosystems. Following this line, national ecological indicators of this kind have been developed in Colombia, incorporating temporality and biophysical vulnerability of the ecosystems [117], leading to a Legacy-adjusted Human Footprint Index (LHFI), that analysed four decades of data resulting in identifying five main change hotspots along the country [118]. Similar attempts to the one described in this Thesis using classification, change detection and spatial indices have been performed in protected areas including Montane Forest and Páramos in Ecuador [119].

Regardless of the nature of the indices, their objective is to encompass a number on a relative scale to account for a given phenomenon. In that sense, the proposition of a novel index must be consistent with the phenomenon that it accounts for, both conceptually and mathematically. Most importantly, an index must be formulated to be informative, as they become powerful tools to condense large and/or complex amounts of information to a single number, that can be interpreted by a wide range of audiences.

# 3   Methodology

A thorough methodology was developed in order to perform an assessment of the level of disturbance of the Páramo ecosystem in the central Eastern Andes of Colombia from satellite Earth Observation (EO) free and open data—available since 1972 from the first Landsat mission—. The methodology was based on three main stages: satellite data was selected and pre-processed for Stage 1, classified and assessed for accuracy for Stage 2, and analysed for change detection for Stage 3, as shown in Figure 3.1.



Figure 3.1: Overall Stages methodological approach

The full-detail schematic of the methodological flowchart can be consulted on Appendix A.1. For practical purposes this flowchart was divided into the three stages and for each stage a full flowchart is presented in the corresponding Sub-Chapter. Assessment Areas are presented in Sub-chapter 3.1, Stage 1 corresponds to Sub-chapters 3.2 and 3.3, Stage 2 corresponds to Sub-Chapter 3.4, while Stage 3 corresponds to Sub-Chapter 3.5 and 3.6.

## 3.1.   Assessment Areas

For the scope of this Thesis, three Protected Areas (PAs) were selected for assessment: Páramo de Guerrero (GUE), Páramo de Rabanal y Río Bogotá (RRB), and Páramo Altiplano Cundiboyacense (ACB). The latter, is composed of small, fragmented areas along the high plateau of *Altiplano Cundiboyacense*—from where it gets its name— and has been sub-divided and grouped in areas belonging to the GUE geometry bounding box (ACB_G), the RRB geometry bounding box (ACB_R), and two other independent areas (ACB_A and ACB_B). Figure 3.2 presents the assessment PAs, and the grouping of the sub-divided ACB areas.



Figure 3.2: Assessment Protected Areas selected for the Thesis

The high plateau of *Altiplano Cundiboyacense* is a region above 2.500 m elevation in the Colombian Eastern Andes AOI with an average temperature of 13.5°C, with monthly oscillations under 1°C, daily variations from 0°C to 25°C, and annual precipitation ranging from 580-1000 mm [120]. Like all Andean Mountain ranges, average temperature decreases with altitude, while rainfall seasonality and micro-climate is highly correlated to the mountain range aspect (West or East facing)—as highlighted by the ecosystem changes noted in Figure 2.2—. Although the Colombian Andean region presents largely a bi-modal rain regime [121]—with dry season between December-February and June-August—, changes due to the El Niño-South Oscillation

and local dry and rainy season regionality can occur, altering the seasonality and the amount of precipitation in the *Altiplano Cundiboyacense* [122].

To provide some historic context, the region sheltered humans over 13.000 years ago, with vestiges of agricultural activities dating back over 5.000 years, with corn and quinoa seeds as the oldest crops registered [123]. By the time of the arrival of Spanish colonisers, it sheltered several heterogenous indigenous populations from the *macrochibcha* linguistic family, whose agricultural practices where diversified spanning the different altitudinal micro-climates, with Spanish chroniclers highlighting the sacred status these indigenous cultures professed for forests, lakes, rivers and mountain-tops [6].

Colonisation led to a transformation of the agricultural practices: while landowners maintained the indigenous practice of farming across micro-climates, they cleared the forests for wood and expanded the pastures as the main land cover and the land use for livestock grazing in the wetlands, engaging in hydraulic modifications through the construction of trenches and ditches, draining wetlands and diverting the river courses [124]. Main indigenous settlements declined or were replaced by nuclear colonial settlements like the towns of Tunja and Bogotá [125], the current capital of Colombia.

Large post-colonial states progressively split among successive generations while significant plots of un-cultivated land were awarded for military victories to civil-war generals during the republic's political turmoil. These states were in turn sub-divided or sub-let as parcels for peasants to cultivate, which resulted in further expansion of the agricultural area towards increasing altitudes [6]. Population growth alongside economic development and an increased demand for produce, coupled with the increased urbanisation and soil degradation in the lower valleys have pushed the agricultural limit upwards into the Páramo [5]. Currently, the *Altiplano Cundiboyacense* is home to over 9 million inhabitants, most of them living in urban areas like Bogotá, that accounts for 7,4 million people, according to the 2018 national census [126].

Just around 70 km north of Bogotá, the main GUE complex polygon PA is found roughly centred in a small natural lake named Laguna Verde. The southern part of this area is what local communities mostly address as the proper *Páramo de Guerrero*, while the northern part is also referred to as *Páramo de Guargua* or *Páramo de Laguna Verde*. The other PAs, including the ACB subdivisions, are smaller mountain tops or escarpment formations where the altitudinal range was considered to harbour Páramo ecosystem according to the Colombian Ministry of Environment and Sustainable Development (MADS) delimitation. RRB complex, roughly centred about 90 km northeast of Bogotá, is composed of three distinct polygon PAs from north to south: *Páramo de Rabanal*, *Páramo de Guacheneque* or *Rio Bogotá*—Source of Bogota River—, and *Páramo Los Cristales*. All of the assessment PAs size and location can be viewed in Table 3.3, and the names of the PAs in italics will refer to the specific polygons of the Páramo Complexes.

Table 3.1: Size and Location of the assessment protected areas.

| Páramo Complex | Size [Ha] | Location |
|---|---|---|
| GUE | 43.418,09 (Total)<br>37.629,60 (*Páramo de Guerrero/Guargua/Laguna Verde*) | 74,00°W - 5,21°N |
| RRB | 24.755,95 (Total)<br>11.202,44 (*Páramo de Rabanal*)<br>10.030,03 (*Páramo de Guacheneque/Rio Bogotá*)<br>3.523,48 (*Páramo de Cristales*) | 73,60°W - 5,40°N<br>73,56°W - 5,18°N<br>73,48°W - 5,15°N |
| ACB | 5.823,19 (Total)<br>4.228,70 (Assessment Total)<br>1.861,55 (ACB_A)<br>1.468,76 (ACB_B)<br>480,62 (ACB_G)<br>417,77 (ACB_R) | 73,72°W - 5,19°N<br>73,42°W - 5,52°N<br>73,85°W - 5,10°N<br>73,55°W - 5,34°N |

Both RRB and GUE complexes share common anthropogenic pressures at different levels, while ACB complex—being a collection of small areas just recently protected—lacks historical records or research about its anthropogenic pressures. As presented in Sub-Chapter 2.1, one of the main pressures affecting the Páramo ecosystem is the expansion of agricultural land. For GUE and RRB, the alternation of high-yield agrochemical-intense potato crops and dual-purpose —meat and milk— cattle grazing has been the most economically favourable option for land use, with an increasing level of mechanised ploughing and commercial-type cultivation beyond the subsistence farming [127]. The introduction of non-native grass to meet cattle grazing needs [128], and a very low prevalence of natural fertilising and sustainable cultivation practices [129] complete the landscape of the agricultural practices within these complexes. Figure 3.3 presents a collection of photographs of agricultural practices in RRB.



*Photos: Robineau O, Châtelet M*

Figure 3.3: North RRB *Páramo de Rabanal* photograph mosaic. a. small-scale farmers planting potatoes; b. spraying a potato field; c. watering a calf; d. entrepreneur planting potatoes; e. herbaceous Páramo [127, Fig. 1]

Another relevant anthropogenic pressure for RRB and GUE complexes is mining. Although being a prohibited activity by law within the delimited Páramo Complexes from 2011 [20], confirmed in 2015 [130] and later ratified by the Colombian Constitutional Court in 2016 [131], mining—both legal and illegal—has a long tradition in *Páramo de Guerrero* and *Páramo de Rabanal* areas[129], [132] and it is still conducted surrounding the government-declared boundaries. Aside from some small-scale surface mining of clay, metal minerals, sand and gravel, the most relevant extractive activity is the sub-surface mining of coal from different beds, mainly *Guaduas* Formation, but also from the *Cacho* Formation and the *Bogotá* Formation [133], [134], [135]. These mining operations are usually coupled with near-site transformation of this coal in coke fuel by anoxic pyrolysis on furnaces to supply national and international markets [129], [134].

Both mining and coke production have detrimental effects on water sources and the soil by acidification, sediment and metal pollution [136], [137], but their effects are difficult to capture from a land cover perspective. Nonetheless, recent research suggests that vegetation covered in black soot from nearby (>2 km) coke furnaces north of *Páramo de Rabanal* exhibit a lower NIR spectral response [28], opening the research to correlated spectral changes of vegetation from this kind of pollution. Figure 3.4 presents a collection of photographs of mining practices in RRB.



Figure 3.4: Mining photograph mosaic of RRB a. Coal mine, b. sedimentation tank for mine water and coal gathering, c. and d. Coke oven furnaces, emissions and coke gathering. [129]

Inter-related with the mining and agriculture activities in the assessment areas is the introduction of exotic vegetation species, both as large-scale plantations and as scattered individuals. Introduction of these exotic species has followed the depletion of native forests supply for wood, erosion control and lower valleys desiccation by means of species of Pines, Acacias and Eucalyptus, among others [138], [139]. In addition to these trees, exotic shrub vegetation such as the Gorse and exotic grass like the Kikuyu represent a risk for native species as they dominate the landscape and hinder the prevalence and composition of the native vegetation [140]. Figure 3.5 presents a pair of photographs of exotic and invasive vegetation species in GUE.



Figure 3.5: Left: Pine (*Pinus sp.*) plantations at the Neusa reservoir along and atop West *Páramo de Guerrero* range (*island patch 73,965°W - 5,147°N, top patch 73,979°W - 5,181°N*). Right: Gorse (*Ulex europaeus*) proliferation near the Neusa reservoir in the Tausa municipality. Taken from [138].

While often disregarded as a relevant anthropogenic disturbance, the flooding of large areas to create reservoirs has negative environmental impacts such as hydrological cycle disruption and methane emissions product of anaerobic organic matter degradation in the sediments [141]. Along all the *Altiplano Cundiboyacense* there are several large water bodies, many of which are reservoirs for flood control and water supply, held back by dam structures like the *Neusa* (*73,95°W - 5,16°N*), *Tominé* (*73,83°W - 4,97°N*), and *El Hato* (*73,908°W - 5,295°N*), all falling within the GUE geometry bounding box. Within the assessment PAs however, only RRB complex has artificial reservoirs within its boundaries at the *Páramo de Rabanal*, with the *Gachaneca I* (*73,564°W - 5,436°N*), *Gachaneca II* (*73,548°W - 5,450°N*), and *Teatinos* (*73,548°W - 5,450°N*) reservoirs.

Finally, the remaining elements of the assessment's Páramo Complexes to be considered are the rural communities that inhabit within government-declared boundaries. Having introduced the historic configuration of its inhabitants at the beginning of this Sub-Chapter, the most recent national census can provide an updated insight, as it incorporated the Páramo inhabitants as a focus population. Table 3.2 summarises some of the most relevant information about the census in the three assessment complexes.

Table 3.2: Páramo Complexes 2018 Census [142].

| Páramo Census | | GUE | RRB | ACB |
|---|---|---|---|---|
| **Residencies** | Total Residences | 1691 | 222 | 641 |
| | (Industry/Commerce/Service/ Agriculture-Agroindustry-Forestry) | (0 / 20 / 0 / 32) | (0 / 0 / 0 / 5) | (1 / 10 / 5 / 11) |
| | Total Mixed Use | 52 | 5 | 27 |
| | Total Housing Units | 1743 | 227 | 668 |
| | **Total Housing Units/ kHa\*** | **40,14** | **9,17** | **114,71** |
| **Non-Residential Plots** | Industry | 0 | 0 | 1 |
| | Commerce | 1 | 0 | 1 |
| | Services | 12 | 4 | 25 |
| | Agriculture-Agroindustry-Forestry | 1941 (53,4%) | 245 (18,6%) | 685 (54,2%) |
| | Institutional | 18 | 0 | 3 |
| | Plot (non-built) | 1411 | 749 | 533 |
| | Park - Green Area | 1 | 0 | 0 |
| | Minero-energetic | 14 (0,38%) | 7 (0,53%) | 4 (0,32%) |
| | Conservation - Env. Protection | 219 (6,02%) | 299 (22,65%) | 7 (0,55%) |
| | Under Construction | 19 | 14 | 6 |
| | No Data | 2 | 2 | 0 |
| | **Total Plots** | **3638** | **1320** | **1265** |
| | **Total Plots/ kHa\*** | **83,79** | **53,32** | **217,23** |
| **Homes (Inhabited Residences)** | **Total Homes** | **968** | **104** | **385** |
| | Electricity Service | 96,4% | 93,8% | 94,2% |
| | Aqueduct Service | 39,1% | 39,2% | 76,9% |
| | Gas Service | 0,6% | 3,1% | 0,5% |
| | Sewage Service | 2,5% | 4,3% | 1,0% |
| | Rubbish Collection Service | 3,4% | 6,2% | 0,5% |
| **Population Characterisation** | **Total Inhabitants** | **3229** | **368** | **1386** |
| | **Total Inhabitants/ kHa\*** | **74,37** | **14,87** | **238,01** |
| | Literacy Rate | 93,3% | 91,8% | 92,9% |
| | Currently in Education (All levels) | 25,0% | 25,1% | 28,2% |
| | Pre-Schooling | 2,7% | 2,1% | 3,4% |
| | Primary Schooling | 62,4% | 54,7% | 60,0% |
| | Middle Schooling | 16,0% | 13,3% | 16,1% |
| | High Schooling | 10,8% | 9,4% | 12,8% |
| | Technical - College Schooling | 0,9% | 0,9% | 1,6% |
| | University Schooling | 0,8% | 2,4% | 1,6% |
| | Post-graduate Schooling | 0,1% | 0,6% | 0,3% |
| | MSc. Schooling | 0,0% | 0,0% | 0,0% |
| | PhD. Schooling | 0,0% | 0,6% | 0,0% |
| | No Education Received | 5,6% | 3,9% | 3,9% |
| | Born in same Municipality | 60,5% | 51,4% | 72,9% |
| | Born in another Municipality | 38,8% | 37,8% | 27,1% |
| | Born in another Country | 0,1% | 0,0% | 0,0% |
| | Resided in Municipality 5 years ago | 96,1% | 92,7% | 97,7% |
| | Resided in Municipality 1 year ago | 92,5% | 77,8% | 88,9% |

*Calculated by dividing over the total Páramo Complex area reported in Table 3.1

It is interesting to note that the area-normalised values allow for a more objective comparison in terms of inhabitants, housing and plots: ACB has a much higher density on all these metrics, followed by GUE and finally RRB. In terms of non-residential plot distribution, although not area-specific, one can note the prevalence of Agriculture-

Agroindustry-Forestry plots on GUE and ACB, around half the total, while RRB is around one fifth. On the other hand, plots distributions for conservation presents a noticeable difference between RRB having around one fifth, while GUE has little over one twentieth, and half a percentual proportion for ACB. The institutional plots are mostly rural schools, which denote the long-standing communities within the GUE area. Service coverage for gas, sewage and rubbish collection are on a one-digit percentual range and below which highlights the struggle for local sanitation and environmental pollution of all these communities. Differences in terms of aqueduct service coverage hint at a higher aggregation of ACB housing, as opposed to more dispersed housing on the other complexes that results in more individual solutions for water procurance. Regarding education, the highest prevalence of schooling is Primary with roughly three-fifths of the population, followed by Middle and High schooling with about one-tenth each. Finally, it is important to note that across complexes most people were born in the same municipality and the vast majority resided there 5 years before the census, which allows us to infer the rooting of the inhabitants to these Páramo areas.

As highlighted both in Sub-Chapter 2.1 about Páramo Ecosystems and in the present Sub-Chapter for the chosen assessment Páramo Complexes, the analysis of the disturbance of Páramo ecosystems spans over different disciplines. As a narrow perspective focused on land cover, this Thesis aims to use some of the state-of-the-art methods to quantify the changes in Páramo extension under the assumption that it represents a direct measure of Páramo quality. One could argue that it is a debatable or biased assumption, but I would argue that just like the struggle for ecologists to define the different ecosystem boundaries, some convention must be adopted, and conventions are part of the human bias. In order to assign data to a class in a hard classification the very classes selected, and their labels, introduce some level of bias. In awareness of this, a substantial effort has been made to be transparent about the assumptions, data processing and modelling decisions made within the methodology, as well as the discussion of the results in the context of the assessment areas described above.

## 3.2. Satellite Data Selection

The work conducted in Stage 1 aimed to retrieve the best available data from the immense Landsat Collection 2 and Sentinel-2 datasets, identifying the imagery for the greatest number of years (Referred to as Epochs, to incorporate the pairing of year and sensor) for the entire Area of Interest (AOI) areas above 2.500 m (~22.000 $Km^2$), and perform all pre-processing necessary to obtain a virtually cloud-free mosaic for further processing. This Sub-Chapter explains the Data Selection procedure, which is the first part of Stage 1, encompassing all the steps before pre-processing, as shown in Figure 2.1.
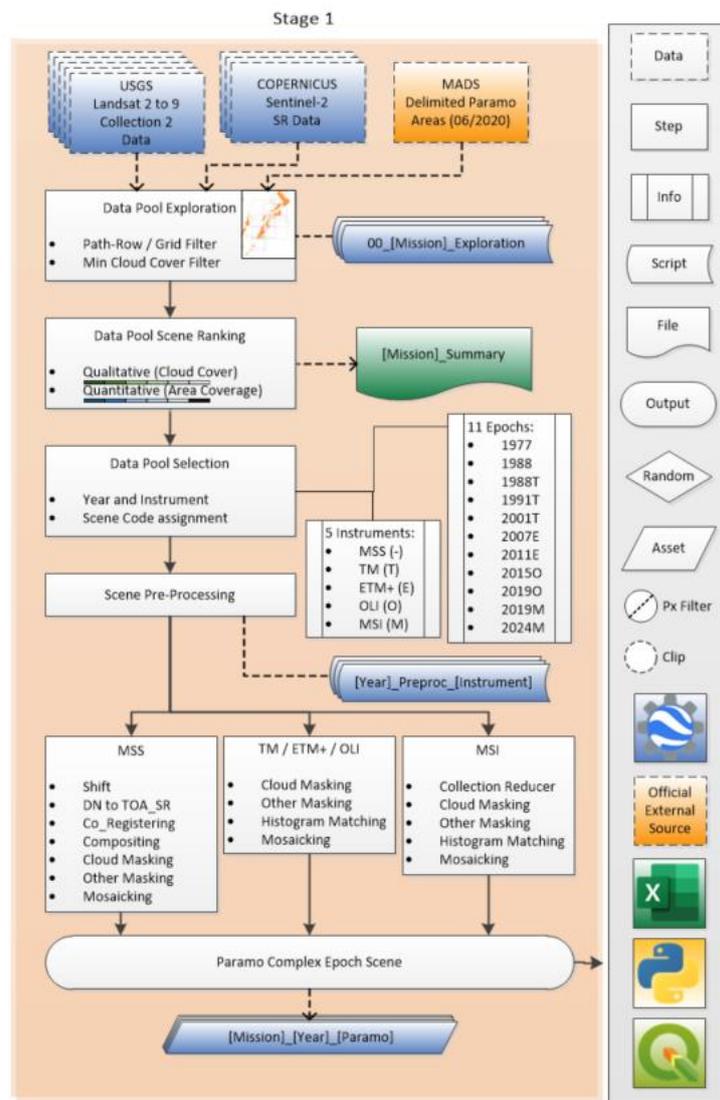


Figure 3.6: Stage 1 methodological flowchart.

As introduced in Sub-Chapter 2.3, both Landsat Archive and Sentinel-2 data products are made available at different levels. This Sub-Chapter will review first the selection of the Landsat data and second the Sentinel-2 data, and the epochs chosen for pre-processing.

Landsat Collection 2 raw (Level 0 with Level 1 metadata) Tier 1 and Tier 2 data was accessed through the Google Earth Engine (GEE) catalogue when no further level data was available, which was the case for the Landsat 1-5 MSS data. For the other Landsat sensors, surface reflectance (SR) data (Level 2) are available in the GEE catalogue as a copy of the USGS Collection 2, Level 2 archive. It is important to note that Landsat 4 TM, 5 TM, and 7 ETM+ SR data are generated using the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) algorithm [143], while Landsat 8 and 9 OLI SR data are generated using the Land Surface Reflectance Code (LaSRC) algorithm [144]. While both the LEDAPS and LaSRC algorithms produce similar SR products, the inputs and methods to do so differ between them[12].

An initial GEE script '*00_Landsat_Exploration*' was developed to search the corresponding Scenes on the Landsat Archive by means of the FAO GAUL 500m Simplified: Global Administrative Units[13] for the Cundinamarca and Boyacá regions. Based on three levels of cloud cover (15%, 20%, 25%)—Available in each scene product metadata— and the dates for each Landsat Mission, the WRS reference grid 7 and 8 paths and the 55, 56, 57 and 58 rows were identified as the scenes covering the AOI. The results were sorted to identify the dates of the least cloudy scenes, which were between December and March, corresponding to one of the AOI dry seasons.

Separate GEE scripts '*00_Landsat_1to5_Exploration*', '*00_Landsat_5_TM_Exploration*' and '*00_Landsat_7to9_Exploration*' were developed to retrieve the best imagery available using the paths and rows previously identified as filters, as well as a wider date filter from October to March, to account for possible regional variation of the dry season. The vector map of the officially declared Páramo Complex areas was retrieved from the environmental information system from Colombia (*SIAC*) [145] and uploaded as GEE assets per Complex for visualisation. The cloud-cover sorted top scenes from the datasets were registered on a Data Pool, curated for the best imagery for different Páramo Complexes along the AOI.

Due to some scenes having different levels of cloud cover over the Páramo Complexes, a ranking system for the Data Pool was designed to incorporate a colour-coded objective and subjective variable to define the study epochs. The objective variable was the percentage of area the scene covered for the given Complex in 6 ranges (*0%, >0-25%, >25-50%, >50-75%, >75-99,9% and 100%*) and the subjective variable was a qualitative rating with 6 cloud cover classes from '*Mostly Covered*' to '*Very Little to None*'. The area percentage of the PAs within the scene was calculated by the script '*Scene_Paramo_Areas*'. The colour coding aided in the decision of whether or not to

---

[12]See Differences in Surface Reflectance Processing Algorithms at:
https://www.usgs.gov/landsat-missions/landsat-collection-2-surface-reflectance
[13] Source of Administrative boundaries: The Global Administrative Unit Layers (GAUL) dataset, implemented by FAO within the CountrySTAT and Agricultural Market Information System (AMIS) projects

incorporate certain scenes with respect to both variables per Complex, considering the trade-off of certain dates being more suitable for some of them and less for others. The Landsat Data Pool and the ranking system can be consulted in Appendix A.3 sorted by date, while the Landsat datasets selected for the Data Pool in the mentioned GEE scripts are summarised in Table 3.3. It should be noted that no Landsat 9 datasets were considered suitable for the analysis.

Table 3.3: Landsat datasets accessed for ranking

| GEE Directory | Description | Availability |
|---|---|---|
| LANDSAT/LM02/C02/T2 [146] | Landsat 2 MSS Collection 2 Tier 2 DN values, representing scaled, calibrated at-sensor radiance. | 1975-01-24T02:12:31 to 1982-03-12T10:52:37 |
| LANDSAT/LM04/C02/T2 [147] | Landsat 4 MSS Collection 2 Tier 2 DN values, representing scaled, calibrated at-sensor radiance. | 11982-08-06T20:50:11 to 1992-10-15T05:31:24 |
| LANDSAT/LT04/C02/T1_L2 [148] | Landsat 4 TM Collection 2 Level 2 Tier 1 SR values, atmospherically corrected surface reflectance and land surface temperature derived from the data produced by the sensor, created with the LEDAPS algorithm (version 3.4.0). | 1982-08-22T14:19:55 to 1993-06-24T14:26:23 |
| LANDSAT/LT05/C02/T1_L2 [149] | Landsat 5 TM Collection 2 Level 2 Tier 1 SR values, atmospherically corrected surface reflectance and land surface temperature derived from the data produced by the sensor, created with the LEDAPS algorithm (version 3.4.0). | 1984-03-05T15:38:45 to 2012-05-05T17:54:06 |
| LANDSAT/LE07/C02/T1_L2 [150] | Landsat 7 ETM+ Collection 2 Level 2 Tier 1 SR values, atmospherically corrected surface reflectance and land surface temperature derived from the data produced by the sensor, created with the LEDAPS algorithm (version 3.4.0). | 1999-05-28T01:02:17 onwards |
| LANDSAT/LC08/C02/T1_L2 [151] | Landsat 7 ETM+ Collection 2 Level 2 Tier 1 SR values, atmospherically corrected surface reflectance and land surface temperature derived from the data produced by the sensor, created with the LEDAPS algorithm (version 3.4.0). | 2013-03-18T15:58:14 onwards |

According to the Appendix A.3 ranking table, the following epochs were defined for the Landsat datasets:

- **1977**: Landsat 2 MSS Tier 2 raw scene collection

- **1988**: Landsat 4 MSS Tier 2 raw scene collection

- **1988T**: Landsat 4 TM Tier 1 Level 2 scene collection

- **1991T**: Landsat 5 TM Tier 1 Level 2 scene collection

- **2001T**: Landsat 5 TM and 7 ETM+ Tier 1 Level 2 scene collection

- **2007E**: Landsat 7 ETM+ Tier 1 Level 2 scene collection

- **2011E**: Landsat 7 ETM+ Tier 1 Level 2 scene collection

- **2015O**: Landsat 8 OLI Tier 1 Level 2 scene collection

- **2019O**: Landsat 8 OLI Tier 1 Level 2 scene collection

For the Sentinel-2 data, only Level 2A data was explored by means of the '*00_Sentinel2_Exploration*' GEE script. Scenes were sorted by the cloudy pixel percentage product of the quality band QA60, and filtered by date from October to March, to account for possible regional variation of the dry season. The scenes were identified belonging to S2 Grid 18N (WJ-WK-WM-XK-XL-XM-YL-YM-YN) and the best imagery per date was selected, considering the overlap with the corresponding Landsat path. Figure 3.7 provides a visual comparison on the Landsat 8 and Sentinel-2 corresponding tiles for the AOI. It should be noted that the previous Landsat 1-7 missions' grid is omitted for clarity, but the overall path and row coverage is the same, with differences in terms of their coverage with respect to Páramo Complexes located at the path intersection.



Figure 3.7: AOI and corresponding Landsat 8 & Sentinel-2 grid tiles.

A similar approach to the Landsat ranking was taken, considering the subjective cloud cover after masking the QA60 pixels flagged as clouds per Páramo Complex. This time with 6 cloud cover classes from '*Mostly Covered*' to '*Very Little to None*' and a black colour code for no coverage. Best imagery per date, per equivalent Landsat Path (7 or 8) were classified A for the best, B for the second best and C for the third best to determine the epochs. Visualisation was done on both GEE and on Sentinel Hub Platform [152] to rank each date set of tiles. The Landsat Data Pool and the ranking

system can be consulted in Appendix A.4, while the Sentinel-2 datasets selected for the Data Pool on the mentioned GEE scripts are summarised in Table 3.4.

Table 3.4: Sentinel-2 datasets accessed for ranking

| GEE Directory | Description | Availability |
|---|---|---|
| COPERNICUS/ S2_SR_HARMONIZED [153] | Sentinel-2 MSI Level2A Landsat-Harmonized Surface Reflectance. Downloaded from SciHub and processed with sen2cor. ESA did not produce L2 data for all L1 assets, and earlier L2 coverage is not global. QA60 is a bitmask band that contained rasterized cloud mask polygons until February 2022, when these polygons stopped being produced. Starting in February 2024, legacy-consistent QA60 bands are constructed from the MSK_CLASSI cloud classification bands | 2017-03-28T00:00:00 onwards |
| COPERNICUS/ S2_CLOUD_PROBABILITY [62] | The S2 cloud probability is created with the sentinel2-cloud-detector library (using LightGBM). All bands are upsampled using bilinear interpolation to 10m resolution before the gradient boost base algorithm is applied. The resulting 0..1 floating point probability is scaled to 0..100 and stored as a UINT8. Areas missing any or all of the bands are masked out. Higher values are more likely to be clouds or highly reflective surfaces (e.g. roof tops or snow). | 2015-06-27T00:00:00 onwards |

According to the Appendix A.4 ranking table, the following epochs were defined for the Sentinel-2 datasets:

- **2019M**: Sentinel-2 MSI Level 2A scene collection

- **2024M**: Sentinel-2 MSI Level 2A scene collection

For each epoch, the scene collection included imagery acquired from December —from the previous year— to March of the same year. A total of eleven epochs were defined, having the years 1988 and 2019 as *pivot* years, that is, years were there is imagery from different sensors. For the year 1988 the date of image acquisition is the same, as both instruments—MSS and TM—were on board the Landsat 4 satellite. For the year 2019, the dates of the Sentinel-2 collection were from similar dates with respect to the Landsat 8, corresponding to the least cloudy weeks of the dry season.

Each scene selected was assigned a code according to the following scheme: [*Mission_Sensor_Path_row_year_letter-grading*] for Landsat data and [*Mission_Level_year_equivalent-landsat-path_letter-grading*] for Sentinel-2 data. Note that same path images from different rows for a given year usually correspond to same date acquisitions. Some code examples are provided for better interpretation:

- **L4_8_56_1988** corresponds to Landsat 4, path 8, row 56 of 1988, the absence of instrument letter follows the equivalent epoch notation, so MSS data have no letter assigned and the absence of the final letter denotes there is no second-best image for this tile. On the contrary, **L4T_8_56_1988** corresponds to Landsat 4 TM

path 8 row 56 of 1988 and again, no final letter denotes there is no second-best image for the tile. Note that the year is the epoch year to which the image is assigned, meaning that the image could have been acquired in December 1987, as is the case of the image **L4T_8_56_1988** (Refer to Appendix A.3).

- **L7E_7_57_2007A** corresponds to Landsat 7, ETM+ path 7, row 57 from 2007, and the letter A denotes it is the best image for this tile, while **L7E_7_57_2007B** would correspond to the second-best image for the tile.

- **S2A_2019_8A** corresponds to Sentinel-2 Level 2A median reducer of all tiles equivalent to Landsat path 8—no actual composition is made as the images are part of the same satellite swath acquisition and are only split on tiles for their size—, and the letter A denotes it is the best set of images for this path, while **S2A_2019_8B** would correspond to the second-best set of images for the path. Note that the year is the epoch year to which the image is assigned, meaning that the image could have been acquired in December 2018, as is the case of the image **S2A_2019_8A** (Refer to Appendix A.4).

## 3.3.  Satellite Data Pre-processing

The work conducted in the first part of Stage 1 retrieved the best available data from the immense Landsat Collection 2 and Sentinel-2 datasets, identifying the imagery for 11 Epochs (1977, 1988, 1988T, 1991T, 2001T, 2007E, 2011E, 2015O, 2019O, 2019M and 2024M) for the entire Area of Interest (AOI) areas above 2.500 m (~22.000 Km$^2$). This Sub-Chapter explains all of the necessary pre-processing required to obtain a virtually cloud-free epoch mosaic for the AOI from the selected images collections in the previous Sub-Chapter. Different Pre-processing procedures were followed for each sensor, although similar steps were adopted for the TM/ETM+/OLI datasets, as shown in Figure 2.1. TM/ETM+ data was combined for the 2001T epoch due to proximity of dates and quality of the ETM+ data, as it was acquired before the Scan Line Corrector (SLC-off) failure—occurred on May 31, 2003— which causes gaps in later acquisitions.

The overall approach to generate the AOI epoch mosaic was to reduce the use of image composites as much as possible. This was decided under the premise that no cloud removal is completely effective and being a cloud-prone area, compositing could lead to the aggregation of hazy pixels around areas where cloud cover was observed to be more prevalent. In other words, the trade-off of the pixels gains from the composite against the potential loss of pixel information by accumulating observations was considered negative. Although this meant that some epochs might have a significant amount of No Data values, the pixels obtained constitute a true observation, or in most cases a set of true observations from close dates. The approach taken was to use the best image for the given date and then use the second-best image to fill the gaps left by the cloud mask. Images from different dates corresponding to different paths were mosaicked in an order depending on the image quality, using the best image as reference and the other as a target for the colour correction.

A detailed pre-processing pipeline table for every sensor is provided in Appendix A.5, summarising the following overall step-by-step explanation. Each scene used is coded according to the coding scheme explained in the previous Sub-Chapter.

**MSS Pre-processing**

Some scenes presented significant geometric anomalies and were shifted according to the Geometric corrections section explained in Sub-Chapter 2.3, using control points from the edges of the *Tominé (73,83°W - 4,97°N)* reservoir and other geographic features located at the intersection of rows on Path 8, and using Equation (2.4).

Once aligned, according to Radiometric corrections sections of  Sub-Chapter 2.3, the DN values of the bands were converted to TOA reflectance by means of the corresponding metadata values and using Equations (2.1) (2.1)and (2.2) for each band. A bi-cubic co-registering of the shifted images was performed for them to match the non-shifted row, and a *minimum* TOA reflectance composited was applied for some scenes with hazy pixels, to reduce their prevalence in the scene overlap.

For the cloud masking, a threshold, buffer and cloud-shadow masking algorithm was developed following the description within the Data Masking section of Sub-Chapter 2.3. Reflectance thresholds adopted varied from 0,18 to 0,22, while buffer was set to 3 pixels for all scenes, with cloud shadow masking distances (along the corresponding azimuth shadow angle) ranging from 300 to 360 m. Faulty dark pixels were removed from the side edges of some scenes, by removing the furthest 3 pixels (~180 m) of the acquisitions.

An elevation mask was defined based on the MERIT Digital Elevation Model (DEM) [154] for elevations lower than 2.500 m[14] and was applied to all scenes, prior to the final mosaic composition, overlaying one scene over the other, with the best imagery on top. The resulting mosaic was assigned a code according to the following scheme: [*Mission_Sensor_year_reflectance-type*], resulting in the following:

- **L2_1977_TOA** corresponding to Landsat 2 MSS for the epoch 1977, in TOA reflectance. The absence of instrument letter follows the equivalent epoch notation. Mosaic was generated using the script '*1977_Preproc_MSS*'. Single date TOA Mosaic (07-01-1977) NIR band combination shown in Figure 3.8 Left.
- **L2_1988_TOA** corresponding to Landsat 2 MSS for the epoch 1988, in TOA reflectance. The absence of instrument letter follows the equivalent epoch notation. Mosaic was generated using the script '*1988_Preproc_MSS*'. Double date TOA Mosaic (22-03-1988, 16-04-1988) NIR band combination shown in Figure 3.8 Right.



Figure 3.8: MSS 1977 and 1988 AOI epoch mosaics generated.

---

[14] The final Elevation mask was obtained by sieving some isolate pixels outside the high plain, this process was performed on the GEE '*DEM_selection*' Script.

Due to the GEE 10-million-pixel restriction for processing, the final epoch mosaic was split by clipping it to the bounding boxes of the assessment Páramo Complexes. The Paramo name abbreviation was added to the resulting corresponding assets. Due to geometric distortions in the 1977 mosaic, further co-registering was performed for each Páramo Complex asset with its 1988 counterpart by means of the Mosaic that was generated using the script '*1977_Preproc_MSS_1988GeoCorr*'. Co-registering was performed by bi-cubic resampling, with offsets ranging 300 to 1440 m (maximum offset allowed when attempting to align the input images), and stiffness ranging from 1 to 6 (out of 10, with higher values representing a more rigid transformation). No atmospheric corrections were performed due to the lack of a reliable method to perform this operation to the mosaics.

**TM, ETM+, and OLI Pre-processing**

Landsat 4 and 5 TM pre-processing required cloud masking, performed by using the QA_PIXEL bitmask for bits 1 (Dilated Cloud), 3 (Cloud) and 4 (Cloud Shadow). QA_PIXEL is a band with pixel quality attributes generated form the CFMASK algorithm, a multi-pass algorithm that uses decision trees to prospectively label pixels in the scene; validates or discards those labels according to scene-wide statistics and creates a cloud shadow mask by iteratively estimating cloud heights and projecting them onto the ground[15].

Some scenes exhibited anomalous edge pixels, that were removed by an inwards edge buffer of 6 pixels (~360 m) or by manually drawn polygons for some cases. The MERIT DEM masking for elevations lower than 2.500 m was also applied to all scenes. Colour correction by histogram matching using a custom function—according to Data integration section of Sub-Chapter 2.3— was performed to second-best scenes and neighbouring rows from different paths, with targets and references denoted in Appendix A.5. The final mosaic was built by overlaying one scene over the other, with the best imagery on top. The resulting mosaic was assigned a code according to the following scheme [*Mission_Sensor_year_reflectance-type*], resulting in:

- **L4T_1988** corresponding to Landsat 4 TM mosaic for the year 1988. The absence of reflectance type denotes that the mosaic is in Surface Reflectance values. Mosaic was generated using the script '*1988_Preproc_TM*'. Three dates SR Mosaic (17-12-1987, 22-03-1988, 16-04-1988) RGB band combination shown in Figure 3.9 Left.
- **L5T_1991** corresponding to Landsat 5 TM mosaic for the year 1991. The absence of reflectance type denotes that the mosaic is in Surface Reflectance values. Mosaic was generated using the script '*1991_Preproc_TM*'. Double date SR Mosaic (16-03-1991, 23-03-1991) RGB band combination shown in Figure 3.9 Centre.
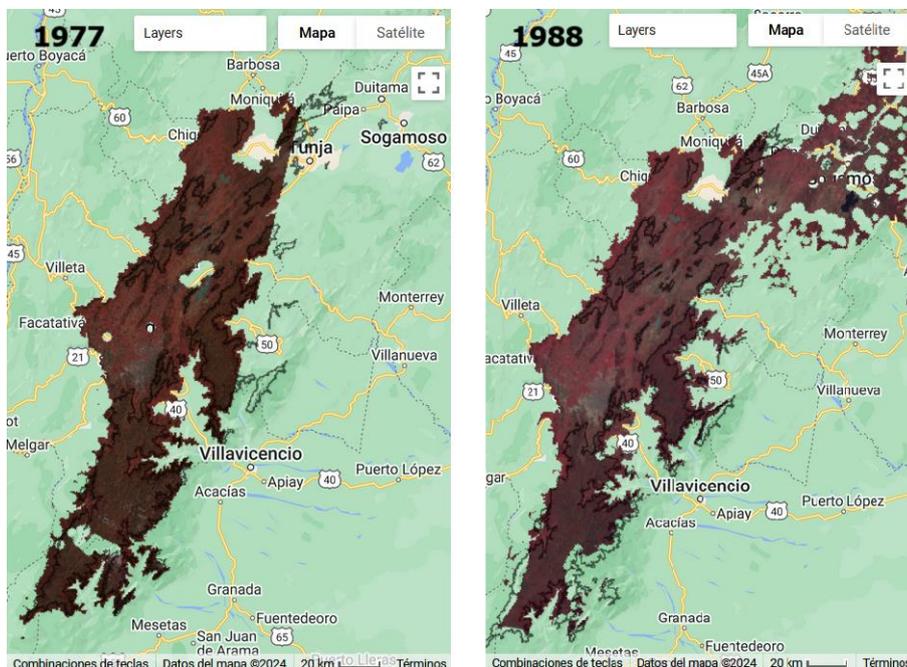
---

[15] https://www.usgs.gov/landsat-missions/cfmask-algorithm

- **L5T_2001** corresponding to Landsat 5 TM mosaic for the year 2001. The absence of reflectance type denotes that the mosaic is in Surface Reflectance values. Mosaic was generated using the script '*2001_Preproc_TM*'. This mosaic covers only Landsat path 8 and covers only ACB_A, GUE and CVS. Single date SR Mosaic (29-01-2001) RGB band combination shown in Figure 3.9 Right.



Figure 3.9: TM 1988, 1991 and 2001 AOI epoch mosaics generated.

- **L5T_L7E_2001** corresponding to Landsat 5 TM and Landsat 7 ETM+ mosaic for the year 2001. The absence of reflectance type denotes that the mosaic is in Surface Reflectance values. Mosaic was generated using the script '*2001_Preproc_TM*'. This mosaic comprises all AOI and completes the cover for ACB_B, CHI, IGM, and RRB. Double date Double sensor SR Mosaic (29-01-2001, 30-01-2001) RGB band combination shown in Figure 3.10 Left.

- **L7E_2007** corresponds to Landsat 7 ETM+ mosaic for the year 2007. The absence of reflectance type denotes that the mosaic is in Surface Reflectance values. Mosaic was generated using the script '*2007_Preproc_TM*'. Four dates SR Mosaic (2007-01-31, 2007-02-07, 2007-02-16, 2007-02-23) RGB band combination shown in Figure 3.10 Centre.

- **L7E_2011** corresponds to Landsat 7 ETM+ mosaic for the year 2011. The absence of reflectance type denotes that the mosaic is in Surface Reflectance values. Mosaic was generated using the script '*2011_Preproc_TM*'. Five dates SR Mosaic (2011-01-01, 2011-01-10, 2011-01-17, 2011-01-26, 2011-02-02) RGB band combination shown in Figure 3.10 Right.

- **L8O_2015** corresponds to Landsat 8 OLI mosaic for the year 2015. The absence of reflectance type denotes that the mosaic is in Surface Reflectance values. Mosaic was generated using the script '*2015_Preproc_OLI*'. Four dates SR

Mosaic (2014-12-28, 2015-01-04, 2015-01-13, 2015-02-21) RGB band combination shown in Figure 3.11 Left.

- **L8O_2019** corresponds to Landsat 8 OLI mosaic for the year 2019. The absence of reflectance type denotes that the mosaic is in Surface Reflectance values. Mosaic was generated using the script '*2019_Preproc_OLI*'. Double date SR Mosaic (30-12-2018, 25-02-2019) RGB band combination shown in Figure 3.11 Right.



Figure 3.10: ETM+ 2001, 2007, and 2011 AOI epoch mosaics generated.



Figure 3.11: OLI 2015, and 2019 AOI epoch mosaics generated.

**MSI Pre-processing**

Sentinel-2 pre-processing required cloud masking, performed by a join between the S2 Level2A dataset and the S2 cloud probability[16] with some scenes masking 30 to 40% and over cloud probabilities. Elevation masking was done by the MERIT DEM, for elevations lower than 2.500 m and was also applied to all tile collections. Colour correction by histogram matching using a custom function—according to Data integration section of Sub-Chapter 2.3— was performed to neighbouring tiles from different Landsat-equivalent paths (7 and 8), with targets and references denoted in Appendix A.5. Final mosaic was built overlaying one collection over the other, with the best imagery on top. The resulting mosaic was assigned a code according to the following scheme [*Mission_Level_year_reflectance-type*], resulting in:

- **S2A_2019** corresponds to Sentinel-2 Level 2A mosaic for the year 2019. The absence of reflectance type denotes that the mosaic is in Surface Reflectance values. Mosaic was generated using the script '*2019_Preproc_MSI*'. Double date SR Mosaic (25-12-2018, 15-02-2019) RGB band combination shown in Figure 3.12 Left.

- **S2A_2024** corresponds to Sentinel-2 Level 2A mosaic for the year 2024. The absence of reflectance type denotes that the mosaic is in Surface Reflectance values. Mosaic was generated using the script '*2024_Preproc_MSI*'. Double date SR Mosaic (23-01-2024, 25-01-2024) RGB band combination shown in Figure 3.12 Right.



Figure 3.12: MSI 2019, and 2024 AOI epoch mosaics generated.

---

[16] ee.ImageCollection("COPERNICUS/S2_CLOUD_PROBABILITY")

## 3.4.  Land Cover Classification & Accuracy Assessment

The work conducted in Stage 2 aimed to perform the best classification results possible for the Landsat Collection 2 and Sentinel-2 mosaics generated as an output in Stage 1, generating the LULC maps for each Páramo Complex of the assessment areas—GUE, RRB, ACB_A and ACB_B— for all epochs, characterising the statistics of the results and performing the accuracy assessment. This Sub-Chapter explains the class number selection procedure, the sample areas selection and characterisation, the application of the chosen RF algorithm to obtain the LULC maps, and the methodology for the statistical analysis and the accuracy assessment of the results obtained, as shown in Figure 3.13.



Figure 3.13: Stage 2 methodological flowchart.

As discussed at the end of Sub-Chapter 3.1, and considering the state of the art in terms of classification typologies mentioned in Sub-Chapter 2.4, one of the central initial decisions for processing the mosaics obtained was the choice of number and type of classes. Instead of following a pre-set schema like the LCCS, CLC or Anderson, the guiding principle adopted for the classification typology was very pragmatic: Finding the *least* number of classes that would provide a descriptive enough schema to identify the main LULC types for the AOI context in the *worst* datasets, these being the MSS data corresponding to epoch **1977** and **1988** (TOA data). This was stated under the assumption that for datasets with improved radiometric and geometric resolution— *better* datasets—, the classification schema would perform even better. This also meant that the resulting classes are more related to the data itself, rather than a fixed prior schema.

To be able to judge the suitability of a given number of classes an unsupervised *K-means* ML algorithm was implemented, according to the description of Sub-Chapter 2.4, for $K = 2$ to $K = 10$ centroids. It was found that around $K = 7 \pm 1$ centroids, the clustered data started representing the most evident land covers in the landscape, capturing the water bodies, agricultural fields, and different kind of vegetation—and their absence on bare surfaces—within the clusters. Figure 3.14 shows the resulting 7-Cluster pixel classification through K-means algorithm for two of the assessment PAs.

Initial sample areas were collected for the different clusters and the associated land cover—Water, Forest, Agriculture, Bare Soil, Grass, Shrub, Built-up, Rock, Wetlands— to assess the spectral signature differences with the MSS data, corresponding to MSS four bands (R, G, NIR1, NIR2). This initial samples were fed to a generic 10-tree RF classifier (with a 0.7/0.3 split for training and testing data) and the results would be compared to the initial clusters, identifying the most confused classes. Following this procedure for a decreasing number of classes, it was noted that some classes were very similar spectrally and the classifier would behave much better when the dimensionality was reduced when merging spectrally similar classes together. The merging of the respective land covers to the least number of classes resulted in the 5-class schema presented in Table 3.5 corresponding to **Water (W)**, **Bare-Rock-Built (B)**, **Agriculture-Grass (A)**, **Forest (F)**, **Shrubland-Wetland (S)**.

Sample areas were collected along all the bounding box extension, and not only inside the PAs, to ensure that all the defined Classes were sufficiently represented, and to prevent a government-declared boundary to possibly bias the sampling process.

Figure 3.14: 7-Cluster K-means pixel aggregation for MSS mosaics for GUE 1977 and 1988 (top) and RRB 1977 and 1988 (bottom).

Table 3.5: Land Use and Land Cover classification scheme used in the present Thesis.

| LULC Class | | | Description | Colour [hex Code] |
|---|---|---|---|---|
| Name | Number | Abbr. | | |
| Water | 0 | **W** | Natural and artificial water bodies. | [#1F9DFF] |
| Bare-Rock-Built | 1 | **B** | Quarries, gravel roads, exposed rock formations and human settlements and infrastructure. Includes greenhouse complexes and high-reflecting artificial surfaces. | [#CACDDE] |
| Agriculture-Grass | 2 | **A** | Crop fields and grass pastures used for farming and grazing. Includes areas of exposed—ploughed—soil and all crop stages. | [#B1EE2D] |
| Forest | 3 | **F** | Dense tree vegetation including native high Andean montane forest, exotic plantations (*Acacia, Eucalyptus, Pinus, etc.*) and riparian vegetation (*Polylepis, Chusquea*). | [#149721] |
| Shrubland-Wetland | 4 | **S** | Sparse vegetation including tussock grasses, shrubs, scrubs, cushion plants, ferns and rosettes (*Calamagrostis, Hypericum, Bromeliaceae, Espeletia, etc.*). Includes water saturated soil around permanent and transitional water bodies. | [#AD8D06] |

Once the 5-class schema was established, and with a rough *a priori* estimate of the LULC class proportions given by the unsupervised classification, a suitable stratified sample design was calculated using Equations (2.8) and (2.9). Due to their reduced size and proximity to the GUE and RRB complexes, ACB_A and ACB_B were subordinated in terms of the classification. This means that the classification algorithm would only be trained with pixels sampled from GUE, but the resulting classification model would be applied to ACB_A, and algorithms trained with pixels sampled from RRB would be applied to classify ACB_B. Considering a target standard deviation $S(\hat{O}) = 0.1$, the resulting minimum sample size for GUE is 932 and for RRB is 706, and the total sample—considering the minimum refers to test pixels, making up 30% of the total—resulted in 3108 and 2354, respectively, as shown in Table 3.6.

Table 3.6: Stratified sample size estimation for Páramo Complexes

| GUE Class | $W_i$ (Area%) | $U_i$ (Target UA) | $S_i$ (Target S) | $W_i*S_i$ | n (min. Test Sample) | N (min. Total Sample) |
|---|---|---|---|---|---|---|
| 0 / W | 2% | 0,95 | 0,218 | 0,0044 | | |
| 1 / B | 7% | 0,8 | 0,400 | 0,0280 | | |
| 2 / A | 48% | 0,9 | 0,300 | 0,1440 | 932 | 3108 |
| 3 / F | 28% | 0,9 | 0,300 | 0,0840 | | |
| 4 / S | 15% | 0,9 | 0,300 | 0,0450 | | |
| RRB Class | $W_i$ (Area%) | $U_i$ (Target UA) | $S_i$ (Target S) | $W_i*S_i$ | n (min. Test Sample) | N (min. Total Sample) |
| 0 / W | 1% | 0,95 | 0,218 | 0,0022 | | |
| 1 / B | 1% | 0,8 | 0,400 | 0,0040 | | |
| 2 / A | 42% | 0,95 | 0,218 | 0,0915 | 562 | 1872 |
| 3 / F | 35% | 0,95 | 0,218 | 0,0763 | | |
| 4 / S | 21% | 0,9 | 0,300 | 0,0630 | | |

With the estimation of the total number of pixels to sample, a process of interpretation of imagery across all epochs was performed to sample the stable areas according to the class overall proportion. Using the GEE scripts '*GUE_ACB_A_Processing*' and '*RRB_ACB_B_Processing*' the following process was repeated until total sampled pixels was above the estimated and all classes area proportions roughly matched the overall:

1) All Páramo complex 11 epochs NIR, RGB and AGRI composites were stacked and displayed on the map.
2) An easily distinguishable area of a given class from the 1977 epoch was located (e.g. the middle of a reservoir for W class).
3) A rectangle was drawn using the GEE tool for the given class, using the cluster layer as additional input in case of boundary doubt.
4) The rectangle was inspected to cover the same class pixels along all epochs, modifying its boundaries when needed. Layer opacity sliders were used for the different epochs and different visualisation composites. When in doubt for the

latest epochs, Google Earth imagery of the location was used to aid interpretation.

5) The final constant-class rectangle was added to the sample area feature collection. Sample areas were estimated to assess if the number of sample pixels for the given class were enough.

6) When not enough the process would start over from step 2) for the same or another class, as needed, trying to sample across all the extension of the mosaic.

The process yielded the across-epochs stable class sample polygons shown in Figure 3.15 and the areas reported in Table 3.7.



Figure 3.15: GUE (Left) and RRB (Right) sample areas.

Table 3.7: Total sample areas for GUE and RRB

| Class | GUE Samples | | A priori | RRB Samples | | A priori |
|---|---|---|---|---|---|---|
| | Area [Ha] | Area [%] | $W_i$ | Area [Ha] | Area [%] | $W_i$ |
| 0 / W | 175,36 | 5,8% | 2% | 51,51 | 7,5% | 1% |
| 1 / B | 249,32 | 8,3% | 7% | 20,98 | 3,0% | 1% |
| 2 / A | 1402,65 | 46,6% | 48% | 234,23 | 34,0% | 42% |
| 3 / F | 728,96 | 24,2% | 28% | 206,92 | 30,0% | 35% |
| 4 / S | 452,04 | 15,0% | 15% | 175,06 | 25,4% | 21% |
| Total [Ha] | 3008,33 | 100% | 100% | 688,7 | 100% | 100% |
| Total [60 m pixels] | 8356 | - | - | 1913 | - | - |

The sample area collection process proved quite slow, as many suitable areas from the 1977 epoch were later changed, or covered by clouds and therefore masked, or covered by the SLC-off gaps of Landsat 7 ETM+ acquisitions or change substantially at each geometric resolution jump (from MSS 60 m to TM/ETM+/OLI 30 m, and then to 10m with Sentinel-2 data). This proved to be more problematic on the RRB Complex, as the initial epoch did not have any data for some of the *Páramo de Guacheneque* and for all *Páramo de Cristales* polygon (refer to Table 3.1).

This sampling effort resulted in a consistent training and test dataset across epochs, allowing the use of the same training dataset polygons to sample the mosaics and to characterise the spectral responses of the classes. The following figures present the sample areas mean spectral signatures per class across all epochs, along with its coefficient of variation $CV = \sigma/\mu$ were $\sigma$ is the sample standard deviation and $\mu$ is the sample mean reflectance. The CV figure was preferred over the typical box and whisker chart due to the number of spectral signatures on each chart (11), providing a better visualisation. A supplementary sample areas interpretation according to the classification scheme and their spectral signatures is included in Appendix B.1.



Figure 3.16: GUE sample areas spectral signatures, from left to right class: W, B, A, F, S. Epoch 11 labels per class omitted for clarity, refer to Appendix B.2.

Figure 3.17: RRB sample areas spectral signatures, from left to right class: W, B, A, F, S. Epoch 11 labels per class omitted for clarity, refer to Appendix B.2.

As stated above, spectral signatures remained consistent across epochs—each line on the top graphs is close to one another and describes a similar pattern— and in between Páramo Complexes. It is also important to note that the 1977 and 1988 epochs, not only had less bands—corresponding to the shorter spectral signatures—but also were expressed in TOA reflectance, so differences from these signatures were expected. It is possible to observe that for most wavelengths the CV values are in the 0.1 to 0.3 range meaning the standard deviation makes up 10 to 30% of the band reflectance mean, except for the blue and red visible bands, that usually presented higher CV values.

Some protuberance was found on the GUE Water spectral signature due to higher reflectance on some water bodies surfaces. This effect was also observed on the RRB Water sample areas, as sometimes the Water surface reflects surrounding clouds or haze, in what is normally a very low reflectance LULC class. This phenomenon also explains why the Water CV is quite variable with respect to other classes as it has smaller values but higher relative variation. All the other classes exhibited a very consistent behaviour in the visible wavelengths with variations on the NIR and SWIR bands. Bare-Rock-Built class exhibited similar NIR wavelength responses, with some epoch variation in the SWIR, probably related to the heterogeneity within the land covers grouped.

Agriculture-Grass class presented a very interesting variation both in GUE an RRB, with alternating increasing and decreasing values for the NIR and SWIR1 reflectance, and a general epoch variation for all SWIR values. This is most likely the result of the A class encompassing all crop stages and bare—ploughed— soil. Finally, the Forest

and Shrubland-Wetland spectral signatures exhibit similar patterns, with the latter having higher reflectance values for both SWIR bands with a lower CV.

Once characterised, the sample areas were randomly split into training (70%) and testing (30%). A RF classifier was then built in GEE and the hyperparameter tuning was performed considering the best performance for the *worst* dataset principle. As explained in Sub-Chapter 2.4 the RF classifier has three hyperparameters. The number of nodes and the number of features sampled was left to GEE default values. This meant the RF classifier would be run for an increasing number of trees for the 1977 epoch, and the Overall Accuracy would be assessed to determine a suitable number of trees. Figure 3.18 shows the OA progression for one random seed and 20 random seed around the max OA, yielding a local optimal **18 trees** for GUE and **17 trees** for RRB. Data can be consulted on '*Processing_RF_init_Tree*' Microsoft Excel spreadsheet.



Figure 3.18: RF hyperparameter tunning for Páramo Complexes

To reduce spatial autocorrelation, a distance filter was implemented as to not include neighbouring pixels within a given distance in the same training dataset. Distance was set to **1000 m** for GUE and **500 m** for RRB. As it was can be noted from Figure 3.18, the random seed defined for the RF classifier has an impact on the OA. This is why the mean of 20 different iterations was used, as it was noted that accuracies tended to be normally distributed around a mean, meaning that a single run could be a value above or below the mean. It was also noted that the implementation of the area-weighted Overall Accuracy ($\hat{O}$) usually resulted in significantly different values.

To capture the variability of the RF model, an increasing set of seeds was used to determine the behaviour of the mean OA and $\hat{O}$, and to assess the normality assumption of these variables and to determine an ideal number of seeds. It should be noted that increasing the number of seeds has a significant effect on the processing time. A good trade-off quantity was **20 random seeds**, for which mean accuracies have been highlighted in Figure 3.19.

Figure 3.19: RF assessment of number of random seeds via OA and Ô distribution

With the set RF hyperparameters and the number of random seeds defined, it was decided that the model iteration strategy would be conducted for each of the 11 epochs as follows:

➢ The pixels of the randomly split sample training areas would be subject to the distance filter, according to the Páramo Complex. The new number of pixels would be computed, as well as the new train-test ratio. A total valid pixel area for the epoch was calculated, as well as the sampling area percentage.

➢ An RF classifier with the given number of trees for each Páramo Complex would be trained with the sampling data and the bounding box mosaic and the subordinate (ACB_A for GUE and ACB_B for RRB) were classified. This was done iteratively for 20 different random seeds.

➢ Results for OA, Ô, User's Accuracy, Producer's Accuracy, F-score, Area per class for the bounding box (including No Data), Area per class (including No Data) for the Government's declared Protected Area and confusion matrix were stored for each iteration. Accuracy metrics were calculated using the error matrix as explained in Sub-Chapter 2.5. These results were exported as plain text to a Microsoft Excel spreadsheet ('*GUE_ACB_A_Results*' and '*RRB_ACB_B_Results*'), from now on referred as the **Raw Results Spreadsheet**.

➤ All iteration values for all variables were statistically analysed to determine their distribution, means and confidence intervals by means of Phyton scripts '*Normality_Testing_GUE*', '*Normality_Testing_RRB*', using an ensemble of the 6 normality tests listed in Table 2.4. Inter-annual or between epochs results were also performed to determine whether class area values were statistically significant across epochs by means of Phyton scripts '*Inter_Annual_Testing_GUE*', '*Inter_Annual _Testing_RRB*', using both normal and non-normal tests listed in Table 2.5 with significance $\alpha = 0,05$. Results were summarised in a Microsoft Excel spreadsheet ('*Processing_GUE*' and '*Processing_RRB*'), from now on referred as the **Processing Spreadsheet**, under the *GUE_Stats/RRB_Stats* pages.

➤ An aggregation of the 20-classification iteration was performed around the **mode**, so each pixel class would be defined by the majority voting. However, a layer of pixel **agreement** was defined as the percentage of iterations that agreed on this mode, to assess the confidence of the final classification. This mode LULC was defined as the final classification result and a new confusion matrix and set of accuracy metrics were calculated. All iteration individual results and aggregated results were centralised in the **Processing Spreadsheet**, under the short epoch code (*77, 88, 88T, 91T, 01T, 07E, 11E, 15O, 19O, 19M and 24M*) _Seeds pages.

➤ All epochs' iterations and mode aggregation results were summarised on the **Processing Spreadsheet**, under the *GUE_All/RRB_All* pages, where the results area charts are generated. As ACB_A and ACB_B bounding boxes were subordinated to GUE and RRB, respectively, their results, along with the results of ACB_G and ACB_R were included in the GUE/RRB **Processing Spreadsheet** (see Figure 3.2).

Resulting LULC cover maps were exported in raster format, with a band including the class number (0 to 4, see Table 3.5), and a band including the agreement expressed in percentage (mode class iterations agreement divided by 20 total). The resulting LULC map was assigned a code according to the following scheme [*Páramo_Short-epoch-code*], and separate maps were generated for the separate ACB assessment areas, resulting in:

▪ *GUE_77, … , GUE_24M ; ACB_A_77, … , ACB_A_24M*

▪ *RRB_77, … , RRB_24M ; ACB_B_77, … , ACB_B_24M*

## 3.5. Change Detection & Temporal Trajectories

Change detection and temporal trajectory analysis was performed exclusively for the pixels within the Government delimited Protected Areas (PAs) and only for the pixels classified with *high confidence* (HC), defined as pixels with agreement equal or greater than 70%. Due to the results of the accuracy assessment, the change was approached from a binary perspective, with Agriculture-Grass class as the focus. This Sub-Chapter explains the binary change detection procedure and the temporal trajectory procedure, corresponding to the first part of Stage 3, as shown in Figure 3.20.



Figure 3.20: Stage 3 methodological flowchart.

From the final LULC maps, the following procedure was followed to obtain the Páramo Complex PA epoch HC binary assets and perform the change detection and temporal trajectory analysis:

> ➤ Final LULC maps were masked for pixels falling outside the corresponding Páramo Complex, as well as those under the agreement threshold (< 70%). Pixels under the agreement threshold were classified as Low Confidence (LC), and those meeting the threshold as High Confidence (HC). Areas inside the PA corresponding to non-Agriculture (Water, Bare-Rock-Built, Forest and Shrubland-Wetland), from now on defined as **Others class**, were quantified, along with Agriculture and No Data. All results were summarised in the **Processing Spreadsheet**, under the *GUE_PA_HC/RRB_PA_HC* pages.

> ➤ For the final Agriculture-Grass HC land cover percentage, a geometric correction factor is included which is calculated on the basis of each Páramo Complex as the factor that equalizes the Agriculture-Grass cover in the pivot years (1988, 2019) where data is obtained for two different sensors at comparable times. This correction is denoted by the (M) symbol as a 'MSI equivalent' cover percentage.

> ➤ Páramo Complex PA epoch HC binary assets were exported in raster format and used as input in the GEE scripts '*GUE_ACB_A_Temporal_T*' and '*RRB_ACB_B_ Temporal_T*'. Due to the considerations explained in Sub-Chapter 2.6, binary change detection was performed to the 1977 to 1988 epochs and the 2019M to 2024M epochs, while a 4-epoch temporal trajectory was analysed between the 1988T and 2019O epochs. For GUE the 4 epochs selected for the temporal trajectory were 1988T-2001T-2015O-2019O, and for RRB 1988T-2001T-2011E-2019O, as the best—and mostly statistically different—epochs. The three overall time steps were defined as **MSS 77-88**, **TEO 88-19** (For the initials of the three instruments TM/ETM+/OLI), and **MSI 19-24**. Results were registered in the **Processing Spreadsheet**, under the *GUE_PA_Change/RRB_PA_Change* pages. Just like with the LULC results, ACB_A and ACB_B, along with the results of ACB_G and ACB_R (see Figure 3.2) were included in the GUE/RRB **Processing Spreadsheet**.

For the binary change detection time steps **MSS 77-88** and **MSI 19-24**, the classes where coded as 1 for Other class and 2 for Agriculture-Grass class. The change categories are defined in Table 3.8.

For the binary temporal trajectory of time step **TEO 88-19**, the change categories are defined in Table 3.9.

Table 3.8: Binary change detection typology

| Bi-temporal Change | | To Class | | Transition Code | Change Name | Colour [Hex code] |
|---|---|---|---|---|---|---|
| | | O / 1 | A / 2 | 101 | Stable Other | [#000000] |
| From Class | O / 1 | 101 | 102 | 102 | Stepped Gain | [#912F40] |
| | A / 2 | 201 | 202 | 201 | Stepped Loss | [#FFC100] |
| | | | | 202 | Stable Agriculture-Grass | [#B1EE2D] |

Table 3.9: Binary temporal trajectory typology

| Transition Code | Change Name | Change Name Simplified | Colour [Hex code] |
|---|---|---|---|
| 1111 | Stable Other | | |
| 1121 | (Stable Other) | Stable Other | [#000000] |
| 1211 | (Stable Other) | | |
| 1112 | Stepped Gain | | |
| 1122 | Stepped Gain | Stepped Gain | [#912F40] |
| 1222 | Stepped Gain | | |
| 1212 | Cyclical | | |
| 1221 | (Cyclical) | Cyclical | [#725AC1] |
| 2112 | (Cyclical) | | |
| 2121 | Cyclical | | |
| 2111 | Stepped Loss | | |
| 2211 | Stepped Loss | Stepped Loss | [#FFC100] |
| 2221 | Stepped Loss | | |
| 2122 | (Stable Agriculture) | | |
| 2212 | (Stable Agriculture) | Stable Agriculture | [#B1EE2D] |
| 2222 | Stable Agriculture | | |

LULC change maps were generated for each of the three overall time steps and exported as raster assets as:

- *GUE_change_MSS, GUE_change_TEO, GUE_change_MSI.*
- *RRB_change_MSS, RRB_change_TEO, RRB_change_MSI.*
- *ACB_A_change_MSS, ACB_A _change_TEO, ACB_A _change_MSI.*
- *ACB_B_change_MSS, ACB_B _change_TEO, ACB_B _change_MSI.*

## 3.6. Agriculture Disturbance Index

The agriculture disturbance index proposed—the Index—aims to capture the effect of the Agricultural LULC within the Páramo Complex PA over the duration of the 47-year period analysed in this Thesis, as a *spatial-temporal* index. The Index is different from the *latest* Agriculture-Grass land cover figure because it is trying to account for the different Agriculture-Grass covers obtained and the positive or negative change for Agriculture-Grass cover within the given time steps. Therefore, it tries to represent a historic overview over time, as opposed to the latest figure that only represents a recent snapshot in time. The aim is to provide a value that incorporates *all the states and changes* in Agriculture-Grass that the Páramo area has undergone over the study period.

The results obtained from the different sensors cannot be directly compared, so the study period is divided into 3 consecutive time steps given the geometric resolution of the sensors (60 m, 30 m and 10 m). The Index will incorporate a variable that accounts for the *initial state* of the Agriculture-Grass cover for the step, a *change factor* that accounts for the changes of the Agriculture-Grass cover within this step and a *weight*.

The Index is defined *per area* and *per time step* as the multiplication of an initial state variable by a change factor. The disturbance index for the whole study period will be a *weighted average* of each step disturbance index. The steps are defined by the comparability of results given the geometric resolution of the sensor: Step 1 is the **MSS 77-88** (60 m), Step 2 is the **TEO 88-19** (30 m) and Step 3 is the **MSI 19-24** (10 m). Equation (3.1) presents the general definition for the step Index.

$$I_{step} = I_{t_1 \rightarrow t_2} = Initial\ state * Change\ Factor \tag{3.1}$$

The best *initial state* variable obtained is the High Confidence (HC) Agriculture-Grass Area cover percentage, calculated with respect to the total PA. For the case of Step 1 and 2 percentages, the geometrically corrected Agriculture-Grass cover percentage—denoted by the (M) symbol— is used. The change factor is designed as to penalize the *net* Agriculture-Grass change in the time step, increasing it or decreasing it at the same percentual change, therefore defined as 1 plus the net change. The *net* change is defined as Agriculture-Grass gain minus Agriculture-Grass loss.

The equation for the first step **MSS 77-88** is therefore defined as in Equation (3.2).

$$I_{MSS} = I_{77 \rightarrow 88} = A_{77(M)}(1 + \Delta A_{77 \rightarrow 88}) \tag{3.2}$$

Where $I_{MSS}$ is the partial index for the first step comprising 1977 to 1988, $A_{77(M)}$ is the High Confidence (HC) Agriculture-Grass Area cover percentage on the year 1977 corrected to be MSI 10 m resolution equivalent, and the net change in Agriculture-Grass is defined in Equation (3.3).

$$\Delta A_{77\to88} = Gain_{77\to88} - Loss_{77\to88} \tag{3.3}$$

Where $Gain_{77\to88}$ and $Loss_{77\to88}$ are percentages of Agriculture-Grass stepped gain and stepped loss (see Table 3.8), calculated with respect to the total PA.

For the middle step **TEO 88-19**, given that it is not a binary change detection, but a multitemporal binary trajectory, another change variable is introduced inside the Change Factor. Given the 16 possible trajectories, simplified to 5 in Table 3.9 (Stable Other, Stepped Agriculture-Grass Gain, Cyclical, Stepped Agriculture-Grass Loss and Stable Agriculture-Grass), it was considered that the factor should incorporate the other non-stable trajectory represented by the *Cyclical*. Equation (3.4) defines this step.

$$I_{TEO} = I_{88\to19} = A_{88(M)}(1 + Cyc_{88\to19} + \Delta A_{88\to19}) \tag{3.4}$$

Where $I_{TEO}$ is the partial index for the second time step comprising 1988 to 2019, $A_{88(M)}$ is the High Confidence (HC) Agriculture-Grass Area cover percentage on the year 1988 corrected to be MSI 10 m resolution equivalent, and the net change in Agriculture-Grass is defined by Equation (3.5).

$$\Delta A_{88\to19} = Stepped\ Gain_{88\to19} - Stepped\ Loss_{88\to19} \tag{3.5}$$

For the last step the disturbance index is defined as in Equation (3.6).

$$I_{MSI} = I_{19\to24} = A_{19M}(1 + \Delta A_{19\to24}) \tag{3.6}$$

Where $I_{MSI}$ is the partial index for the first step comprising 2019 to 2024, $A_{19M}$ is the High Confidence (HC) Agriculture-Grass Area cover percentage on the year 2019, and the net change in Agriculture-Grass is defined by Equation (3.7).

$$\Delta A_{19\to24} = Gain_{19\to24} - Loss_{19\to24} \tag{3.7}$$

The general disturbance index for a given PA results in the weighted average of the previous indexes, yielding Equation (3.8).

$$I_{PA} = \frac{\sum_{All\ time\ steps}(I_{step} * W_i)}{\sum W_i} * 100 \tag{3.8}$$

Developing Equation (3.8) it becomes Equation (3.9).

$$I_{PA} = \frac{A_{77(M)}(1 + \Delta A_{77\to88})W_{MSS} + A_{88(M)}(1 + Cyc_{88\to19} + \Delta A_{88\to19})W_{TEO} + A_{19M}(1 + \Delta A_{19\to24})W_{MSI}}{W_{MSS} + W_{TEO} + W_{MSI}} * 100 \tag{3.9}$$

The weights of each time step index are envisaged in such way as to follow these principles:

$$Weight \propto Sensor\ Radiometric\ Resolution$$
$$Weight \propto Time\ Period\ in\ Years$$

The weight for each step is therefore defined as in Equation (3.10).

$$W_i = b * \Delta t \tag{3.10}$$

Where $b$ is the numbers of bits of the spectral bands used for the classification (see Table 2.1) and $\Delta t$ is the difference between the last and initial year of the step. The Equation above yields for the different time steps:

$$W_{MSS} = 6 * (1988 - 1977) = 6 * 9 = 54$$

$$W_{TEO} = 8 * (2019 - 1988) = 8 * 31 = 248$$

$$W_{MSI} = 12 * (2024 - 2019) = 12 * 5 = 60$$

To illustrate the result in terms of absolute weights:

$$\frac{W_{MSS}}{\sum W_i} = \frac{54}{54 + 248 + 60} = 0,1492$$

$$\frac{W_{TEO}}{\sum W_i} = \frac{248}{54 + 248 + 60} = 0,6851$$

$$\frac{W_{MSI}}{\sum W_i} = \frac{60}{54 + 248 + 60} = 0,1657$$

Therefore Equation (3.9) can be simplified to Equation **¡Error! No se encuentra el origen de la referencia.**.

$$I_{PA} = A_{77(M)}(1 + \Delta A_{77 \to 88})14,92 + A_{88(M)}(1 + Cyc_{88 \to 19} + \Delta A_{88 \to 19})68,51 + A_{19M}(1 + \Delta A_{19 \to 24})16,57 \tag{3.11}$$

The Index is *bounded* between 0 and 100 as all the variables are defined as percentages with respect to total PA. The index is *optimistic* as it doesn't consider No Data (N.D.) values that could be accounting for more Agricultural-Grass LULC, it only accounts for the HC pixels of the class with one of the highest accuracies—therefore, trying to reduce the model variance—. It also allows for *net* loss of agriculture to *reduce* the partial step Index, in given case.

The weight definition rewards the *radiometric resolution* of the bands used for classification as the spectral analysis of the sample areas (see Sub-Chapter 3.4 and Appendix B.2) show similar behaviour between classes at similar wavelengths. It also rewards the amount of *time in each time step* as a proportion of the whole study period.

The introduction of the spectral resolution in the weights, despite having made a *correction* to the Agriculture-Grass area cover percentage, responds to the fact that the oldest datasets have a *higher uncertainty* due to the co-registration of the images, the lack of atmospheric correction (TOA reflectance was used for classification), the custom-made cloud and cloud-shadow masking, and the difficulty of sample areas interpretation, among others (see Sub-Chapter 3.3 and Appendix B.1). It is the

consideration of the author that the processing pipeline followed becomes more robust and its results more reliable, as time and instrument progress, and is therefore given more weight in the average. The assumption is that older results are considered less reliable and recent ones are considered the most reliable, and the bit resolution is used as a mean for this end.

According to its mathematical formulation, the Index should be interpreted as a representative—and optimistic lower bound— persistent Agriculture-Grass LULC percentage of the given area for the whole study period. This means that an Index of value 50 for a given area should be interpreted as: '*At least half the area has been regularly under a LULC of Agriculture-Grass for the last 47 years, and at present could be greater*'.

# 4  Results

This chapter presents the Páramo areas Land Use/Land Cover results obtained from the satellite imagery mosaics, their change in time, along with the corresponding accuracies and error estimation. It also calculates the disturbance Index for the different Páramo areas.

Before presenting the classification results obtained, it is pertinent to review some of the caveats of the class schema chosen for the present work. Following the methodology description in Sub-Chapter 3.4 and the classification scheme presented in Table 3.5, some key clarifications need to be made for the interpretation of the results.

❖ There are inherent limitations related to the Land Cover Typology adopted:

   o Bare-Rock-Built class incorporates both *natural* and *artificial* land cover, so this class extent cannot be indicative of anthropogenic disturbance exclusively.

   o Forest class incorporates both *native* and *exotic* tree vegetation, therefore it is not possible to assess the disturbance due to the incorporation of the exotic and/or invasive species.

   o Shrubland-Wetland includes Páramo vegetation as well as *other* water-saturated sparse vegetation land cover. Consequently, this class extent cannot be a sole indicator of Páramo extent.

❖ There are also limitations related to the datasets and pre-processing in a cloud-prone and mountainous area such as the one studied:

   o No Data (N.D.) values, mostly due to *cloud* and *cloud shadow* masked pixels, and due to *gaps* from the ETM+ sensor malfunction (SLC-off), that constrain the analysis.

   o Pixel confusion due ridge shadow, as well as haze and cloud shadows omitted by the cloud-masking algorithms used

Some considerations have been made to address these limitations. The positive trade-off of the simplicity of the 5-class scheme adopted is that it performs well both in the *worst* and the *best* datasets, a key aspect to a long-spanning macroanalysis as the one developed. Also, the Agriculture-Grass class proved to be not only the *simplest* to interpret, but also the most *informative* and *accurate*. This is what led to the change detection and temporal trajectories to be focused on an *agriculture/non-agriculture* binary scheme. Lastly, special attention was given to N.D. values and area percentages, as well as area error estimations and confidence intervals reported.

# 4.1.   Land Cover Classification & Accuracy Assessment

This Sub-Chapter will present the classification LULC maps for each Páramo Complex of the assessment areas—GUE, RRB, ACB_A and ACB_B— for all 11 epochs (**1977**, **1988**, **1988T**, **1991T**, **2001T**, **2007E**, **2011E**, **2015O**, **2019O**, **2019M** and **2024M**), the statistical characterisation and its corresponding accuracy assessment. These results meet the first of the four key objectives defined in Sub-Chapter 1.2: **Classify and establish a baseline land cover map for the Páramo Complex**.

The baseline LULC map for all assessment areas correspond to the **1977** epoch, except for ACB_B, where the oldest records could only be retrieved for the **1988** epoch. For visualisation purposes, LULC maps of the assessment areas were divided into the GUE bounding box a—that includes ACB_G—, and the bounding boxes of ACB_A, ACB_B and RRB —that includes ACB_R— (refer to Figure 3.2). It should be noted that as explained in Sub-Chapter 3.4, GUE classifier was applied to ACB_A, and RRB classifier was applied to ACB_B.

The results will be organised as follows: First, a baseline classification results for GUE and RRB will be analysed, along with their statistical results and accuracy assessments. Second, the baseline, the *pivots* and the most recent epoch Mode LULC maps— generated in QGIS[17] with the exported GEE outputs—will be presented. Third, an analysis of the whole study period (1977-2024) class accuracy-derived areas and the high confidence results for the PAs pixel-count results will be performed, highlighting overall trends. Fourth, the general accuracy and class-specific accuracy behaviours will be interpreted.

**1.  Baseline classification results for GUE 1977**

With an average Overall Accuracy of 91,94% and area-weighted Overall Accuracy (Ô) of 90,00% for the 20 RF model iterations, the results for the GUE baseline epoch were considered satisfactory. The Mode aggregation sample counts ($n_{ij}$) error matrix and estimated area proportions ($\hat{p}_{ij}$) are presented in Table 4.1 and Table 4.2, respectively.

Table 4.1: GUE 1977 Mode sample count error matrix.

| | | Test dataset | | | | | |
| | Map | 0 / W | 1 / B | 2 / A | 3 / F | 4 / S | Total |
|---|---|---|---|---|---|---|---|
| | 0 / W | 56 | 0 | 0 | 0 | 0 | 56 |
| | 1 / B | 0 | 88 | 4 | 0 | 0 | 92 |
| Classified | 2 / A | 0 | 44 | 1723 | 22 | 19 | 1808 |
| | 3 / F | 0 | 10 | 8 | 564 | 71 | 653 |
| | 4 / S | 0 | 0 | 5 | 37 | 199 | 241 |
| | Total | 56 | 142 | 1740 | 623 | 289 | 2850 |

Table 4.2: GUE 1977 Mode estimated area proportions error matrix.

---

[17] QGIS.org (2024). QGIS Geographic Information System. Open Source Geospatial Foundation Project. http://qgis.org

| | | Test dataset | | | | | |
|---|---|---|---|---|---|---|---|
| **Map** | | **0 / W** | **1 / B** | **2 / A** | **3 / F** | **4 / S** | **Total%** |
| | **0 / W** | **1,6** | 0,0 | 0,0 | 0,0 | 0,0 | **1,6** |
| | **1 / B** | 0,0 | **8,2** | 0,4 | 0,0 | 0,0 | **8,6** |
| Classified | **2 / A** | 0,0 | 1,0 | **40,7** | 0,5 | 0,4 | **42,7** |
| | **3 / F** | 0,0 | 0,5 | 0,4 | **26,5** | 3,3 | **30,7** |
| | **4 / S** | 0,0 | 0,0 | 0,3 | 2,5 | **13,6** | **16,5** |
| **Total%** | | **1,6** | **9,7** | **41,7** | **29,5** | **17,4** | **100,0** |

For every epoch, statistical aggregation both with respect the mode and the mean of the iterations was included result summary for comparison. The results summary was compiled in a Table divided in four distinct sections: The main Páramo Complex results (GUE), the subordinated Páramo Complex results (ACB_A), the main Páramo Complex Protected Area (GUE_PA) results and the subordinated Páramo Complex (ACB_A_PA) results. Table 4.3 presents the result summary for GUE/ACB_A 1977.

Table 4.3: GUE/ACB_A 1977 epoch classification results summary

| **20 Seed Mode** | | **0 / W** | **1 / B** | **2 / A** | **3 / F** | **4 / S** | | | |
|---|---|---|---|---|---|---|---|---|---|
| **GUE 1977** | Total% | 1,57 | 9,75 | 41,75 | 29,53 | 17,40 | | **Ô** | **90,57%** |
| | (SE%) | 0,00 | 0,28 | 0,34 | 0,57 | 0,56 | | **(SE)** | **0,642%** |
| | Prod% | 100,00 | 84,53 | 97,38 | 89,67 | 78,27 | | | |
| | (SE%) | 0,00 | 3,46 | 0,65 | 1,88 | 5,34 | | | |
| | User% | 100,00 | 95,65 | 95,30 | 86,37 | 82,57 | | | |
| | (SE%) | 0,00 | 2,14 | 0,50 | 1,34 | 2,45 | | | |
| | F-score | 100,00 | 75,21 | 97,13 | 88,40 | 75,09 | | | |
| Mode Pixel count | Area [Ha] | 3023,82 | 16605,68 | 82237,27 | 59100,58 | 31804,03 | | Total | 192771,38 |
| Accuracy Derived | Âⱼ [Ha] | 3023,82 | 18790,11 | 80476,90 | 56928,98 | 33551,58 | | Total | 192771,4 |
| | (SE)[Ha] | 0,00 | 542,33 | 665,01 | 1104,57 | 1077,60 | | | |
| Mean Pixel count | Area [Ha] | 3006,01 | 16398,72 | 82480,83 | 59105,12 | 31780,71 | | Total | 192771,39 |
| | (SE)[Ha]* | 7,61 | 107,93 | 181,41 | 192,53 | 124,38 | | | |
| **ACB_A 1977** | | | | | | | | | |
| Mode Pixel count | Area [Ha] | 273,33 | 944,96 | 13685,26 | 3857,87 | 3293,95 | | Total | 22055,37 |
| | Area% | 1,24 | 4,28 | 62,05 | 17,49 | 14,93 | | | |
| Mean Pixel count | Area [Ha] | 272,60 | 950,74 | 13627,66 | 3901,49 | 3302,88 | | Total | 22055,37 |
| | (SE)[Ha]* | 0,91 | 5,90 | 30,28 | 27,70 | 14,39 | | | |
| **GUE 1977 PA** | | | | | | | | | |
| Mode Pixel count | Area [Ha] | 259,15 | 66,38 | 9890,15 | 18503,01 | 14506,12 | | Total | 43224,81 |
| | Area% | 0,60 | 0,15 | 22,88 | 42,81 | 33,56 | | | |
| Mean Pixel count | Area [Ha] | 253,65 | 81,32 | 9886,53 | 18578,00 | 14425,31 | | Total | 43224,81 |
| | (SE)[Ha]* | 4,19 | 2,89 | 45,51 | 51,17 | 37,41 | | | |
| **ACB_A 1977 PA** | | | | | | | | | |
| Mode Pixel count | Area [Ha] | 0,26 | 4,53 | 921,50 | 663,63 | 261,73 | | Total | 1851,65 |
| | Area% | 0,01 | 0,24 | 49,77 | 35,84 | 14,13 | | | |
| Mean Pixel count | Area [Ha] | _0,15_ | 4,74 | 914,09 | 670,26 | 262,39 | | Total | 1851,64 |
| | (SE)[Ha]* | 0,04 | 0,08 | 3,33 | 3,29 | 1,39 | | | |

*Pixel count SE was calculated as $SE = \sigma/\sqrt{n}$, where $\sigma$ is the standard deviation and $n$, the number of iterations.

Underlined results in italics indicate that the 20 iteration results were not normally distributed.

One of the intentions for including both median and mode results was to compare the mean area pixel count with the accuracy-derived estimator $\hat{A}_j$ and their Standard Errors. It can be noted that most mode and mean pixel-count values fall outside the confidence interval (±1,96*SE) of the $\hat{A}_j$, and their SE* are one order of magnitude under those of the accuracy-derived counterpart. The mean SE* can be interpreted as

proportional to the variance of the iteration ensemble, and therefore it tends to be reduced with the number of random seeds (e.g. it would be even smaller for 30 or 40 iterations) as opposed to the SE of the statistical inference. It should be noted that the results for the PAs, do not have statistical inference-based SE, as they are a sub-sample of the total map, with a different area distribution.

It was noticed that small area values for a class tended not to be normally distributed, as they often corresponded to near-zero number of pixels, like class W area for ACB_A 1977 PA mean pixel count. These non-normally distributed mean values are denoted by italics and underlining on the results summary. All the other variables are normally distributed, according to the normality test ensemble described in Table 2.4.

Regarding the *individual class accuracy metrics* and their SE, a comparison from the mode and the mean aggregation was also made by means of their respective scatter plots. Figure 4.1 shows the improvement in accuracy achieved by the mode aggregation, which is noticeable in the Bare-Rock-Built, Forest and Shrubland-Wetland classes.



Figure 4.1: GUE 1977 Accuracy Scatter Plot, Mode (Left) vs. Mean (Right)

All these mean vs. mode aggregation comparisons from the results support the approach taken in the methodology and confirm they do improve the classification.

**Baseline classification results for RRB 1977**

With an average OA= 92,62% and Ô=95,77% for the 20 RF model iterations, the results for the RRB baseline epoch were considered satisfactory, despite the missing path of data for the epoch (Landsat path 7). The Mode aggregation sample counts ($n_{ij}$) error matrix and estimated area proportions ($\hat{p}_{ij}$) are presented in Table 4.4 and Table 4.5, respectively.

Table 4.4: RRB 1977 Mode sample count error matrix.

| | Test dataset | | | | | |
|---|---|---|---|---|---|---|
| **Map** | **0 / W** | **1 / B** | **2 / A** | **3 / F** | **4 / S** | **Total** |
| **0 / W** | **114** | 0 | 0 | 1 | 0 | **115** |
| **1 / B** | 0 | **15** | 0 | 0 | 0 | **15** |
| **2 / A** | 0 | 0 | **178** | 0 | 0 | **178** |
| **3 / F** | 0 | 0 | 1 | **139** | 4 | **144** |
| **4 / S** | 0 | 0 | 3 | 41 | **200** | **244** |
| **Total** | **114** | **15** | **182** | **181** | **204** | **696** |

*Classified (row label)*

Table 4.5: RRB 1977 Mode estimated area proportions error matrix.

| | Test dataset | | | | | |
|---|---|---|---|---|---|---|
| **Map** | **0 / W** | **1 / B** | **2 / A** | **3 / F** | **4 / S** | **Total%** |
| **0 / W** | **0,3** | 0,0 | 0,0 | 0,0 | 0,0 | **0,3** |
| **1 / B** | 0,0 | **0,5** | 0,0 | 0,0 | 0,0 | **0,5** |
| **2 / A** | 0,0 | 0,0 | **40,7** | 0,0 | 0,0 | **40,7** |
| **3 / F** | 0,0 | 0,0 | 0,3 | **44,4** | 1,3 | **46,0** |
| **4 / S** | 0,0 | 0,0 | 0,2 | 2,1 | **10,2** | **12,5** |
| **Total%** | **0,3** | **0,5** | **41,2** | **46,5** | **11,5** | **100,0** |

*Classified (row label)*

For this results summary the main Páramo Complex results (RRB), the subordinated Páramo Complex results (ACB_B), the main Páramo Complex Protected Area (RRB_PA) results and the subordinated Páramo Complex (ACB_B_PA) are presented in Table 4.6.

Table 4.6: RRB/ACB_A 1977 epoch classification results summary

| 20 Seed Mode | | 0 / W | 1 / B | 2 / A | 3 / F | 4 / S | | | |
|---|---|---|---|---|---|---|---|---|---|
| **RRB 1977** | Total% | 0,25 | 0,52 | 41,18 | 46,55 | 11,51 | | Ô | **96,15%** |
| | (SE%) | 0,00 | 0,00 | 0,33 | 0,76 | 0,70 | | (SE) | **0,769%** |
| | Prod% | 100,00 | 100,00 | 98,85 | 95,49 | 88,89 | | | |
| | (SE%) | 0,00 | 0,00 | 0,80 | 0,92 | 5,72 | | | |
| | User% | 99,13 | 100,00 | 100,00 | 96,53 | 81,97 | | | |
| | (SE%) | 0,87 | 0,00 | 0,00 | 1,53 | 2,47 | | | |
| | F-score | 99,56 | 100,00 | 98,89 | 85,54 | 89,29 | | | |
| Mode Pixel count | Area [Ha] | 159,15 | 321,18 | 25369,79 | 28698,96 | 7779,69 | | Total | 62328,77 |
| Accuracy Derived | $\hat{A}_j$ [Ha] | 157,77 | 321,18 | 25664,74 | 29011,10 | 7173,99 | | Total | 62328,8 |
| | (SE)[Ha] | 1,38 | 0,00 | 206,05 | 475,80 | 437,18 | | | |
| Mean Pixel count | Area [Ha] | 159,98 | 331,03 | 25344,01 | 28524,83 | 7968,93 | | Total | 62328,78 |
| | (SE)[Ha]* | 0,85 | 9,94 | 75,44 | 88,15 | 42,93 | | | |
| **ACB_B 1977** | | | | | | | | | |
| Mode Pixel count | Area [Ha] | - | - | - | - | - | | Total | 0,00 |
| | Area% | - | - | - | - | - | | | |
| Mean Pixel count | Area [Ha] | - | - | - | - | - | | Total | 0,00 |
| | (SE)[Ha]* | - | - | - | - | - | | | |
| **RRB 1977 PA** | | | | | | | | | |
| Mode Pixel count | Area [Ha] | 137,53 | 47,78 | 2197,47 | 11950,63 | 4637,07 | | Total | 18970,48 |
| | Area% | 0,72 | 0,25 | 11,58 | 63,00 | 24,44 | | | |
| Mean Pixel count | Area [Ha] | 136,83 | 48,75 | 2197,44 | 11863,67 | 4723,79 | | Total | 18970,48 |
| | (SE)[Ha]* | 0,45 | 2,48 | 19,42 | 26,45 | 20,76 | | | |
| **ACB_B 1977 PA** | | | | | | | | | |
| Mode Pixel count | Area [Ha] | - | - | - | - | - | | Total | 0,00 |
| | Area% | - | - | - | - | - | | | |
| Mean Pixel count | Area [Ha] | - | - | - | - | - | | Total | 0,00 |
| | (SE)[Ha]* | - | - | - | - | - | | | |

*Pixel count SE was calculated as $SE = \sigma/\sqrt{n}$, where $\sigma$ is the standard deviation and $n$, the number of iterations.

It can be noted that in this case, the mode and mean pixel-count values do fall within the confidence interval (±1,96*SE) of the $\widehat{A_J}$. However, their SE* is still one order of magnitude under those of the accuracy-derived counterpart when the accuracy is not 100%. All variable distributions resulted normal, according to the tests performed.

Lastly, regarding the *individual class accuracy metrics* and their SE, Figure 4.2 shows the improvement in accuracy achieved by the mode aggregation, noticeable in the Forest and Shrubland-Wetland classes.



Figure 4.2: RRB 1977 Accuracy Scatter Plot, Mode (Left) vs. Mean (Right)

The error matrices and results summaries for the pivot years epochs and the most recent epoch (**1988**, **1988T**, **2019O**, **2019M**, **2024M**) can be consulted in Appendix B.3. For the rest of the epochs, please refer to the '*Processing_GUE*' and '*Processing_RRB*' Microsoft Excel spreadsheets.

## 2. LULC Maps

The resulting LULC mode map from the classification and the agreement map for GUE are shown in Figures 4.3 to 4.7, and for RRB are shown in Figures 4.8 to 4.13.

Figure 4.3: GUE and ACB_G 1977 LULC mode Map classification results.



Figure 4.4: GUE and ACB_G 1988 LULC mode Map classification results.

Figure 4.5: GUE and ACB_G 1988T LULC mode Map classification results.



Figure 4.6: GUE and ACB_G 2019O LULC mode Map classification results.

Figure 4.7: GUE and ACB_G 2019M LULC mode Map classification results.



Figure 4.8: RRB, ACB(_A,_B,_R) 1977 LULC mode Map classification results.

Figure 4.9: RRB, ACB(_A,_B,_R) 1988 LULC mode Map classification results.



Figure 4.10: RRB, ACB(_A,_B,_R) 1988T LULC mode Map classification results.

Figure 4.11: RRB, ACB(_A,_B,_R) 2019O LULC mode Map classification results.



Figure 4.12: RRB, ACB(_A,_B,_R) 2019M LULC mode Map classification results.

Figure 4.13: RRB, ACB(_A,_B,_R) 2024M LULC mode Map classification results.

The remaining LULC and agreement maps (**1991T**, **2011E**, **2007E**, **2011E**, **2015O**) can be consulted in Appendix B.4.

With respect to the agreement LULC Map, it should be noted that although low agreement pixels are scattered across all areas, they are located mostly on land cover class interfaces/boundaries.

### 3. Study period results

Accuracy-derived areas per class, as well as their SE, were normalised over the *total valid* area of the bounding box scene. This area corresponds to the one with elevation above 2.500 m within the bounding boxes, this is, the maximum amount of area if there were no other masked pixels. The maximum area value was used to estimate the percentage of masked/invalid pixels (No Data pixels) and to provide a *common benchmark* for all epochs, as the amount of total valid pixels varies from epoch to epoch. The same approach was taken with respect to the PAs area percentages reported, using the total areas already shown in Table 3.1. To provide a unique reference, all normalisation areas are grouped in Table 4.7.

Table 4.7: Normalisation areas for reported area percentages

| Description | Max. Valid Area [Ha] |
|---|---|
| GUE bounding box, with elevation mask ≥2.500 m | 199.190 |
| GUE PA Polygons total | 43.418,09 |
| ACB_A bounding box, with elevation mask ≥2.500 m | 22.166,28 |
| ACB_A PA Polygons total | 1.861,55 |
| ACB_G PA Polygons total | 480,62 |
| RRB bounding box, with elevation mask ≥2.500 m | 110.245 |
| RRB PA Polygons total | 24.755,95 |
| ACB_B bounding box, with elevation mask ≥2.500 m | 6.319,92 |
| ACB_B PA Polygons total | 1.468,76 |
| ACB_R PA Polygons total | 417,77 |

Accuracy-derived area percentages for GUE and RRB bounding box area presented in the LULC maps was estimated for all epochs, resulting in  Figure 4.14 and Figure 4.15, respectively. In both figures, it can be noted that the Standard Errors are not that significant with respect to the LULC class proportions.

The apparent LULC class trends are difficult to capture when compared to the percentage of No Data (N.D.) values, particularly for the RRB case. For GUE, the most prevalence of N.D. values is in the middle of the study period, which does not allow to determine if the apparent changes of proportions are due to changes in land cover of due to the missing data. In both cases the most prevalent class is Agriculture-Grass which overall, presents a tendency to increase.



Figure 4.14: GUE bounding box results for the study period.

Figure 4.15: RRB bounding box results for the study period.

More insight of the class behaviour is drawn when the results are considered only within the boundaries of the Government-delimited PAs. Although this means that no inference SE can be properly estimated, it also means that the agreement layer produced can be integrated into the analysis using the High Confidence (HC) and Low Confidence (LC) area percentages. A further step was taken as to incorporate the effect of N.D. values, as the error bars shown in all the following PA charts represent the *maximum* value the HC area proportion could reach if both the N.D. pixels and the LC pixels, followed the same class distribution from all the valid pixels. This is a reasonable assumption considering that for the PAs, the total valid pixels represent more than 90% of the total area, with some few exceptions. Figures 4.16, 4.17 and 4.18 show the results for GUE, ACB_G and ACB_A, respectively.



Figure 4.16: GUE Protected Area results for the study period. High Confidence (HC) areas are composed of pixels with agreement equal or greater than 70%.

Figure 4.17: ACB_G Protected Area results for the study period. High Confidence (HC) areas are composed of pixels with agreement equal or greater than 70%.



Figure 4.18: ACB_A Protected Area results for the study period. High Confidence (HC) areas are composed of pixels with agreement equal or greater than 70%.

As a general observation, it can be noted that the maximum values incorporated in the bars are helpful when trying to find overall trends and that each PA has its own different composition and class evolution in time. This evolution is harder capture for the least prevalent classes like W or B. For the case of GUE, it can be noted that F and S areas have tended to decrease, while A has had a steady increase in proportion. The same overall trends are followed by ACB_A, with the difference being the much higher prevalence of the A class. ACB_G has had fluctuations on A and F class proportions, but there is a tendency for the S class to decrease, while the B class has increased.

Figures 4.19, 4.20 and 4.21 show the results for RRB, ACB_R and ACB_B, respectively. It is worth pointing out that RRB is the only PA were F is the dominant class, with some fluctuations, and slight declines in S cover as well as a slight increase in A cover proportion. ACB_R does exhibit an increase in A class proportion, a slight decline in the F class and a very low prevalence of the S class. ACB_B in contrast, presents stability on it's a and F classes, with a slight decline of the S class.

Figure 4.19: RRB Protected Area results for the study period. High Confidence (HC) areas are composed of pixels with agreement equal or greater than 70%.



Figure 4.20: ACB_R Protected Area results for the study period. High Confidence (HC) areas are composed of pixels with agreement equal or greater than 70%.



Figure 4.21: ACB_B Protected Area results for the study period. High Confidence (HC) areas are composed of pixels with agreement equal or greater than 70%.

With respect to all the assessment areas evaluated, it is worth mentioning that none has a dominant Shrubland-Wetland cover, as it would be expected of a Páramo ecosystem Protected Area. In fact, the ACB_A, ACB_B and ACB_G PAs have a dominant Agriculture-Grass area instead, while RRB is dominated by Forest and both GUE and ACB_G distributions of the main classes, A, F and S, with none of them surpassing 50% of the cover proportion.

It is important to remind the reader that because these epoch results are a distribution and not a pixel-to-pixel comparison, so definitive conclusions about the class transformation cannot be extracted. However, the baseline for the areas can now be compared to the following epochs and the overall context of each PA is defined.

## 4. General accuracy behaviours

Given that all epochs share the same testing dataset it is possible to perform an analysis about the class accuracies for the whole study period. This allows for a cumulative sample count error matrix to be generated, as the sum of all epoch's individual matrix. Table 4.8 and Table 4.9 present the cumulative sample error matrix for GUE and RRB, respectively. A colour scale has been added to the samples to aid interpretation incorporating the Quartiles (Q1-Q3), the 0,7 Percentile (P0,7), mean value and maximum value.

Table 4.8: All-epochs cumulative sample error matrix for GUE

|  | | Test dataset | | | | |
|---|---|---|---|---|---|---|
| GUE_All | 0 / W | 1 / B | 2 / A | 3 / F | 4 / S | Total |
| 0 / W | 2183 | 0 | 0 | 1 | 0 | 2184 |
| 1 / B | 0 | 2813 | 269 | 1 | 30 | 3113 |
| 2 / A | 0 | 1691 | 65482 | 754 | 501 | 68428 |
| 3 / F | 0 | 57 | 702 | 22455 | 1165 | 24379 |
| 4 / S | 0 | 132 | 214 | 1345 | 7267 | 8958 |
| Total | 2183 | 4693 | 66667 | 24556 | 8963 | 107062 |
| Colour | | Q1-Q2 | Q2 - P0,7 | P0,7 - Q3 | Q3 - AVG | AVG - MAX |
| Range | 0 | 1-10 | 11-66 | 67-121 | 121-228 | 228-3380 |

(Classified — row axis label)

Table 4.9: All-epochs cumulative sample error matrix for RRB

|  | | Test dataset | | | | |
|---|---|---|---|---|---|---|
| RRB_All | 0 / W | 1 / B | 2 / A | 3 / F | 4 / S | Total |
| 0 / W | 3415 | 213 | 410 | 142 | 0 | 4180 |
| 1 / B | 0 | 511 | 35 | 1 | 1 | 548 |
| 2 / A | 0 | 1 | 6809 | 49 | 20 | 6879 |
| 3 / F | 2 | 11 | 159 | 5084 | 113 | 5369 |
| 4 / S | 0 | 105 | 249 | 468 | 8541 | 9363 |
| Total | 3417 | 841 | 7662 | 5744 | 8675 | 26339 |
| Colour | | Q1-Q2 | Q2 - P0,7 | P0,7 - Q3 | Q3 - AVG | AVG - MAX |
| Range | 0 | 1-62 | 63-260 | 261-391 | 392-1039 | 1040-15730 |

(Classified — row axis label)

Both cumulative matrices highlight the constant confusion between the S and F class, which have been already recognised as being spectrally similar. With respect to GUE, the confusion dynamic that follows is the B class being confused for A class, and for RRB is B class being confused for W class. If this was operated as a unique matrix (e.g. the result from a single map product) from a producer's accuracy perspective, the worst class for both PAs is the B (both ~60%). In contrast, from a user's accuracy approach, the worst class is S for GUE and W for RRB (both ~81%). User's accuracies tend to be higher and close to the targets established in the sampling design, except for the S class in GUE and the W class in RRB.

Although Overall Accuracy is high from the cumulative exercise (~92-94%), and for all the individual LULC products for GUE [88,2-95,6%] and RRB [93,6-98,8%], it is important to analyse class-specific behaviour for all epochs. For this reason, an all-epoch Producer's vs. User's accuracy scatter plots were built for GUE (Figure 4.22) and RRB (Figure 4.23). The codification for each data point is defined as follows: [*Class*]_[*Year(2-digit)*]_[*Sensor*]. Cumulative results were also included as GUE_All/RRB_All.



Figure 4.22: GUE all-epoch User's vs. Producer's accuracies.

Figure 4.23: RRB all-epoch User's vs. Producer's accuracies.

Several observations from the scattergrams presented can be made:

➢ Agriculture-Grass is the most consistently classified, with high precision, high accuracies and similar omission and commission errors (marked by the plot diagonal), and with small and similar SE. All user's and producer's accuracies are in the vicinity of 97%.

➢ Water is the class with the highest accuracies, highest precision and lowest SE, with the notable exceptions of 2011 and 2015 data points for RRB. Producer's accuracy for the class was almost absolute.

➢ Bare-Rock-Built and Shrubland-Wetland classes were the most inconsistent with low precision and larger SE. Although the latter presented a better balance for omission and commission errors.

➢ Forest class was a middle ground with respect to all other classes with respect to accuracies, precision, SE and consistency.

➢ Best overall performing classes were Water and Agriculture-Grass, followed by Forest, Shrubland-Wetland and Bare-Rock-Built.

➢ No evident correlations were found with respect to spectral signature CV, sensor or year, although the worst data points for both PAs corresponded to S_07_ETM.

## 4.2.   Change Detection & Temporal Trajectories

This Sub-Chapter will present the change detection and temporal trajectory of the LULC classes for each Páramo Complex of the assessment areas—GUE, RRB, ACB_A and ACB_B— for three different time steps (**1977- 1988**, **1988T-2019O**, and **2019M-2024M**). These results meet the second of the four key objectives defined in Sub-Chapter 1.2: **Quantify the degree of change in Páramo land cover over time**.

For visualisation purposes, bi-temporal change maps and temporal trajectory maps of the assessment areas were divided into the GUE bounding box a—that includes ACB_G—, and the bounding boxes of ACB_A, ACB_B and RRB —that includes ACB_R— (refer to Figure 3.2). This follows the convention also adopted for the LULC map products from the previous Sub-Chapter and are also generated in QGIS with the exported GEE output.

The results will be organised as follows: First, the binary scheme reduction of the classification will be implemented and the 'MSI equivalent' conversion for all epochs will be performed. Second, the statistics of the Agriculture-Grass class for the study period will be analysed. Third, the bi-temporal change and temporal trajectory maps are presented. Fourth, the time step results are analysed.

### 1.   Binary Scheme reduction

Following the caveats described at the beginning of this Chapter and considering the per-class accuracy obtained for Agriculture-Grass along all epochs in the previous Sub-Chapter, it was decided to reduce the classification to a binary scheme prior to the change analysis. This consisted in merging all the other classes into a new one, Not Agriculture-Grass or Other class.

Figures 4.24 to 4.27 show the binary scheme results for GUE and ACB_G, ACB_A, RRB and ACB_R, and ACB_B, along with a snapshot of the GEE interface for the corresponding PA with all the binary maps stacked on top of each other at 30% opacity. Following the same approach to incorporate the effect of N.D. values, the error bars shown in all the charts represent the *maximum* value the HC area proportion could reach if both the N.D. pixels and the LC pixels, followed the same binary class distribution from all the valid pixels.

Figure 4.24: GUE and ACB_G binary results for the study period.



Figure 4.25: ACB_A binary results for the study period.



Figure 4.26: RRB and ACB_R binary results for the study period.

Figure 4.27: ACB_B binary results for the study period.

This last set of charts and their corresponding stacked maps generated in GEE provide a clearer picture of the evolution of the Agriculture-Grass land cover proportion than their 5-class chart counterpart (Figures 4.16 to 4.21). It confirms the previous observation of some PAs undergoing either an increase of Agriculture-Grass cover (case for GUE, ACB_A, ACB_R) or a fluctuation around a given percentage (case for ACB_G, RRB and ACB_B). Besides this, the corresponding snapshot provides a glimpse of the locations of the Agriculture-Grass pixels throughout the study period and already denoted the agricultural vocation of land for the ACB_A and ACB_B PAs.

There is a final issue to address to provide the best estimation for the Agriculture-Grass area proportion in these areas and it has to do with the *pivot* years. To verify the geometric resolution related variability, the *pivot* years provide two different estimations for comparable dates, for images acquired by different sensors at different geometric resolutions. In principle, these estimations should be similar, but inspecting the estimations for the 1988 pivot (epochs **1988**-**1988T)** and 2019 pivot (epochs **2019O**-**2019M)**, significant differences are found.

To provide a well-round Agriculture-Grass estimation to report, it was decided for all the epochs to be converted to the '*MSI equivalent*'. This was performed by means of a single factor per pivot year, one that multiplying the oldest area proportion yielded the most recent one. Only High Confidence areas were used for the conversion factor estimation.

As an example, for GUE the estimation for the Agriculture-Grass HC area proportion in **2019O** was 36,78% and for **2019M** was 41,02%. Dividing 41,02 over 36,78 results in a factor of 1,12. This factor multiplies all estimations of TEO (TM, ETM+, OLI) sensors to convert them to MSI equivalent—TEO to MSI factor—. Same operation would be performed between the corrected 1988T proportion and the 1988 proportion to obtain the MSS to MSI factor. Table 4.10 summarises all conversion factors employed in the corrections.

Table 4.10: Pivot years MSI equivalent conversion factors.

| PA | TEO to MSI factor | MSS to MSI factor |
|----|-------------------|-------------------|
| GUE | 1,12 | 1,42 |
| ACB_G | 1,26 | 1 |
| ACB_A | 1,04 | 1,59 |
| RRB | 0,88 | 1,33 |
| ACB_R | 1,4 | 1,45 |
| ACB_B | 0,92 | 1,01 |

Due to its large prevalence of N.D. pixels for almost all PAs, epoch **2007E** was not included in the converted area proportions. A comparison between the Agriculture-Grass area proportions before and after the conversion to MSI equivalent are shown in Figure 4.28 for GUE, ACB_G and ACB_A and Figure 4.29 for RRB, ACB_R and ACB_B.



Figure 4.28: GUE, ACB_G and ACB_A High Confidence (HC) area percentage conversion.

Figure 4.29: RRB, ACB_R and ACB_B High Confidence (HC) area percentage conversion.

According to the converted Agriculture-Grass cover area proportions:

❖ GUE baseline was 34,1% in 1977, 40,8% in 1988, 41,0% in 2019 and 41,4% in 2024.

❖ ACB_G baseline was 27,0% in 1977, 45,3% in 1988, 36,5% in 2019 and 31,9% in 2024.

❖ ACB_A baseline was 78,7% in 1977, 74,6% in 1988, 74,2% in 2019 and 74,2% in 2024.

❖ RRB baseline was 9,9% in 1977, 12,3% in 1988, 15,8% in 2019 and 16,3% in 2024.

❖ ACB_R baseline was 55,4% in 1977, 69,0% in 1988, 67,8% in 2019 and 69,2% in 2024.

❖ ACB_B baseline was 69,9% in 1988, 75,7% in 2019 and 79,2% in 2024.

## 2. Agriculture-Grass class statistics

The normality tests for all variables analysed in this study was not conclusive, as some might exhibit a normal behaviour at some epoch and a non-normal behaviour in another. The test for statistical significance of the difference in class area values along epochs yielded different results under the normality/non-normality assumptions. Although both ANOVA one-way test and Kruskal-Wallis's test coincided that at least one of the data samples being compared has a different mean or different rank, their respective post-hoc paired tests (Tukey-HSD and Dunn) flagged different pairs as having a statistically significant difference.

A stand was taken to use the Kruskal-Wallis and Dunn's tests as the final resulting LULC maps are based on a mode aggregation, and it was determined that some of the class areas do not follow a normal distribution. Dunn's test was used as a guide as to choose the best 4 steps for the temporal trajectory between 1988 and 2019, but other factors such as the amount and location of N.D./LC values were also considered. GUE and RRB tests results are summarised on Table 4.11 and Table 4.12, respectively.

Table 4.11: GUE inter-annual statistical testing

| Kruskal-Wallis test | | | Dunn test | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| GUE | H-statistic | 216,31 | reject $H_0$ | 1988 | 1988T | 1991T | 2001T | 2007E | 2011E | 2015O | 2019O | 2019M | 2024M |
| Class 2 / A | | | 1977 | Yes | Yes | No | No | No | Yes | Yes | Yes | Yes | Yes |
| | $\alpha=$ | 0,05 | 1988 | | No | No | No | Yes | No | No | No | Yes | Yes |
| | | | 1988T | | | Yes | Yes | Yes | No | No | Yes | No | No |
| | p-value | 6,31E-41 | 1991T | | | | No | No | No | Yes | No | Yes | Yes |
| | | | 2001T | | | | | No | Yes | Yes | No | Yes | Yes |
| | reject $H_0$ | Yes | 2007E | | | | | | Yes | Yes | No | Yes | Yes |
| | | | 2011E | | | | | | | No | No | No | Yes |
| | | | 2015O | | | | | | | | No | No | No |
| | | | 2019O | | | | | | | | | Yes | Yes |
| | | | 2019M | | | | | | | | | | No |
| | | | 2024M | | | | | | | | | | |

Table 4.12: RRB inter-annual statistical testing

| Kruskal-Wallis test | | | Dunn test | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| RRB | H-statistic | 216,05 | reject $H_0$ | 1988 | 1988T | 1991T | 2001T | 2007E | 2011E | 2015O | 2019O | 2019M | 2024M |
| Class 2 / A | | | 1977 | No | Yes | No | Yes | No | Yes | Yes | Yes | Yes | Yes |
| | $\alpha=$ | 0,05 | 1988 | | No | No | Yes | No | Yes | No | No | Yes | Yes |
| | | | 1988T | | | No | Yes | No | No | No | No | Yes | Yes |
| | p-value | 7,19E-41 | 1991T | | | | Yes | No | Yes | Yes | Yes | Yes | Yes |
| | | | 2001T | | | | | Yes | No | No | Yes | No | No |
| | reject $H_0$ | Yes | 2007E | | | | | | Yes | Yes | No | Yes | Yes |
| | | | 2011E | | | | | | | No | No | No | No |
| | | | 2015O | | | | | | | | No | No | Yes |
| | | | 2019O | | | | | | | | | No | No |
| | | | 2019M | | | | | | | | | | No |
| | | | 2024M | | | | | | | | | | |

### 3.  Change Maps

The three overall time steps were defined as **MSS 77-88**, **TEO 88-19** and **MSI 19-24**. Binary change detection was performed to the **MSS 77-88** and **MSI 19-24** time steps, while a 4-epoch temporal trajectory was analysed for **TEO 88-19**. For GUE the 4 epochs selected for the temporal trajectory were **1988T**-**2001T**-**2015O**-**2019O**, and for RRB **1988T**-**2001T**-**2011E**-**2019O**. The binary HC map for the starting epoch and the resulting change map for each time step are shown together. For GUE the maps are shown in Figures 4.30 and 4.31, and for RRB are shown in Figures 4.32 to 4.33.

Figure 4.30: GUE 1977 binary HC map and 1977-1988 change results.



Figure 4.31: GUE 1988T binary HC map and 1988T-2019O change results.

Figure 4.32: RRB, ACB(_A,_B,_R) 1977 binary HC map and 1977-1988 change results.



Figure 4.33: RRB, ACB(_A,_B,_R) 1988T binary HC map and 1988T-2019O change results

The remaining binary HC and change maps (**2019O** and **MSI 19-24**) are presented in Appendix B.4, as no significant change is estimated in this last time step.

## 4. Time step results

Change classes were normalised over the *total valid* area of the bounding box scene, as it was done for the PA LULC results, by the areas shown in Table 4.7. The joint No Data (N.D.) and Low Confidence pixels from all the epochs included in the time step are merged into ND/LC class. It should be noted that, being a pixel-to-pixel comparison, this class groups all pixels that have no valid counterpart to be compared, either because of masking or because its counterpart was a low confidence one. This means that only HC vs. HC pixels are compared, resulting in an optimistic lower boundary for change.

Following the change typology described in Table 3.8 and Table 3.9, all time steps results are presented by Protected Area in Figures 4.34 to 4.36. It can be stated that the change analysis has solidified some of the overall trends adverted in the previous Sub-Chapter, as all PAs exhibited some level of **increment** in the stable Agriculture-Grass proportions. This increase is nevertheless, not a sided one, as the gain-loss of Agriculture-Grass has been dynamic, except for the ACB_G 77-88 gain (~20%) which was one of the highest recorded.



Figure 4.34: GUE, ACB_G time step change results.

Figure 4.35: RRB, ACB_R time step change results.



Figure 4.36: ACB_A, ACB_B time step change results.

The following highest Agriculture-Grass gains (~11-12%) mostly belong to the first time step, with progressive reductions in the following steps. This indicates that although Agriculture-Grass class and its stable areas has been increasing, its gain rate has been steadily **decelerating**. Proof of this is that the final time step has very similar

one-digit percentages for Agriculture-Grass gain-loss values and that the difference of the area proportion for the Agriculture-Grass class were not statistically significant.

Cyclical areas are not very prevalent and represent roughly only one percentual point of the change proportion. Stable Other areas have had different trends, increasing for some of the PAs, and decreasing for others. Being such a heterogeneous aggregation of classes prevents extracting any more conclusions from this stable area proportion.

An interesting insight is offered when only the change classes, namely Agriculture-Grass gain, loss and cyclical are displayed on a map. This allows for the change areas to stand on top of the context of the base map, allowing also for the stable areas to be appreciated. To provide the reader with what the author considers as the most meaningful maps from the Thesis, two GEE generated maps are shown below. Both are a combination of two different snapshots from different scripts manually merged in a single map. The first one (Figure 4.37) corresponds to the temporal trajectory of the second time step (1988T to 2019O) spanning 31 years. The second one (Figure 4.38) correspond to the superposition of all the time step changes (most recent changes on top) spanning the full 47 years of the study period.



Figure 4.37: GEE snapshots of AOI TEO 88-19 temporal trajectory (only change classes).

1977-2024

**Change Detection**
- [1->2] Agriculture-Grass Gain
- [<->] Cyclical
- [2->1] Agriculture-Grass Loss

Figure 4.38: GEE snapshots of AOI with change classes for the whole study period.

## 4.3.   Agriculture Disturbance Index

This Sub-Chapter will calculate the Agriculture Disturbance Index—the Index— detailed in Sub-Chapter 3.6 for different areas. These results meet the third and fourth key objectives defined in Sub-Chapter 1.2: **Develop a Disturbance Index that encompasses spatial-temporal changes to appropriately determine Páramo quality** and **Identify the most pristine and most disturbed Páramo Complex and Páramo areas.**

According to the mathematical development, the Index is calculated for GUE using Equation (3.11) as an example of its application. All assessment Protected Areas Index results are summarised in Table 4.13.

$$I_{GUE} = A_{77(M)}(1 + \Delta A_{77\to88})14,92 + A_{88(M)}(1 + Cyc_{88\to19} + \Delta A_{88\to19})68,51 + A_{19M}(1 + \Delta A_{19\to24})16,57$$

$$I_{GUE} = 0,341(1 + 0,0402)14,92 + 0,408(1 + 0,011 + 0,0051)68,51 + 0,41(1 - 1,001)16,57$$

$$I_{GUE} = (0,3549)14,92 + (0,4151)68,51 + (0,4098)16,57$$

$$I_{GUE} = \mathbf{40,52}$$

According to the suggested interpretation: '*At least 40,52 percent of the government-delimited area for the Páramo de Guerrero Complex has been regularly under a LULC of Agriculture-Grass for the last 47 years, and at present could be greater*'.

Table 4.13: Agriculture Disturbance Index and variables for the assessment Protected Areas

| PA | $A_{77(M)}$ | $\Delta A_{77\to88}$ | $A_{88(M)}$ | $\Delta A_{88\to19}$ | $Cyc_{88\to19}$ | $A_{19M}$ | $\Delta A_{19\to24}$ | Index |
|---|---|---|---|---|---|---|---|---|
| GUE | 0,341 | 0,0402 | 0,408 | 0,0051 | 0,0110 | 0,410 | -0,0010 | **40,52** |
| ACB_G | 0,270 | 0,1395 | 0,453 | -0,0592 | 0,0118 | 0,365 | -0,0395 | **39,96** |
| ACB_A | 0,787 | -0,0161 | 0,746 | 0,0069 | 0,0092 | 0,742 | 0,0011 | **75,82** |
| RRB | 0,099 | 0,0086 | 0,123 | 0,0281 | 0,0112 | 0,158 | 0,0051 | **12,85** |
| ACB_R | 0,554 | 0,0161 | 0,690 | 0,0856 | 0,0151 | 0,678 | 0,0134 | **71,84** |
| ACB_B | - | 0,0000 | 0,698 | 0,0492 | 0,0160 | 0,757 | 0,0247 | **75,02** |

The same Index development was performed for the subdivision of the GUE, ACB_G and ACB_A assessment PAs for the rural districts (*Veredas*) of municipalities with over 20 Ha of jurisdiction, and the results are presented in Table 4.14.

Table 4.14: Agriculture Disturbance Index for Protected Areas Municipality rural Districts.

| PA | REG. | MUNICIPALITY | VEREDA (DISTRICT) | DISTRICT AREA [Ha] | DISTRICT PA [Ha] | Index |
|---|---|---|---|---|---|---|
| ACB_A | CUN | CHOCONTÁ | HATO FIERO | 1035,20 | 252,31 | **64,47** |
| ACB_A | CUN | CHOCONTÁ | SANTA BARBARA | 1238,93 | 231,89 | **71,41** |

| PA | REG. | MUNICIPALITY | VEREDA (DISTRICT) | DISTRICT AREA [Ha] | DISTRICT PA [Ha] | Index |
|---|---|---|---|---|---|---|
| ACB_A | CUN | CHOCONTÁ | CALIENTE | 680,62 | 59,44 | 69,20 |
| ACB_A | CUN | CHOCONTÁ | PUEBLO VIEJO | 1543,02 | 44,92 | 81,60 |
| ACB_A | CUN | CUCUNUBÁ | JUAITOQUE | 351,06 | 104,56 | 76,64 |
| ACB_A | CUN | CUCUNUBÁ | LA TOMA | 504,01 | 64,19 | 77,64 |
| ACB_A | CUN | CUCUNUBÁ | HATO DE ROJAS | 634,81 | 41,74 | 80,18 |
| ACB_A | CUN | CUCUNUBÁ | ATRAVIESAS | 462,24 | 39,42 | 76,98 |
| ACB_A | CUN | LENGUAZAQUE | CHIRVANEQUE | 959,82 | 49,90 | 80,03 |
| ACB_A | CUN | LENGUAZAQUE | FARACÍA - PANTANITOS | 1167,13 | 45,43 | 76,34 |
| ACB_A | CUN | SUESCA | HATILLO | 1672,03 | 423,94 | 80,86 |
| ACB_A | CUN | SUESCA | TAUSAQUIRA | 1089,03 | 227,25 | 37,28 |
| ACB_A | CUN | SUESCA | CUAYÁ | 1169,01 | 185,50 | 62,34 |
| ACB_A | CUN | SUESCA | OVEJERAS | 1653,59 | 43,47 | 46,70 |
| ACB_G | CUN | GACHANCIPÁ | SAN JOSÉ | 2045,54 | 44,96 | 19,69 |
| ACB_G | CUN | NEMOCÓN | AGUA CLARA | 1150,75 | 75,88 | 0,43 |
| ACB_G | CUN | SUESCA | PIEDRAS LARGAS | 1396,60 | 170,48 | 37,54 |
| ACB_G | CUN | SUESCA | CUAYÁ | 1169,01 | 163,97 | 72,65 |
| GUE | BOY | BUENAVISTA | PISMAL | 1460,24 | 32,60 | 27,01 |
| GUE | CUN | CARMEN DE CARUPA | CASA BLANCA | 1969,05 | 1965,85 | 17,17 |
| GUE | CUN | CARMEN DE CARUPA | SALINAS | 2174,91 | 1794,95 | 15,49 |
| GUE | CUN | CARMEN DE CARUPA | MORTINO | 1573,13 | 1197,40 | 59,56 |
| GUE | CUN | CARMEN DE CARUPA | LA ESPERANZA | 2145,03 | 1043,20 | 20,24 |
| GUE | CUN | CARMEN DE CARUPA | LLANO GRANDE | 1734,46 | 540,21 | 22,36 |
| GUE | CUN | CARMEN DE CARUPA | SANTUARIO | 811,77 | 490,91 | 55,22 |
| GUE | CUN | CARMEN DE CARUPA | CHARQUIRÁ | 1546,27 | 423,54 | 63,39 |
| GUE | CUN | CARMEN DE CARUPA | ALTO DE MESA | 1249,30 | 368,22 | 68,06 |
| GUE | CUN | CARMEN DE CARUPA | TUDELA | 1160,43 | 357,73 | 57,27 |
| GUE | CUN | CARMEN DE CARUPA | PAPAYO | 842,96 | 353,74 | 89,63 |
| GUE | CUN | CARMEN DE CARUPA | SANTA DORA | 839,20 | 333,17 | 48,03 |
| GUE | CUN | CARMEN DE CARUPA | BUGRE | 2267,37 | 237,42 | 11,01 |
| GUE | CUN | CARMEN DE CARUPA | SALITRE | 655,56 | 204,38 | 30,66 |
| GUE | CUN | CARMEN DE CARUPA | SAN JOSÉ | 1153,36 | 134,24 | 3,78 |
| GUE | CUN | CARMEN DE CARUPA | EL RATO | 1306,31 | 130,26 | 10,05 |
| GUE | CUN | CARMEN DE CARUPA | ALISAL | 978,79 | 120,67 | 90,41 |
| GUE | CUN | CARMEN DE CARUPA | SAN AGUSTÍN | 571,98 | 118,79 | 31,27 |
| GUE | CUN | CARMEN DE CARUPA | PERQUIRA | 1920,92 | 98,41 | 3,77 |
| GUE | CUN | CARMEN DE CARUPA | CHEGUA | 900,10 | 91,65 | 18,34 |
| GUE | CUN | CARMEN DE CARUPA | SAN ANTONIO | 867,42 | 35,83 | 1,05 |
| GUE | CUN | COGUA | PARAMO ALTO | 2195,33 | 2046,15 | 74,19 |
| GUE | CUN | COGUA | QUEBRADA HONDA | 2198,45 | 630,89 | 40,98 |
| GUE | CUN | COGUA | RODAMONTAL | 2313,77 | 369,63 | 18,78 |
| GUE | CUN | COGUA | CARDONAL | 932,65 | 210,60 | 15,05 |
| GUE | CUN | COGUA | CASA BLANCA | 1149,60 | 101,60 | 14,71 |
| GUE | CUN | COGUA | PATASICA | 1026,55 | 36,51 | 2,05 |

| PA | REG. | MUNICIPALITY | VEREDA (DISTRICT) | DISTRICT AREA [Ha] | DISTRICT PA [Ha] | Index |
|---|---|---|---|---|---|---|
| GUE | CUN | CUCUNUBÁ | PEÑAS | 739,49 | 94,16 | 19,30 |
| GUE | CUN | CUCUNUBÁ | PUEBLO VIEJO | 1245,33 | 58,25 | 20,33 |
| GUE | CUN | FÚQUENE | CENTRO | 1133,52 | 58,14 | 25,11 |
| GUE | CUN | FÚQUENE | CHINZAUQUE | 1215,26 | 38,11 | 31,68 |
| GUE | CUN | PACHO | CANADA | 1582,35 | 886,03 | 30,37 |
| GUE | CUN | PACHO | NEGRETE | 2067,38 | 669,88 | 18,38 |
| GUE | CUN | PACHO | EL BOSQUE | 1877,88 | 608,44 | 18,97 |
| GUE | CUN | PACHO | EL CABRERO | 1848,12 | 332,59 | 27,83 |
| GUE | CUN | PACHO | CERRO NEGRO | 1306,03 | 67,44 | 3,01 |
| GUE | CUN | SAN CAYETANO | MORTIÑO | 596,62 | 566,73 | 60,60 |
| GUE | CUN | SAN CAYETANO | QUEBRADAS | 1236,45 | 513,00 | 20,09 |
| GUE | CUN | SAN CAYETANO | CARDONAL | 577,25 | 486,14 | 34,78 |
| GUE | CUN | SAN CAYETANO | BOCA DE MONTE | 867,04 | 46,91 | 59,74 |
| GUE | CUN | SIMIJACA | TAQUIRA | 2452,46 | 69,63 | 14,09 |
| GUE | CUN | SIMIJACA | DON LOPE | 1076,37 | 69,31 | 3,75 |
| GUE | CUN | SUBACHOQUE | EL GUAMAL | 3246,13 | 2145,50 | 48,96 |
| GUE | CUN | SUBACHOQUE | PANTANO DE ARCE | 1522,49 | 1168,24 | 61,92 |
| GUE | CUN | SUBACHOQUE | EL PARAMO | 2404,45 | 685,93 | 59,08 |
| GUE | CUN | SUBACHOQUE | RINCON SANTO | 1941,07 | 584,90 | 27,59 |
| GUE | CUN | SUBACHOQUE | LA UNION | 769,14 | 187,70 | 51,66 |
| GUE | CUN | SUBACHOQUE | CASCAJAL | 1168,93 | 107,74 | 91,10 |
| GUE | CUN | SUBACHOQUE | EL TOBAL | 717,81 | 77,01 | 77,15 |
| GUE | CUN | SUBACHOQUE | CANICA ALTA | 259,09 | 56,08 | 1,39 |
| GUE | CUN | SUBACHOQUE | ALTANIA | 894,57 | 26,50 | 79,91 |
| GUE | CUN | SUPATÁ | SANTA BARBARA | 3143,47 | 290,26 | 5,59 |
| GUE | CUN | SUPATÁ | LAJAS | 1350,02 | 69,57 | 12,97 |
| GUE | CUN | SUSA | NUTRIAS | 1313,60 | 822,65 | 46,76 |
| GUE | CUN | SUSA | EL TABLON | 934,99 | 743,51 | 49,64 |
| GUE | CUN | SUSA | PAUNITA | 1292,53 | 365,86 | 57,62 |
| GUE | CUN | SUSA | MATA REDONDA | 571,82 | 120,14 | 56,80 |
| GUE | CUN | SUSA | APOSENTOS | 863,44 | 84,76 | 43,88 |
| GUE | CUN | SUSA | LA FRAGUA | 749,87 | 68,03 | 5,10 |
| GUE | CUN | SUSA | TIMINGUITA | 449,90 | 30,83 | 32,41 |
| GUE | CUN | SUSA | CASCADAS | 686,92 | 28,02 | 6,30 |
| GUE | CUN | SUTATAUSA | PEÑAS DE CAJON | 281,95 | 75,26 | 8,64 |
| GUE | CUN | SUTATAUSA | PALACIO | 955,80 | 70,42 | 8,31 |
| GUE | CUN | SUTATAUSA | NOVOA | 509,00 | 66,98 | 5,87 |
| GUE | CUN | SUTATAUSA | CONCUBITA | 441,66 | 54,51 | 8,98 |
| GUE | CUN | SUTATAUSA | PEÑAS DE BOQUERON | 500,25 | 39,04 | 5,64 |
| GUE | CUN | SUTATAUSA | MOCHILA | 276,14 | 23,14 | 30,74 |
| GUE | CUN | TABIO | RIO FRI OOCCIDENTAL | 2533,21 | 229,27 | 50,35 |
| GUE | CUN | TABIO | EL SALITRE | 689,34 | 48,79 | 14,00 |
| GUE | CUN | TAUSA | EL SALITRE | 2766,08 | 2381,37 | 46,62 |

| PA | REG. | MUNICIPALITY | VEREDA (DISTRICT) | DISTRICT AREA [Ha] | DISTRICT PA [Ha] | Index |
|---|---|---|---|---|---|---|
| GUE | CUN | TAUSA | SAN ANTONIO | 1917,45 | 1917,45 | 55,52 |
| GUE | CUN | TAUSA | SABANEQUE | 1646,78 | 1646,77 | 36,23 |
| GUE | CUN | TAUSA | LLANO GRANDE | 1767,69 | 849,70 | 5,99 |
| GUE | CUN | TAUSA | PARAMO ALTO | 1572,11 | 701,64 | 57,43 |
| GUE | CUN | TAUSA | LAGUNITAS | 1623,78 | 654,37 | 61,79 |
| GUE | CUN | TAUSA | LA FLORIDA | 1998,29 | 550,39 | 11,96 |
| GUE | CUN | TAUSA | PARAMO BAJO | 1354,64 | 443,67 | 24,53 |
| GUE | CUN | TAUSA | LOS PINOS | 350,68 | 291,80 | 38,35 |
| GUE | CUN | TAUSA | PUEBLO VIEJO | 766,73 | 183,48 | 5,46 |
| GUE | CUN | TAUSA | PAJARITO | 898,63 | 89,70 | 23,49 |
| GUE | CUN | TAUSA | CHORRILLO | 385,07 | 52,32 | 16,24 |
| GUE | CUN | TAUSA | RASGATA ALTO | 814,33 | 27,73 | 7,73 |
| GUE | CUN | UBATÉ | VOLCAN | 1587,86 | 190,60 | 91,82 |
| GUE | CUN | UBATÉ | SOAGA | 830,36 | 76,91 | 48,31 |
| GUE | CUN | ZIPAQUIRÁ | PARAMO DE GUERRERO | 3384,37 | 3304,44 | 62,08 |
| GUE | CUN | ZIPAQUIRÁ | VENTALARGA | 2270,56 | 1775,34 | 31,04 |
| GUE | CUN | ZIPAQUIRÁ | EL EMPALIZADO | 1721,92 | 1207,68 | 36,66 |
| GUE | CUN | ZIPAQUIRÁ | SAN ISIDRO | 1588,23 | 233,34 | 39,01 |
| GUE | CUN | ZIPAQUIRÁ | RIO FRIO | 1044,39 | 128,82 | 13,00 |
| GUE | CUN | ZIPAQUIRÁ | SAN ANTONIO | 638,69 | 86,91 | 9,75 |

Considering the top 10 rural districts (*veredas*) with the highest areas [~1700-3300 Ha], it is evident there is a great dissimilarity in the persistence of agricultural disturbances, as shown in Table 4.15.

Table 4.15: Top 10 highest PA districts results

| VEREDA (DISTRICT) | Index |
|---|---|
| PARAMO DE GUERRERO | 62,08 |
| EL SALITRE | 46,62 |
| EL GUAMAL | 48,96 |
| PARAMO ALTO | 74,19 |
| CASA BLANCA | 17,17 |
| SAN ANTONIO | 55,52 |
| SALINAS | 15,49 |
| VENTALARGA | 31,04 |
| SABANEQUE | 36,23 |
| EL EMPALIZADO | 36,66 |

Finally, the top 30 (least disturbed) and bottom 30 (most disturbed) ranked rural districts (*veredas*) are presented in Table 4.16.

Table 4.16: Top 30 best and worst ranked rural districts results

| VEREDA (DISTRICT) | Index | VEREDA (DISTRICT) | Index |
|---|---|---|---|
| AGUA CLARA | 0,43 | VOLCAN | 91,82 |
| SAN ANTONIO | 1,05 | CASCAJAL | 91,10 |
| CANICA ALTA | 1,39 | ALISAL | 90,41 |
| PATASICA | 2,05 | PAPAYO | 89,63 |
| CERRO NEGRO | 3,01 | PUEBLO VIEJO | 81,60 |
| DON LOPE | 3,75 | HATILLO | 80,86 |
| PERQUIRA | 3,77 | HATO DE ROJAS | 80,18 |
| SAN JOSÉ | 3,78 | CHIRVANEQUE | 80,03 |
| LA FRAGUA | 5,10 | ALTANIA | 79,91 |
| PUEBLO VIEJO | 5,46 | LA TOMA | 77,64 |
| SANTA BARBARA | 5,59 | EL TOBAL | 77,15 |
| PEÑAS DE BOQUERON | 5,64 | ATRAVIESAS | 76,98 |
| NOVOA | 5,87 | JUAITOQUE | 76,64 |
| LLANO GRANDE | 5,99 | FARACÍA - PANTANITOS | 76,34 |
| CASCADAS | 6,30 | PARAMO ALTO | 74,19 |
| RASGATA ALTO | 7,73 | CUAYÁ | 72,65 |
| PALACIO | 8,31 | SANTA BARBARA | 71,41 |
| PEÑAS DE CAJON | 8,64 | CALIENTE | 69,20 |
| CONCUBITA | 8,98 | ALTO DE MESA | 68,06 |
| SAN ANTONIO | 9,75 | HATO FIERO | 64,47 |
| EL RATO | 10,05 | CHARQUIRÁ | 63,39 |
| BUGRE | 11,01 | CUAYÁ | 62,34 |
| LA FLORIDA | 11,96 | PARAMO DE GUERRERO | 62,08 |
| LAJAS | 12,97 | PANTANO DE ARCE | 61,92 |
| RIO FRIO | 13,00 | LAGUNITAS | 61,79 |
| EL SALITRE | 14,00 | MORTIÑO | 60,60 |
| TAQUIRA | 14,09 | BOCA DE MONTE | 59,74 |
| CASA BLANCA | 14,71 | MORTINO | 59,56 |
| CARDONAL | 15,05 | EL PARAMO | 59,08 |
| SALINAS | 15,49 | PAUNITA | 57,62 |

# 5  Discussion

This work aims to contribute to a growing body of research focused on the monitoring of tropical high-elevation ecosystems in Colombia, and particularly to the monitoring of Colombian Páramos.

Research agrees on an evident Páramo ecosystem decline, mostly due to anthropogenic causes. The quantification reported and trends described for this decline vary depending on the research scope. As mentioned throughout the development of the state of the art and methods, the methodology followed in classified remote sensing images is essential to the results reported.

The present Chapter compares both the methodological approach, and the results obtained, with comparisons to similar research. The following Sub-Chapters will discuss these aspects mirroring Chapter 4 structure.

## 5.1.  Land Cover

**Caveats of Pre-Processing and class schema definition**

The pre-processing methodology adopted for this Thesis aimed at favouring single observation data over time aggregation compositions. Single observation data examples, where authors refer to the specific scenes processed [155] are not common. Scene composites were limited to a minimum and only performed to fill no data values or mosaic different scene paths. This is usually not the approach taken for cloud-prone areas, where monthly, annual or even biannual compositions are preferred [28], [93], [156].

No topographic and no atmospheric correction was performed on Landsat MSS images, and geometric corrections (shifts, co-registration) were performed in between MSS images. Several other possible approaches could have been taken with respect to this imagery, possibly affecting the final classification results.

Finally, the data-centred class definition and minimal classes approach, constrained by the limitations of the oldest datasets is a pragmatic approach, were not found in the literature reviewed. Most of the research consulted opts for LCCS [84], the Anderson classification system [86], CLC [87] or derived local variations, sometimes without specifying the reason for the modifications. There is a need to establish consensus regarding which system is better suited for the Páramo context.

To the authors knowledge, no other research has focused on trying to harness *all* the historical satellite remote sensing datasets for such a long study period, minimising the use of pixel compositing and making use of the *least* number of classes to provide a descriptive enough schema to identify the *main* LULC types for the Páramo context.

**Sample areas and spectral characterisation**

A comprehensive guide on the sampling process was included in the present work, along with the extensive references and inclusion of the scripts, to support the open-access spirit of the Thesis and allow for replicability, improvement and re-purposing of the methodology. Recent GEE centred research allows for training and testing datasets to be accessed [93], and the actual trends aim for this to become standard practice. The constant training across all-epochs datasets made available are encouraged to be consulted, corrected and further expanded for future research.

The sample areas and spectral characterisation aims to contribute to the definition of the Páramo spectral signature, although in the present work this was not possible to be isolated. However, the reported Shrubland-Wetland class presents similarities to the ones in the Supplementary Table 1A of Murad et al.[28].

**Classification algorithm and accuracy assessment**

The RF iterative ensemble proved effective in terms of per-class User's and Producer's accuracy improvement and to capture the model's variability. Although all classifications were trained with and validated against the same reference datasets, the iterations led to significantly different results. This behaviour is also reported in the literature [156] and highlights the need for observance of both the datasets and the algorithm selected for classification. Other research has obtained comparable or even better OAs with SVM-based classification [28], [93], [157].

The per-class accuracy assessment revealed that there is a progression on the User's and Producer's accuracy ranking between the 5 classes defined. From Water being the most accurate for both metrics, followed by Agriculture-Grass, Forest, Shrubland-Wetland and Bare-Rock-Built, as evidenced by the scatter plots produced. This behavior corresponds to equivalent classes for the main world-coverage land cover maps [82]. The per-class accuracy and the search for the most accurate description of Páramo PAs evolution became the reason to choose the Agriculture-Grass class as the focus in the binary system.

Finally, it must be noted that this assessment did not use external or complementary validation data and relied on Google Earth high-resolution imagery to support the sample areas interpretation. This highlights the need for regional and context-specific data [158], and future work is encouraged to pursue validation against authoritative and regional LULC products where available.

**Agricultural LULC within Páramo Protected Areas**

The assessment revealed that between 1977 and 2024, Agriculture-Grass areas increased in most of the protected areas with some of them being historically dominated by this land cover: GUE: **34,1%** to **41,4%**; RRB: **9,9%** to **16,3%**; ACB_A: **78,7%** to **74,2%**; ACB_B: **69,9%** (year 1988) to **79,2%**; ACB_R: **55,4%** to **69,2%**; ACB_G: **27,0%** to **31,9%.**

This is consistent with the research cited at the end of the Páramo introductory Sub-Chapter (page 15), and with the average of 18% transformed areas in Eastern Andes Range Páramos, the worst Range of the Andes for Colombia in this aspect [159].

The results highlight differences within areas belonging to the same Páramo Complex, as not all are suffering disturbances related to agriculture. RRB for instance, has a much lower agriculture prevalence and growth, but it is still highlighted as a threatened Páramo by other disturbances, such as forestry and mining [28], [129], [160].

The results denote that the Altiplano Cundiboyacense Páramo Complex (ACB) areas have a long-standing agricultural vocation and there is no evidence for significant land cover of Shrubland-Wetland class, that contains Páramo. This questions the effectiveness of hard-boundary protection when the historical presence of inhabitants and their activities is disregarded. An interesting discussion on this subject can be found in [161].

## 5.2. Land Cover Change

The schema reduction resulting from the focus on Agriculture-Grass class accuracy and the High Confidence areas defined by the pixel agreement, were leveraged as means to achieve the most accurate description of change. Overall trends appreciated in the epoch-to-epoch LULC were confirmed with the bi-temporal and temporal trajectory analysis, showing a steady increase of Agriculture-Class stable areas.

The Agriculture-Grass gain/loss behaviour was dynamic, with similar patterns to those observed for GUE in [28], with that being said the cited research only considers the *Páramo de Guerrero* largest polygon delimited. This means that the evolution of the land cover cannot be reduced solely to an aggressive expansion of agriculture over the other classes. In fact, some net agriculture changes found were negative, adding to the dynamism that the land cover evolution can present.

Among this dynamic behaviour, the results show a clear trend of deceleration of change. This means that although Agriculture-Grass class has mostly grown in most of the assessment areas, there has been a progressive reduction on the rates of

agriculture gain and loss. Further research is needed to determine whether this deceleration is correlated with economical or public-policy factors.

Finally, it is important to highlight that Páramo is an ecosystem that does not exist in isolation, but it is interrelated with the high Andean forests and even with agricultural practices, to some extent. An interesting insight on this matter is presented by the avian biodiversity loss in Páramo and the effect of deforestation on surrounding areas [162]. There is not a complete incompatibility between human activities and natural ecosystems, and this is one of the reasons a multidisciplinary and holistic approach for ecosystem monitoring and conservation is needed.

## 5.3.  Disturbance Index

Being a novel and context-specific proposition, there are no direct comparisons for the Agricultural Disturbance Index. However, the use of indices is complementary to the sole area or area proportion reporting. Proof of that is the index rankings presented for the rural districts of municipalities with jurisdiction on GUE, that show huge disparities in terms of disturbance within the same protected area. These results are also easily interpretable and accessible for a wide range of audiences.

Other spatial indexes that were not implemented in the work conducted are recommended for complementing the evolution of the protected areas such as Fragmentation, Simpson, Shannon, Patch size, Isolation, among others [160].

# 6   Conclusions and Future Work

Páramo is a unique, beautiful and fragile Ecosystem, pivotal to water and carbon regulation. Conservation efforts for this ecosystem in Colombia have focused on hard-lining boundaries for Protected Areas, in response to Páramo decline due to agriculture, grazing and mining. Remote Sensing monitoring of this ecosystem is cost-effective, operational and much needed. However, it is essential to establish consensus regarding the most appropriate and effective classification system for the Páramo context. Moreover, further work needs to be undertaken to pursue a definition of the Páramo spectral signature, to be able to effectively and accurately map the true extent of the Páramo.

This work can conclude that from a land-cover perspective, Protected Areas in the central Eastern Andes have undergone an expansion of Agriculture-Grass for the past 47 years. This expansion behaviour is different among the Protected Areas included in the assessment, with some of them having a dominant agricultural vocation at the start of the study period. The change analysis showed that stable Agriculture-Grass areas also steadily increased, with a dynamic gain/loss behaviour, and a clear decelerating trend. However, further research is needed to determine whether this deceleration is correlated with economical or public-policy factors.

The main contribution of this work, beyond the results reported, is the methodological approach focused on data and accuracy driven choices, detailed processing steps, open-access resources and replicability. Expansion, correction and re-purposing of the methodology is actively encouraged. The author hopes that this work will offer a key contribution and baseline for further multidisciplinary research on ecosystem monitoring and conservation that is so vital for the future of Páramo.

# Bibliography

[1] R. Hofstede and L. Llambí, 'Plant Diversity in Páramo—Neotropical High Mountain Humid Grasslands', in *Encyclopedia of the World's Biomes*, vol. 1, Elsevier, 2020, pp. 362–372. doi: 10.1016/B978-0-12-409548-9.11858-5.

[2] J. L. Luteyn, *Páramos: a checklist of plant diversity, geographical distribution and botanical literature*. in Memoirs of the New York Botanical Garden, no. 84. The New York Botanical Garden Press, 1999. Accessed: Dec. 13, 2023. [Online]. Available: https://issuu.com/jpintoz/docs/1999_luteyn_pmos-checklpldivgeodistbotlit_memnybg8

[3] O. Zúñiga-Escobar, A. Uribe V, A. M. Torres-González, R. Cuero-Guependo, and J. A. Peña-Óspina, 'Assessment of the impact of anthropic activities on carbon storage in soils of high montane ecosystems in Colombia', *Agronomía Colombiana*, vol. 31, pp. 112–119, 2013.

[4] R. Hofstede, 'Health state of Páramos: an effort to correlate science and practice', *Lyonia*, vol. 6, no. 1, 2004.

[5] W. Buytaert *et al.*, 'Human impact on the hydrology of the Andean páramos', *Earth-Science Reviews*, vol. 79, no. 1, pp. 53–72, Nov. 2006, doi: 10.1016/j.earscirev.2006.06.002.

[6] F. Cárdenas Tamara *et al.*, *El Páramo un ecosistema de alta montaña*. Santa Fé de Bogotá: Fundación Ecosistemas Andinos, Gobernación de Boyacá, 1995. Accessed: Dec. 14, 2023. [Online]. Available: https://babel.banrepcultural.org/digital/collection/p17054coll10/id/2819/

[7] G. Peyre, D. Osorio, R. François, and F. Anthelme, 'Mapping the páramo land-cover in the Northern Andes', *International Journal of Remote Sensing*, vol. 42, no. 20, pp. 7777–7797, Oct. 2021, doi: 10.1080/01431161.2021.1964709.

[8] J. Cuatrecasas, 'Aspectos de la vegetación natural de Colombia', *Revista de la Academia Colombiana de Ciencias Exactas, Físicas y Naturales*, vol. 10, no. 40, pp. 221–264, 1958, doi: 10.18257/raccefyn.570.

[9] A. M. Cleef, 'The vegetation of the páramos of the Colombian Cordillera Oriental', *Diss. Bot.*, vol. 61, pp. 1–320, Jan. 1981.

[10] T. Van der Hammen *et al.*, Eds., *Studies on Tropical Andean Ecosystems/Estudios De Ecosistemas Tropandinos*, vol. 1/2/3/4/5/6/7, 7 vols. Berlin-Stuttgart: J. Cramer (Borntraeger), 1983to2008.

[11] T. Van der Hammen, J. D. Pabón Caicedo, H. J. Gutiérrez Rey, and J. C. Alarcón Hincapié, 'El cambio Global y los Ecosistemas de Alta Montaña de Colombia', *Páramos y Ecosistemas Alto Andinos de Colombia en condiciones HotSpot & Global Climatic Tensor*, vol. 3, pp. 163–209, 2002.

[12] J. C. Alarcón Hincapié *et al.*, 'Transformaciones de las coberturas vegetales en los páramos de Colombia: un punto de partida para la evaluación ambiental', *Transformación y cambio en el uso del suelo en los Páramos de Colombia en las últimas décadas*, vol. 4, pp. 210–222, 2002.

[13] R. Hofstede, 'The effects of grazing and burning on soil and plant nutrient concentrations in Colombian Páramo grasslands', *Plant and Soil*, vol. 173, no. 1, pp. 111–132, 1995, doi: 10.1007/BF00155524.

[14] H. J. Gutiérrez Rey, 'Aproximación a un modelo para la evaluación de la vulnerabilidad de las coberturas vegetales de Colombia ante un posible cambio climático utilizando Sistemas de Información Geográfica SIG con énfasis en la vulnerabilidad de las coberturas nival y de páramo de Colombia', vol. 5, pp. 335–377, 2002.

[15] M. F. Cárdenas Agudelo, 'Ecohydrology of paramos in Colombia: vulnerability to climate change and land use', Universidad Nacional de Colombia, 2016. Accessed: Dec. 14, 2023. [Online]. Available: https://repositorio.unal.edu.co/handle/unal/59136

[16] IAvH *et al.*, *Atlas de Páramos de Colombia*, 1st ed. Instituto de Investigación de recursos biológicos Alexander von Humboldt, 2007.

[17] IAvH, 'Cartografía 2013 de los Páramos de Colombia: Diversidad, territorio e historia'. Accessed: Dec. 16, 2023. [Online]. Available: http://www.humboldt.org.co/es/actualidad/item/109-nueva-cartografia-de-los-paramos-de-colombia-diversidad-territorio-e-historia

[18] IDEAM, 'Paramos Delimitados', Catálogo de Mapas Sistema de Información Ambiental de Colombia - SIAC. Instituto de Hidrología, Meteorología y Estudios Ambientales de Colombia. Accessed: Dec. 15, 2023. [Online]. Available: http://www.siac.gov.co/catalogo-de-mapas

[19]   C. Buitrago Gómez, J. B. Carrasco Leal, G. Moreno Pinzón, and E. Pardo Pardo, 'Cambio, uso y transforamación de las actividades económicas de los ecosistemas de Alta Montaña', *Transformación y cambio en el uso del suelo en los Páramos de Colombia en las últimas décadas*, pp. 274–295, 2002.

[20]   C. N. de la R. Colombia, *Art. 202 - Ley 1450 Por la cual se expide el Plan Nacional de Desarrollo 2011 - 2014.* 2011. [Online]. Available: https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=43101

[21]   MADS, *Res. No. 886 del 18/05/2018 Por la cual se adoptan los lineamientos para la zonificación y régimen de usos en las áreas de páramos delimitados y se establecen las directrices para diseñar, capacitar y poner en marcha programas de sustitución y reconversión de las actividades.* 2018. [Online]. Available: https://www.minambiente.gov.co/documento-normativa/resolucion-0886-de-2018/

[22]   C. N. de la R. Colombia, *Ley 1930 de 2018 Por medio de la cual se dictan disposiciones para la gestión integral de los Páramos en Colombia.* 2018. [Online]. Available: https://www.minambiente.gov.co/documento-normativa/ley-1930-de-2018-julio-27/

[23]   W. Tobler, 'Measuring spatial resolution', *Proceedings, Land Resources Information Systems Conference*, pp. 12–16, Jan. 1987.

[24]   E. A. Páez Lugo, 'Análisis multitemporal, proceso de zonificación y régimen de usos de los complejos de páramos delimitados a escala 1:25.000 en la jurisdicción CAR', Universidad Distrital Francisco José de Caldas, 2018. Accessed: Dec. 15, 2023. [Online]. Available: http://repository.udistrital.edu.co/handle/11349/14733

[25]   D. A. Aguirre Nieto, 'Identificación de las condiciones actuales del Páramo de Guerrero en la jurisdicción CAR – Dirección Regional Ubaté mediante fotografía satelital de acceso libre y SIG', presented at the I Seminario Nacional de Medio Ambiente e Innovación Ambiental, Corporación Autónoma Regional de Cundinamarca - CAR, 2018.

[26]   D. A. Aguirre Nieto, 'Colombia Guerrero Paramo Complex land cover classification with satellite imagery and machine learning algorithms', Politecnico di Milano, 2022.

[27]   J. F. Camargo Pérez and S. Chacón Ramírez, 'Estudio de los cambios en el uso del suelo en el Páramo de Guerrero, Tausa-Cundinamarca periodo 2007-2016', Universidad Distrital Francisco José de Caldas, 2018. Accessed: Dec. 13, 2023. [Online]. Available: http://repository.udistrital.edu.co/handle/11349/15583

[28]   C. A. Murad, J. Pearse, and C. Huguet, 'Multitemporal monitoring of paramos as critical water sources in Central Colombia', *Sci Rep*, vol. 14, no. 1, p. 16706, Jul. 2024, doi: 10.1038/s41598-024-67563-z.

[29]   CRCSI, 'Data—Basics and Acquisition.', in *Earth Observation: Data, Processing and Applications*, vol. 1A, 3 vols, B. A. Harrison, D. L. B. Jupp, M. M. Lewis, B. C. Forster, N. Mueller, C. Smith, S. Phinn, D. Hudson, I. Grant, and I. Coppa, Eds., Melbourne: Cooperative Research Centre for Spatial Information, 2016. Accessed: Dec. 15, 2023. [Online]. Available: https://www.eoa.org.au/earth-observation-textbooks

[30]   B. A. Harrison and D. L. B. Jupp, *Introduction to Remotely Sensed Data*. CSIRO Publishing, 1989. Accessed: Aug. 12, 2024. [Online]. Available: https://www.waterstones.com/book/introduction-to-remotely-sensed-data/b-a-harrison/d-l-b-jupp/9780643049918

[31]   NASA, 'What is Remote Sensing? | Earthdata', Earth Science Data Systems. Accessed: Aug. 12, 2024. [Online]. Available: https://www.earthdata.nasa.gov/learn/backgrounders/remote-sensing

[32]   C. Brown, L. N. Connor, J. Lillibridge, N. Nalli, and R. V. Legeckis, 'An introduction to satellite sensors, observations and techniques', 2005, pp. 21–50.

[33]   A. Mahood *et al.*, 'Ten simple rules for working with high resolution remote sensing data', *Peer Community Journal*, vol. 3, Jan. 2023, doi: 10.24072/pcjournal.223.

[34]   RESTEC, 'Various type of the Earth observation satellites (Japan and overseas)', Remote Sensing Technology Center of Japan. Accessed: Aug. 13, 2024. [Online]. Available: https://www.restec.or.jp/en/knowledge/sensing/sensing-3.html

[35]   CRCSI, 'Processing—Basic Image Operations', in *Earth Observation: Data, Processing and Applications*, vol. 2A, 3 vols, B. A. Harrison, D. L. B. Jupp, M. M. Lewis, B. C. Forster, T. Sparks, S. Phinn, N. Mueller, and G. Byrne, Eds., Melbourne: Cooperative Research Centre for Spatial Information, 2017. Accessed: Dec. 15, 2023. [Online]. Available: https://www.eoa.org.au/earth-observation-textbooks

[36]   EOS Data Analytics, 'Spatial Analysis Of Geospatial Data: Processing And Use Cases', Earth Observing System Data Analytics. Accessed: Aug. 14, 2024. [Online]. Available: https://eos.com/blog/spatial-analysis/

[37]   CRCSI, 'Processing—Image Rectification', in *Earth Observation: Data, Processing and Applications*, vol. 2B, 3 vols, B. A. Harrison, D. L. B. Jupp, M. M. Lewis, T. Sparks, N. Mueller, and G. Byrne, Eds., Melbourne: Cooperative Research Centre for Spatial Information, 2018. Accessed: Dec. 15, 2023. [Online]. Available: https://drive.google.com/file/d/1fTIzqK9XuCvDTTQ3W0Owcinc8uvI8yfC/view?usp=sharing&usp=embed_facebook

[38] G. Myhre, C. E. L. Myhre, B. H. Samset, and T. Storelvmo, 'Aerosols and their Relation to Global Climate and Climate Sensitivity', *Nature Education Knowledge*, vol. 4, no. 5, 2013, Accessed: Aug. 14, 2024. [Online]. Available: https://www.nature.com/scitable/knowledge/library/aerosols-and-their-relation-to-global-climate-102215345/

[39] CRCSI, 'Data—Image Interpretation', in *Earth Observation: Data, Processing and Applications*, vol. 1B, 3 vols, B. A. Harrison, D. L. B. Jupp, M. M. Lewis, B. C. Forster, I. Coppa, N. Mueller, D. Hudson, S. Phinn, C. Smith, J. Anstee, I. Grant, A. G. Dekker, C. Ong, and I. Lau, Eds., Melbourne: Cooperative Research Centre for Spatial Information, 2017. Accessed: Dec. 15, 2023. [Online]. Available: https://www.eoa.org.au/earth-observation-textbooks

[40] S. Liang and J. Wang, Eds., 'Chapter 4 - Atmospheric correction of optical imagery', in *Advanced Remote Sensing (Second Edition)*, Academic Press, 2020, pp. 131–156. doi: 10.1016/B978-0-12-815826-5.00004-0.

[41] G. Doxani *et al.*, 'Atmospheric Correction Inter-comparison eXercise, ACIX-II Land: An assessment of atmospheric correction processors for Landsat 8 and Sentinel-2 over land', *Remote Sensing of Environment*, vol. 285, p. 113412, Feb. 2023, doi: 10.1016/j.rse.2022.113412.

[42] J. Li *et al.*, 'Deep learning in multimodal remote sensing data fusion: A comprehensive review', *International Journal of Applied Earth Observation and Geoinformation*, vol. 112, p. 102926, Aug. 2022, doi: 10.1016/j.jag.2022.102926.

[43] R. E. Wolfe, D. P. Roy, and E. Vermote, 'MODIS land data storage, gridding, and compositing methodology: Level 2 grid', *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, no. 4, pp. 1324–1338, Jul. 1998, doi: 10.1109/36.701082.

[44] R. R. Downs and S. W. Olding, *EOSDIS Processing Levels RFC*, Program, Feb. 13, 2024. doi: https://doi.org/10.5067/DOC/ESCO/ESDS-RFC-050v1.

[45] W. B. Cohen and S. N. Goward, 'Landsat's Role in Ecological Applications of Remote Sensing', *BioScience*, vol. 54, no. 6, pp. 535–545, Jun. 2004, doi: 10.1641/0006-3568(2004)054[0535:LRIEAO]2.0.CO;2.

[46] USGS, 'Spectral Bandpasses for all Landsat Sensors | U.S. Geological Survey'. Accessed: Jun. 08, 2023. [Online]. Available: https://www.usgs.gov/media/images/spectral-bandpasses-all-landsat-sensors

[47] M. A. Wulder *et al.*, 'Fifty years of Landsat science and impacts', *Remote Sensing of Environment*, vol. 280, p. 113195, Oct. 2022, doi: 10.1016/j.rse.2022.113195.

[48] CEOS and ESA, 'Satellite earth observations in support of the sustainable development goals', *The CEOS Earth Observation Handbook*, 2018, Accessed: Dec. 16, 2023. [Online]. Available: https://eohandbook.com/sdg/

[49] ESA, 'Copernicus Programme'. Accessed: Aug. 18, 2024. [Online]. Available: https://sentiwiki.copernicus.eu/web/copernicus-programme

[50] ESA, 'Sentinel-2', Sentinel-2 SentiWiki chapter. Accessed: Aug. 15, 2024. [Online]. Available: https://sentiwiki.copernicus.eu/web/sentinel-2

[51] M. A. Wulder *et al.*, 'The global Landsat archive: Status, consolidation, and direction', *Remote Sensing of Environment*, vol. 185, pp. 271–283, Nov. 2016, doi: 10.1016/j.rse.2015.11.032.

[52] C. J. Crawford *et al.*, 'The 50-year Landsat collection 2 archive', *Science of Remote Sensing*, vol. 8, p. 100103, Dec. 2023, doi: 10.1016/j.srs.2023.100103.

[53] USGS, 'Landsat Project Statistics | U.S. Geological Survey'. Accessed: Aug. 19, 2024. [Online]. Available: https://www.usgs.gov/landsat-missions/landsat-project-statistics

[54] USGS, 'Landsat Acquisitions | U.S. Geological Survey'. Accessed: Aug. 19, 2024. [Online]. Available: https://www.usgs.gov/landsat-missions/landsat-acquisitions

[55] ESA, 'Sentinel-2 User Handbook'. European Space Agency, 2015. [Online]. Available: https://sentinel.esa.int/documents/247904/685211/Sentinel-2_User_Handbook

[56] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore, 'Google Earth Engine: Planetary-scale geospatial analysis for everyone', *Remote Sensing of Environment*, 2017, doi: 10.1016/j.rse.2017.06.031.

[57] M. A. Wulder *et al.*, 'Current status of Landsat program, science, and applications', *Remote Sensing of Environment*, vol. 225, pp. 127–147, May 2019, doi: 10.1016/j.rse.2019.02.015.

[58] G. Chander, B. L. Markham, and D. L. Helder, 'Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors', *Remote Sensing of Environment*, vol. 113, no. 5, pp. 893–903, May 2009, doi: 10.1016/j.rse.2009.01.007.

[59] D. Radočaj, J. Obhođaš, M. Jurišić, and M. Gašparović, 'Global Open Data Remote Sensing Satellite Missions for Land Monitoring and Conservation: A Review', *Land*, vol. 9, no. 11, Art. no. 11, Nov. 2020, doi: 10.3390/land9110402.

[60] S. Foga *et al.*, 'Cloud detection algorithm comparison and validation for operational Landsat data products', *Remote Sensing of Environment*, vol. 194, pp. 379–390, Jun. 2017, doi: 10.1016/j.rse.2017.03.026.

[61] X. Gao, H. Chi, J. Huang, Y. Han, Y. Li, and F. Ling, 'Comparison of Cloud-Mask Algorithms and Machine-Learning Methods Using Sentinel-2 Imagery for Mapping Paddy Rice in Jianghan Plain', *Remote Sensing*, vol. 16, no. 7, Art. no. 7, Jan. 2024, doi: 10.3390/rs16071305.

[62] ESA, 'Sentinel-2: Cloud Probability | Earth Engine Data Catalog'. Accessed: Jul. 16, 2023. [Online]. Available: https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_CLOUD_PROBABILITY

[63] U. Gandhi, 'Histogram Matching in Google Earth Engine', Spatial Thoughts Blog. Accessed: Aug. 21, 2024. [Online]. Available: https://spatialthoughts.com/2020/07/14/histogram-matching-gee/

[64] M. Restelli, 'Machine Learning Course', presented at the Introduction, Politecnico di Milano, 2022.

[65] J. MacQueen, 'Some methods for classification and analysis of multivariate observations', *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, pp. 281–297, 1967.

[66] C. Piech, 'K means', Stanford CS221. Accessed: Aug. 24, 2024. [Online]. Available: https://stanford.edu/~cpiech/cs221/handouts/kmeans.html

[67] C. M. Bishop, *Pattern Recognition and Machine Learning*. Cambridge: Springer, 2006. Accessed: Aug. 25, 2024. [Online]. Available: https://link.springer.com/book/9780387310732

[68] L. Breiman, 'Bagging Predictors', *Machine Learning*, vol. 24, pp. 123–140, 1996.

[69] IBM, 'What Is Random Forest? | IBM'. Accessed: Aug. 25, 2024. [Online]. Available: https://www.ibm.com/topics/random-forest

[70] L. Breiman, 'Random Forests', *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.

[71] Humboldt-Universität, 'Image classification - Random Forests', Image classification - Random Forests. Accessed: Jan. 08, 2024. [Online]. Available: https://pages.cms.hu-berlin.de/EOL/geo_rs/S08_Image_classification2.html

[72] C. Gómez, J. C. White, and M. A. Wulder, 'Optical remotely sensed time series data for land cover classification: A review', *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 116, pp. 55–72, Jun. 2016, doi: 10.1016/j.isprsjprs.2016.03.008.

[73] S. V. Stehman, 'Practical implications of design-based sampling inference for thematic map accuracy assessment', *Remote Sensing of Environment*, vol. 72, no. 1, pp. 35–45, 2000, doi: 10.1016/S0034-4257(99)00090-5.

[74] S. V. Stehman and G. M. Foody, 'Key issues in rigorous accuracy assessment of land cover products', *Remote Sensing of Environment*, vol. 231, p. 111199, Sep. 2019, doi: 10.1016/j.rse.2019.05.018.

[75] W. G. Cochran, *Sampling Techniques*, 3rd ed. John Wiley & Sons, 1977. Accessed: Feb. 27, 2024. [Online]. Available: https://www.academia.edu/29684662/Cochran_1977_Sampling_Techniques_Third_Edition

[76] D. Phiri, M. Simwanda, S. Salekin, V. R. Nyirenda, Y. Murayama, and M. Ranagalage, 'Sentinel-2 Data for Land Cover/Use Mapping: A Review', *Remote Sensing*, vol. 12, no. 14, Art. no. 14, Jan. 2020, doi: 10.3390/rs12142291.

[77] M. Buchhorn, M. Lesiv, N.-E. Tsendbazar, M. Herold, L. Bertels, and B. Smets, 'Copernicus Global Land Cover Layers—Collection 2', *Remote Sensing*, vol. 12, no. 6, Art. no. 6, Jan. 2020, doi: 10.3390/rs12061044.

[78] D. Zanaga *et al.*, 'ESA WorldCover 10 m 2020 v100'. Zenodo, Oct. 20, 2021. doi: 10.5281/zenodo.5571936.

[79] D. Zanaga *et al.*, 'ESA WorldCover 10 m 2021 v200'. Zenodo, Oct. 28, 2022. doi: 10.5281/zenodo.7254221.

[80] K. Karra, C. Kontgis, Z. Statman-Weil, J. C. Mazzariello, M. Mathis, and S. P. Brumby, 'Global land use / land cover with Sentinel 2 and deep learning', in *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, Jul. 2021, pp. 4704–4707. doi: 10.1109/IGARSS47720.2021.9553499.

[81] C. F. Brown *et al.*, 'Dynamic World, Near real-time global 10 m land use land cover mapping', *Scientific Data*, vol. 9, no. 1, p. 251, Jun. 2022, doi: 10.1038/s41597-022-01307-4.

[82] Z. S. Venter, D. N. Barton, T. Chakraborty, T. Simensen, and G. Singh, 'Global 10 m Land Use Land Cover Datasets: A Comparison of Dynamic World, World Cover and Esri Land Cover', *Remote Sensing*, vol. 14, no. 16, Art. no. 16, Jan. 2022, doi: 10.3390/rs14164101.

[83] Y. Wang, Y. Sun, X. Cao, Y. Wang, W. Zhang, and X. Cheng, 'A review of regional and Global scale Land Use/Land Cover (LULC) mapping products generated from satellite remote sensing', *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 206, pp. 311–334, Dec. 2023, doi: 10.1016/j.isprsjprs.2023.11.014.

[84] A. Di Gregorio and L. J. Jansen, 'Land Cover Classification System (LCCS): classification concepts and user manual', *FAO, Rome*, 1998.

[85] M. C. Hansen and B. Reed, 'A comparison of the IGBP DISCover and university of maryland 1 km global land cover products', *International Journal of Remote Sensing*, vol. 21, no. 6–7, pp. 1365–1373, 2000, doi: 10.1080/014311600210218.

[86]     J. R. Anderson, *A land use and land cover classification system for use with remote sensor data*, vol. 964. US Government Printing Office, 1976.

[87]     Copernicus, 'Corine Land Cover', Land Monitoring Service. Accessed: Aug. 26, 2024. [Online]. Available: https://land.copernicus.eu/en/products/corine-land-cover

[88]     IDEAM, *Leyenda Nacional de Coberturas de la Tierra. Metodología CORINE Land Cover adaptada para Colombia Escala 1:100.000*. Bogotá D.C.: Instituto de Hidrología, Meteorología y Estudios Ambientales, 2010.

[89]     IDEAM, 'Coberturas Nacionales'. [Online]. Available: http://www.ideam.gov.co/web/ecosistemas/ coberturas-nacionales

[90]     IDEAM, 'Mapa de Cobertura de la Tierra. Adaptación Corine Land Cover. República de Colombia. Escala 1:100.000. Periodo 2018', Instituto de Hidrología, Meteorología y Estudios Ambientales, Bogotá D.C., 2021. [Online]. Available: https://www.colombiaenmapas.gov.co/?e=-99.51609485351166,-13.695728388253686,-48.97898547852508,23.416278879819167,4686&b=igac&u=0&t=43&servicio=881

[91]     IDEAM, IGAC, IAvH, Invemar, SINCHI, and IIAP, *Ecosistemas continentales, costeros y marinos de Colombia*. Bogotá D.C.: Instituto de Hidrología, Meteorología y Estudios Ambientales, 2007. [Online]. Available: http://www. invemar.org.co/documents/10182/0/ecosistemas_continentales_costeros_y_marinos.      pdf/cad522d4-7fbc-4d8c-85e2-5b3120a48e9e.

[92]     IDEAM, IGAC, IAvH, Invemar, SINCHI, and IIAP, 'Ecosistemas continentales, costeros y marinos de Colombia v. 2.1', Instituto de Hidrología, Meteorología y Estudios Ambientales, Bogotá D.C., 2017. [Online]. Available: http://www.ideam.gov.co/documents/11769/222663/E_ECCMC_Ver21_100K.pdf/addc175f-3ac6-415b-9b9e-a1c4368b5b3e

[93]     A. González-González, N. Clerici, and B. Quesada, 'A 30 m-resolution land use-land cover product for the Colombian Andes and Amazon using cloud-computing', *International Journal of Applied Earth Observation and Geoinformation*, vol. 107, p. 102688, Mar. 2022, doi: 10.1016/j.jag.2022.102688.

[94]     J. Cihlar, 'Land cover mapping of large areas from satellites: Status and research priorities', *International Journal of Remote Sensing*, vol. 21, no. 6–7, pp. 1093–1114, 2000, doi: 10.1080/014311600210092.

[95]     M. Story and R. G. Congalton, 'Accuracy assessment: a user's perspective.', *Photogrammetric Engineering & Remote Sensing*, vol. 52, no. 3, pp. 397–399, 1986.

[96]     P. Olofsson, G. M. Foody, M. Herold, S. V. Stehman, C. E. Woodcock, and M. A. Wulder, 'Good practices for estimating area and assessing accuracy of land change', *Remote Sensing of Environment*, vol. 148, pp. 42–57, May 2014, doi: 10.1016/j.rse.2014.02.015.

[97]     G. M. Foody, 'Status of land cover classification accuracy assessment', *Remote Sensing of Environment*, vol. 80, no. 1, pp. 185–201, Apr. 2002, doi: 10.1016/S0034-4257(01)00295-4.

[98]     P. Olofsson, G. M. Foody, S. V. Stehman, and C. E. Woodcock, 'Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation', *Remote Sensing of Environment*, vol. 129, pp. 122–131, Feb. 2013, doi: 10.1016/j.rse.2012.10.031.

[99]     P. Coppin, I. Jonckheere, K. Nackaerts, B. Muys, and E. Lambin, 'Review ArticleDigital change detection methods in ecosystem monitoring: a review', *International Journal of Remote Sensing*, vol. 25, no. 9, pp. 1565–1596, May 2004, doi: 10.1080/0143116031000101675.

[100]   K. S. Willis, 'Remote sensing change detection for ecological monitoring in United States protected areas', *Biological Conservation*, vol. 182, pp. 233–242, Feb. 2015, doi: 10.1016/j.biocon.2014.12.006.

[101]   R. D. Swetnam, 'Rural land use in England and Wales between 1930 and 1998: Mapping trajectories of change with a high resolution spatio-temporal dataset', *Landscape and Urban Planning*, vol. 81, no. 1, pp. 91–103, May 2007, doi: 10.1016/j.landurbplan.2006.10.013.

[102]   S. S. Shapiro and M. B. Wilk, 'An analysis of variance test for normality (complete samples)†', *Biometrika*, vol. 52, no. 3–4, pp. 591–611, Dec. 1965, doi: 10.1093/biomet/52.3-4.591.

[103]   T. W. Anderson and D. A. Darling, 'Asymptotic Theory of Certain "Goodness of Fit" Criteria Based on Stochastic Processes', *The Annals of Mathematical Statistics*, vol. 23, no. 2, pp. 193–212, Jun. 1952, doi: 10.1214/aoms/1177729437.

[104]   T. Dodge, 'Kolmogorov–Smirnov Test', in *The Concise Encyclopedia of Statistics*, New York: Springer, 2008.

[105]   R. B. D'Agostino, 'An Omnibus Test of Normality for Moderate and Large Size Samples', *Biometrika*, vol. 58, no. 2, pp. 341–348, 1971, doi: 10.2307/2334522.

[106]   R. D'Agostino and E. S. Pearson, 'Tests for Departure from Normality. Empirical Results for the Distributions of b2 and √ b1', *Biometrika*, vol. 60, no. 3, pp. 613–622, 1973, doi: 10.2307/2335012.

[107]  C. M. Jarque and A. K. Bera, 'Efficient tests for normality, homoscedasticity and serial independence of regression residuals', *Economics Letters*, vol. 6, no. 3, pp. 255–259, Jan. 1980, doi: 10.1016/0165-1765(80)90024-5.

[108]  H. W. Lilliefors, 'On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown', *Journal of the American Statistical Association*, vol. 62, no. 318, pp. 399–402, 1967, doi: 10.2307/2283970.

[109]  W. H. Kruskal and W. A. Wallis, 'Use of Ranks in One-Criterion Variance Analysis', *Journal of the American Statistical Association*, vol. 47, no. 260, pp. 583–621, 1952, doi: 10.2307/2280779.

[110]  O. J. Dunn, 'Multiple Comparisons Using Rank Sums', *Technometrics*, vol. 6, no. 3, pp. 241–252, 1964, doi: 10.2307/1266041.

[111]  R. A. Fisher, 'On the "Probable Error" of a Coefficient of Correlation Deduced from a Small Sample', 1921, Accessed: Aug. 30, 2024. [Online]. Available: https://digital.library.adelaide.edu.au/dspace/handle/2440/15169

[112]  J. W. Tukey, 'Comparing Individual Means in the Analysis of Variance', *Biometrics*, vol. 5, no. 2, pp. 99–114, 1949, doi: 10.2307/3001913.

[113]  D. Montero, C. Aybar, M. D. Mahecha, F. Martinuzzi, M. Söchting, and S. Wieneke, 'A standardized catalogue of spectral indices to advance the use of remote sensing in Earth system research', *Sci Data*, vol. 10, no. 1, p. 197, Apr. 2023, doi: 10.1038/s41597-023-02096-0.

[114]  J. W. Rouse, R. H. Haas, J. A. Schell, and D. W. Deering, 'Monitoring vegetation systems in the Great Plains with ERTS', Jan. 1974. Accessed: Aug. 30, 2024. [Online]. Available: https://ntrs.nasa.gov/citations/19740022614

[115]  L. Hannah, J. L. Carr, and A. Lankerani, 'Human disturbance and natural habitat: a biome level analysis of a global data set', *Biodivers Conserv*, vol. 4, no. 2, pp. 128–155, Mar. 1995, doi: 10.1007/BF00137781.

[116]  E. W. Sanderson, M. Jaiteh, M. A. Levy, K. H. Redford, A. V. Wannebo, and G. Woolmer, 'The Human Footprint and the Last of the Wild: The human footprint is a global map of human influence on the land surface, which suggests that human beings are stewards of nature, whether we like it or not', *BioScience*, vol. 52, no. 10, pp. 891–904, Oct. 2002, doi: 10.1641/0006-3568(2002)052[0891:THFATL]2.0.CO;2.

[117]  A. Etter, C. A. McAlpine, L. Seabrook, and K. A. Wilson, 'Incorporating temporality and biophysical vulnerability to quantify the human spatial footprint on ecosystems', *Biological Conservation*, vol. 144, no. 5, pp. 1585–1594, May 2011, doi: 10.1016/j.biocon.2011.02.004.

[118]  C. A. Correa Ayram, A. Etter, J. Díaz-Timoté, S. Rodríguez Buriticá, W. Ramírez, and G. Corzo, 'Spatiotemporal evaluation of the human footprint in Colombia: Four decades of anthropic impact in highly biodiverse ecosystems', *Ecological Indicators*, vol. 117, p. 106630, Oct. 2020, doi: 10.1016/j.ecolind.2020.106630.

[119]  S. López, M. F. López-Sandoval, A. Gerique, and J. Salazar, 'Landscape change in Southern Ecuador: An indicator-based and multi-temporal evaluation of land use and land cover in a mixed-use protected area', *Ecological Indicators*, vol. 115, p. 106357, Aug. 2020, doi: 10.1016/j.ecolind.2020.106357.

[120]  A. M. Groot, L. Herrera, S. Mora, and Á. Botiva Contreras, 'V. La altiplanicie cundiboyacense', in *Colombia prehispánica: regiones arqueológicas*, Bogotá: Colcultura, Instituto Colombiano de Antropología, 1989. Accessed: Sep. 04, 2024. [Online]. Available: https://babel.banrepcultural.org/digital/collection/p17054coll10/id/2781

[121]  V. Urrea, A. Ochoa, and O. Mesa, 'Seasonality of Rainfall in Colombia', *Water Resources Research*, vol. 55, no. 5, pp. 4149–4162, 2019, doi: 10.1029/2018WR023316.

[122]  D. A. Gómez-Latorre, G. A. Araújo-Carrillo, and Y. R. Leguizamón, 'Regionalización de patrones de lluvias para periodos multianuales secos y húmedos en el Altiplano Cundiboyacense de Colombia. | RClimatol', *Revista de Climatología*, vol. 22, pp. 162–177, 2022.

[123]  M. C. Schrimpff, 'En busca de los primeros agricultores del altiplano cundiboyacense', *Maguaré*, no. 5, 1987.

[124]  K. G. Mora Pacheco, 'Agriculture and Livestock in Wetlands in the Bogota Plateau (Colombia), Eighteenth Century. Land Use and Wetland Management', in *Environmental History in the Making: Volume II: Acting*, C. Joanaz de Melo, E. Vaz, and L. M. Costa Pinto, Eds., Cham: Springer International Publishing, 2017, pp. 3–13. doi: 10.1007/978-3-319-41139-2_1.

[125]  P. M. Argüello García, 'El patrón de poblamiento prehispánico en el altiplano cundiboyacense y sus consecuencias en la constitución de la sociedad colonial', *Revista Colombiana de Antropología*, vol. 59, no. 3, Art. no. 3, Sep. 2023, doi: 10.22380/2539472X.2467.

[126]  DANE, 'Censo Nacional de Población y Vivienda 2018', Departamento Administrativo Nacional de Estadística, Colombia, 2019. [Online]. Available: https://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/censo-nacional-de-poblacion-y-vivenda-2018

[127]  O. Robineau, M. Châtelet, M. Isabelle, C. Soulard, and J. Posner, 'WHAT ARE FARMERS' DEVELOPMENT OPTIONS IN THE CONTEXT OF PÁRAMO CONSERVATION ? A CASE STUDY FROM COLOMBIA', *Innovation and Sustainable Development in Agriculture and Food*, Jun. 2010.

[128]    N. León Rodríguez, 'EL PÁRAMO DE GUERRERO: CONFLICTOS ENTRE CONSERVACIÓN Y REPRIMARIZACIÓN DE SU ECONOMÍA', *REVGEO*, vol. 2, no. 47E, Feb. 2012, Accessed: Sep. 04, 2024. [Online]. Available: https://www.revistas.una.ac.cr/index.php/geografica/article/view/2533

[129]    IAvH, CAR, CORPOBOYACÁ, and CORPOCHIVOR, 'Estudio del estado actual del Paramo de Rabanal'. Instituto de Investigación de recursos biológicos Alexander von Humboldt, 2008. [Online]. Available: https://faunacorpochivor.gov.co/wp-content/uploads/2016/04/Estudio-del-estado-actual-del-paramo-de-Rabanal.pdf

[130]    C. N. de la R. Colombia, *Art. 173 - Ley 1753 Por la cual se expide el Plan Nacional de Desarrollo 2015 - 2018*. 2015. [Online]. Available: https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=61933

[131]    G. S. Ortiz Delgado, *Sentencia C-035 de 2016*. 2016.

[132]    M. Rodríguez and C. Ardila, 'Impacto socioeconómico y ambiental de la pequeña minería del carbón en los municipios de Zipaquirá, Nemocón, Cogua, Suesca, Tausa, Sutatausa, Cucunubá y Lenguazaque', 1999.

[133]    C. Duarte and W. Monroy, 'Exploración gas metano asociado al carbón. Área Checua-Lenguazaque. Sector GMAC Boquerón de Tausa-La Pluma, Cucunubá', 2012.

[134]    C. L. Guatame and G. Sarmiento, 'Interpretación del ambiente sedimentario de los carbones de la formación guaduas en el sinclinal checua-lenguazaque a partir del análisis petrográfico', *Geología Colombiana*, vol. 29, pp. 41–57, 2004.

[135]    T. Moore *et al.*, 'Petrographic and geochemical characteristics of selected coal seams from the Late Cretaceous-Paleocene Guaduas Formation, Eastern Cordillera Basin, Colombia', *International Journal of Coal Geology*, vol. 259, p. 104042, Jun. 2022, doi: 10.1016/j.coal.2022.104042.

[136]    O. Aduvire, 'Drenaje ácido de mina generación y tratamiento', *Instituto Geológico y Minero de España Dirección de Recursos Minerales y Geoambiente*, vol. 140, 2006.

[137]    M. D. González-Martínez, C. Huguet, J. Pearse, N. McIntyre, and L. A. Camacho, 'Assessment of potential contamination of Paramo soil and downstream water supplies in a coal-mining region of Colombia', *Applied Geochemistry*, vol. 108, p. 104382, Sep. 2019, doi: 10.1016/j.apgeochem.2019.104382.

[138]    Conservación Iternacional and CAR, 'DECLARATORIA Y FORMULACIÓN DEL PLAN DE MANEJO DE UN ÁREA DE CONSERVACIÓN EN JURISDICCIÓN DEL PÁRAMO DE GUERRERO'. Conservación Internacional Colombia, 2008. [Online]. Available: https://faunacorpochivor.gov.co/wp-content/uploads/2016/04/Estudio-del-estado-actual-del-paramo-de-Rabanal.pdf

[139]    CORPOCHIVOR, 'DISTRITO REGIONAL DE MANEJO INTEGRADO (DRMI) PARAMO DE CRISTALES, CASTILLEJO O GUACHANEQUE'. Instituto de Investigación de recursos biológicos Alexander von Humboldt, 2011. [Online]. Available:    https://faunacorpochivor.gov.co/wp-content/uploads/2016/04/Documento-si%CC%81ntesis-declaratoria-DRMI-Cristales-Castillejo-o-Guachaneque.pdf

[140]    M. F. Mora Goyes, J. A. Rubio, R. Ocampo Gutiérrez, and J. I. Barrera Cataño, *Catálogo de Especies Invasoras del Territorio CAR*. Pontificia Universidad Javeriana, 2018.

[141]    S. Ma *et al.*, 'Autochthonous organic matter input in reservoirs: Limited methane oxidation in sediments fails to suppress methane emission', *Science of The Total Environment*, vol. 945, p. 174122, Oct. 2024, doi: 10.1016/j.scitotenv.2024.174122.

[142]    DANE, 'INFORME TÉCNICO CARACTERIZACIÓN POBLACIONAL ÁREAS DE PÁRAMO A PARTIR DE LOS RESULTADOS DEL CENSO NACIONAL DE POBLACIÓN Y VIVIENDA CNPV 2018', Departamento Administrativo Nacional de Estadística, Colombia, 2020.

[143]    J. G. Masek *et al.*, 'A Landsat Surface Reflectance Dataset for North America, 1990–2000', *IEEE Geosci. Remote Sensing Lett.*, vol. 3, no. 1, pp. 68–72, Jan. 2006, doi: 10.1109/LGRS.2005.857030.

[144]    E. Vermote, C. Justice, M. Claverie, and B. Franch, 'Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product', *Remote Sensing of Environment*, vol. 185, pp. 46–56, Nov. 2016, doi: 10.1016/j.rse.2016.04.008.

[145]    MADS, 'Paramos Delimitados Junio 2020', 2020. [Online]. Available: https://www.arcgis.com/sharing/rest/content/items/9631ed8c44274baa824e6277276de48f/data

[146]    USGS, 'USGS Landsat 2 MSS Collection 2 Tier 2 Raw Scenes | Earth Engine Data Catalog'. Accessed: Jul. 16, 2023. [Online]. Available: https://developers.google.com/earth-engine/datasets/catalog/LANDSAT_LM02_C02_T2

[147]    USGS, 'USGS Landsat 4 MSS Collection 2 Tier 2 Raw Scenes | Earth Engine Data Catalog'. Accessed: Jul. 16, 2023. [Online]. Available: https://developers.google.com/earth-engine/datasets/catalog/LANDSAT_LM04_C02_T2

[148]    USGS, 'USGS Landsat 4 Level 2, Collection 2, Tier 1 | Earth Engine Data Catalog'. Accessed: Jul. 16, 2023. [Online]. Available: https://developers.google.com/earth-engine/datasets/catalog/LANDSAT_LT04_C02_T1_L2

[149] USGS, 'USGS Landsat 5 Level 2, Collection 2, Tier 1 | Earth Engine Data Catalog'. Accessed: Jul. 16, 2023. [Online]. Available: https://developers.google.com/earth-engine/datasets/catalog/LANDSAT_LT05_C02_T1_L2

[150] USGS, 'USGS Landsat 7 Level 2, Collection 2, Tier 1 | Earth Engine Data Catalog'. Accessed: Jul. 16, 2023. [Online]. Available: https://developers.google.com/earth-engine/datasets/catalog/LANDSAT_LE07_C02_T1_L2

[151] USGS, 'USGS Landsat 8 Level 2, Collection 2, Tier 1 | Earth Engine Data Catalog'. Accessed: Jul. 16, 2023. [Online]. Available: https://developers.google.com/earth-engine/datasets/catalog/LANDSAT_LC08_C02_T1_L2

[152] Sinergise Ltd., 'EO-Browser', Sentinel-hub EO-Browser. Accessed: May 18, 2023. [Online]. Available: https://apps.sentinel-hub.com/eo-browser/?zoom=6&lat=6.56639&lng=-73.45459&themeId=DEFAULT-THEME

[153] ESA, 'Harmonized Sentinel-2 MSI: MultiSpectral Instrument, Level-2A | Earth Engine Data Catalog'. Accessed: Jul. 16, 2023. [Online]. Available: https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS/S2_SR_HARMONIZED

[154] D. Yamazaki *et al.*, 'A high-accuracy map of global terrain elevations', *Geophysical Research Letters*, vol. 44, no. 11, pp. 5844–5853, 2017, doi: 10.1002/2017GL072874.

[155] V. J. García, C. O. Márquez, T. M. Isenhart, M. Rodríguez, S. D. Crespo, and A. G. Cifuentes, 'Evaluating the conservation state of the Páramo ecosystem: An object-based image analysis and CART algorithm approach for central Ecuador', *Heliyon*, vol. 5, no. 10, p. e02701, Oct. 2019, doi: 10.1016/j.heliyon.2019.e02701.

[156] T. N. Phan, V. Kuch, and L. W. Lehnert, 'Land Cover Classification using Google Earth Engine and Random Forest Classifier—The Role of Image Composition', *Remote Sensing*, vol. 12, no. 15, Art. no. 15, Jan. 2020, doi: 10.3390/rs12152411.

[157] A. Septianto and A. S. Rand, 'Comparison of machine learning algorithms for land use and land cover analysis using google earth engine (case study: wanggu watershed)', *International Journal of Remote Sensing and Earth Sciences (Denpasar)*, vol. 19, no. 2, pp. 197–197, Jan. 2023, doi: 10.30536/j.ijreses.2022.v19.a3803.

[158] M. G. Tulbure, P. Hostert, T. Kuemmerle, and M. Broich, 'Regional matters: On the usefulness of regional land-cover datasets in times of global change', *Remote Sensing in Ecology and Conservation*, vol. 8, no. 3, pp. 272–283, 2022, doi: 10.1002/rse2.248.

[159] J. Burbano-Girón, M. A. M. Berbeo, C. G. Montoya, C. A. Cruz-Rodríguez, and J. M. Ochoa-Quintero, 'Estado de conservación de los Páramos en Colombia', *Instituto Humbolt*, vol. 2002–2009, 2012.

[160] Y. Forero-Gómez, P. Gil Leguizamón, and M. Morales-Puentes, 'Conectividad estructural entre los Páramos de Guacheneque y Los Cristales, complejo Rabanal-río Bogotá, Colombia', *Revista de Teledetección*, vol. 0, p. 65, Dec. 2020, doi: 10.4995/raet.2020.13946.

[161] P. Ungar, 'Assembling an Ecosystem: The Making of State: Páramos: in Colombia', *Conservation and Society*, vol. 19, no. 2, p. 119, Jun. 2021, doi: 10.4103/cs.cs_19_103.

[162] S. C. Mills *et al.*, 'Avian biodiversity losses from grazing of high Andean páramo', *Biological Conservation*, vol. 286, p. 110298, Oct. 2023, doi: 10.1016/j.biocon.2023.110298.

# A  Appendix A

This Appendix includes supplementary figures and tables from the Thesis

## A.1.  Methodological Flowchart

The full methodological flowchart of the Thesis, composed by the three stages depicted in Figure 3.1 is shown below. All files and scripts are shared following the open-source spirit of the research conducted. A detailed list of all these elements can be found on the following appendix section.

**Legend (right panel):** Data, Step, Info, Script, File, Output, Random, Asset, Px Filter, Clip, Official External Source

**Data sources:**
- USGS Landsat 2 to 9 Collection 2 Data
- COPERNICUS Sentinel-2 SR Data
- MADS Delimited Paramo Areas (06/2020)

**Data Pool Exploration**
- Path-Row / Grid Filter
- Min Cloud Cover Filter
→ 00_[Mission]_Exploration

**Data Pool Scene Ranking**
- Qualitative (Cloud Cover)
- Quantitative (Area Coverage)
→ [Mission]_Summary

**Data Pool Selection**
- Year and Instrument
- Scene Code assignment

**11 Epochs:**
- 1977
- 1988
- 1988T
- 1991T
- 2001T
- 2007E
- 2011E
- 2015O
- 2019O
- 2019M
- 2024M

**5 Instruments:**
- MSS (-)
- TM (T)
- ETM+ (E)
- OLI (O)
- MSI (M)

**Scene Pre-Processing**

[Year]_Preproc_[Instrument]

**MSS**
- Shift
- DN to TOA_SR
- Co_Registering
- Compositing
- Cloud Masking
- Other Masking
- Mosaicking

**TM / ETM+ / OLI**
- Cloud Masking
- Other Masking
- Histogram Matching
- Mosaicking

**MSI**
- Collection Reducer
- Cloud Masking
- Other Masking
- Histogram Matching
- Mosaicking

**Paramo Complex Epoch Scene**
→ [Mission]_[Year]_[Paramo]

---

**For Each Paramo Complex Scene**

**Visualization & Sample Area Selection**
- NIR / RGB / AGRI Visualization
- MSS Pixel Clustering (TOA Data)
- Across-Epochs Stable Sample Area Manual Polygon Selection (Rectangles)
- Stratified Sampling following rough MSS LULC proportions

**5 LULC Classes:**
- Water (0 / W)
- Bare-Rock-Built (1 / B)
- Agriculture-Grass (2 / A)
- Forest (3 / F)
- Shrubland-Wetland (4 / S)

→ [Paramo]_Processing

[Class]

**Sample Area Spectral Characterization**

Split: 0,7 / 0,3 → Train set / Test set

**Random Forest Parameter Selection**
'Best OA for worst Dataset' -> 1977 MSS
- Number of Trees

x20 → Seed

Distance [m] → Train set / Test set

**Random Forest Paramo LULC Supervised Classification & Accuracy Assessment**

**Results Statistical Analysis**
- Normality_Testing_[Paramo]
- Inter_Annual_Testing_[Paramo]
→ [Paramo]_Results

**LULC Aggregation:**
- Average
- Mode + Agreement

**Paramo Complex Mode LULC**
→ [Paramo]_[Year(2-digit)]

Scene / PA Geometry

**Mode Accuracy Assessment**

**Pixel Count Class Area Calculation**

→ [Paramo]_PA_[Year(2-digit)]

**Accuracy Derived Class Area Calculation**

**High Confidence (Agreement ≥ 70%) Area Calculation**

→ Processing_[Paramo]

→ [Paramo]_PA_[Year(2-digit)]_HC

**Binary Conversion (A / Not A)**

→ [Paramo]_PA_[Year(2-digit)]__C2B_HC

---

**Subdivision Disturbance Index**
→ [Paramo]_Subdiv_Index

[Paramo]_Results

**Administrative and Hydrological Subdivision Analysis**
- CAR River Basin
- IGAC Administrative Division

**Paramo Disturbance Index**
→ [Paramo]_change_[Step]
→ Processing_[Paramo]

**Change Class:**
- Stable O
- A Gain
- A Loss
- Stable A

**Change Class:**
- Stable O
- A Gain
- Cyclical
- A Loss
- Stable A

**Binary Change Detection:**
- MSS 1977-1988
- MSI 2019-2024

**Binary Temporal Trajectory:**
- TM/ETM/OLI 1988T to 2019O

→ [Paramo]_Temporal_T

**Binary LULC Class:**
- Other (1/O)
- Agriculture-Grass (2/A)

**Paramo Complex PA Epoch HC Binary Assets**

## A.2. Files & Scripts

List of all files and scripts used in the Thesis. Assets can be accessed via the GEE scripts. All are available on the following repository: https://github.com/Daenel35/Paramo-Land-Cover-Monitoring

| File | File Type | Script Name | Script Type |
|---|---|---|---|
| Thesis Satellite Data Pool: | Microsoft Excel | 00_Landsat_Exploration | GEE JavaScript |
| LANDSAT_Summary | Excel Page | 00_Landsat_1to5_Exploration | GEE JavaScript |
| Proportions | Excel Page | 00_Landsat_5_TM_Exploration | GEE JavaScript |
| Pre_Processing MSS | Excel Page | 00_Landsat_7to9_Exploration | GEE JavaScript |
| Pre_ProcessingTM | Excel Page | 00_Sentinel2_Exploration | GEE JavaScript |
| Pre_Processing ETM | Excel Page | 1977_Preproc_MSS | GEE JavaScript |
| Pre_Processing OLI | Excel Page | 1977_Preproc_MSS_1988_GeoCorr | GEE JavaScript |
| SENTINEL_Summary | Excel Page | 1988_Preproc_MSS | GEE JavaScript |
| Pre_Processing MSI | Excel Page | 1988_Preproc_TM | GEE JavaScript |
| Processing_RF_init_Tree | Microsoft Excel | 1991_Preproc_TM | GEE JavaScript |
| GUE_ACB_A_Results | Microsoft Excel | 2001_Preproc_TM | GEE JavaScript |
| GUE_ACB_A_Subdiv_index | Microsoft Excel | 2007_Preproc_ETM | GEE JavaScript |
| Processing_GUE | Microsoft Excel | 2011_Preproc_ETM | GEE JavaScript |
| RRB_ACB_B_Results | Microsoft Excel | 2015_Preproc_OLI | GEE JavaScript |
| Processing_RRB | Microsoft Excel | 2019_Preproc_OLI | GEE JavaScript |
| | | 2019_Preproc_MSI | GEE JavaScript |
| | | 2024_Preproc_MSI | GEE JavaScript |
| | | DEM_selection | GEE JavaScript |
| | | GUE_ACB_A_Processing | GEE JavaScript |
| | | GUE_ACB_A_Temporal_T | GEE JavaScript |
| | | RRB_ACB_B_Processing | GEE JavaScript |
| | | RRB_ACB_B_Temporal_T | GEE JavaScript |
| | | Scene_Paramo_Areas | GEE JavaScript |
| | | Normality_Testing_GUE_ACB_A | Python |
| | | Normality_Testing_RRB_ACB_B | Python |
| | | Inter_Annual_Testing_GUE | Python |
| | | Inter_Annual_Testing_RRB | Python |

# A.3. Landsat Data Pool Database

Supplementary table for Landsat Data Pool.

### Páramo Complex Area Cloud Cover Qualitative Rating

Very Little to None ................................................ Mostly Covered

### Proportion of Páramo Complex Area Covered in Scene

| 100% | 75-99% | 50-75% | 25-50% | 1-25% | 0% |

| | | | | | PÁRAMO COMPLEX | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Mission & Inst.** | **Directory (LANDSAT/…)** | **Year** | **Code** | **Proc. Level** | **Res. [m]** | **GUE** | **RRB** | **CHI** | **IGM** | **ACB** | **CVS** | **Notes** |
| **Landsat 2 MSS** | LM02/C02/T2/LM02_008056_19770107 | 1977 | L2_8_56_1977 | Tier 2 Raw | 60 | | | | | | | 1977 MSS REFERENCE, No clouds on interest areas |
| **Landsat 2 MSS** | LM02/C02/T2/LM02_008057_19770107 | 1977 | L2_8_57_1977 | Tier 2 Raw | 60 | | | | | | | 1977 MSS REFERENCE, Image Shifted, scattered clouds in CVS |
| **Landsat 2 MSS** | LM02/C02/T2/LM02_008058_19770107 | 1977 | L2_8_58_1977 | Tier 2 Raw | 60 | | | | | | | 1977 MSS REFERENCE, Image Shifted, scattered clouds in CVS, completes South CVS |
| Landsat 4 MSS | LM04/C02/T2/LM04_008056_19840123 | 1984 | - | Tier 2 Raw | 60 | | | | | | | SEC, Image Shifted, haze |
| Landsat 5 TM | LT05/C02/T1_L2/LT05_008056_19850322 | 1985 | - | Tier 1 Level 2 | 30 | | | | | | | SEC, Clouds on north GUE, all over IGM and scattered on RRB |
| Landsat 5 TM | LT05/C02/T1_L2/LT05_008057_19850322 | 1985 | - | Tier 1 Level 2 | 30 | | | | | | | SEC, Clouds south of CVS on most edges and south of CHI |
| **Landsat 4 TM** | LT04/C02/T1_L2/LT04_008056_19871217 | 1987 | L4T_8_56_1987 | Tier 1 Level 2 | 30 | | | | | | | 1988 TM REFERENCE, December 1987 better for RRB than March 1988 |
| **Landsat 4 TM** | LT04/C02/T1_L2/LT04_007056_19880111 | 1988 | L4T_7_56_1988 | Tier 1 Level 2 | 30 | | | | | | | 1988 TM REFERENCE, Scattered clouds Northwest GLR, West RRB and Southwest TBM |
| **Landsat 4 TM** | LT04/C02/T1_L2/LT04_007057_19880111 | 1988 | L4T_7_57_1988 | Tier 1 Level 2 | 30 | | | | | | | 1988 TM REFERENCE, Clouds East visible CHI, completes the West corner CHI |

| Mission & Inst. | Directory (LANDSAT/…) | Year | Code | Proc. Level | Res. [m] | PÁRAMO COMPLEX | | | | | | Notes |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | GUE | RRB | CHI | IGM | ACB | CVS | |
| **Landsat 4 MSS** | LM04/C02/T2/LM04_008056_19880322 | 1988 | L4_8_56_1988 | Tier 2 Raw | 60 | | | | | | | **1988 MSS REFERENCE, clouds on south and east RRB** |
| **Landsat 4 MSS** | LM04/C02/T2/LM04_008057_19880322 | 1988 | L4_8_57_1988 | Tier 2 Raw | 60 | | | | | | | **1988 MSS REFERENCE, scattered clouds in CVS** |
| **Landsat 4 MSS** | LM04/C02/T2/LM04_008058_19880322 | 1988 | L4_8_58_1988 | Tier 2 Raw | 60 | | | | | | | **1988 MSS REFERENCE, clouds in CVS, completes South CVS** |
| **Landsat 4 TM** | LT04/C02/T1_L2/LT04_008056_19880322 | 1988 | L4T_8_56_1988 | Tier 1 Level 2 | 30 | | | | | | | **1988 TM REFERENCE, clouds on south and east RRB** |
| **Landsat 4 TM** | LT04/C02/T1_L2/LT04_008057_19880322 | 1988 | L4T_8_57_1988 | Tier 1 Level 2 | 30 | | | | | | | **1988 TM REFERENCE, scattered clouds in CVS** |
| **Landsat 4 TM** | LT04/C02/T1_L2/LT04_008058_19880322 | 1988 | L4T_8_58_1988 | Tier 1 Level 2 | 30 | | | | | | | **1988 TM REFERENCE, clouds in CVS, completes South CVS** |
| **Landsat 4 MSS** | LM04/C02/T2/LM04_007056_19880416 | 1988 | L4_7_56_1988 | Tier 2 Raw | 60 | | | | | | | **1988 MSS REFERENCE, Clouds northwest GLR, scattered clouds, completes West GLR** |
| **Landsat 4 MSS** | LM04/C02/T2/LM04_007057_19880416 | 1988 | L4_7_57_1988 | Tier 2 Raw | 60 | | | | | | | **1988 MSS REFERENCE, Clouds East visible CHI, completes West corner CHI** |
| Landsat 4 TM | LT04/C02/T1_L2/LT04_008057_19891222 | 1989 | - | Tier 1 Level 2 | 30 | | | | | | | SEC, Peripheral Clouds, better than 88 for south CVS |
| Landsat 5 TM | LT05/C02/T1_L2/LT05_008056_19891230 | 1989 | - | Tier 1 Level 2 | 30 | | | | | | | SEC, Peripheral Clouds, better than 88 for middle RRB |
| **Landsat 5 TM** | LT05/C02/T1_L2/LT05_007056_19910316 | 1991 | L5T_7_56_1991 | Tier 1 Level 2 | 30 | | | | | | | **1991 TM REFERENCE, scattered clouds, completes East IGM ACB and RRB** |
| **Landsat 5 TM** | LT05/C02/T1_L2/LT05_007057_19910316 | 1991 | L5T_7_57_1991 | Tier 1 Level 2 | 30 | | | | | | | **1991 TM REFERENCE, completes East CHI** |
| **Landsat 5 TM** | LT05/C02/T1_L2/LT05_008056_19910323 | 1991 | L5T_8_56_1991 | Tier 1 Level 2 | 30 | | | | | | | **1991 TM REFERENCE, Clouds on west GUE and north IGM** |
| **Landsat 5 TM** | LT05/C02/T1_L2/LT05_008057_19910323 | 1991 | L5T_8_57_1991 | Tier 1 Level 2 | 30 | | | | | | | **1991 TM REFERENCE, Clouds south of CVS on most edges and south of CHI** |
| Landsat 5 TM | LT05/C02/T1_L2/LT05_008056_19980105 | 1998 | - | Tier 1 Level 2 | 30 | | | | | | | Haze on north GUE, scattered clouds all over others |
| **Landsat 5 TM** | LT05/C02/T1_L2/LT05_008056_20010129 | 2001 | L5T_8_56_2001 | Tier 1 Level 2 | 30 | | | | | | | **2001 TM REFERENCE, clouds on west GUE** |

| Mission & Inst. | Directory (LANDSAT/...) | Year | Code | Proc. Level | Res. [m] | PÁRAMO COMPLEX | | | | | | Notes |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | GUE | RRB | CHI | IGM | ACB | CVS | |
| **Landsat 5 TM** | LT05/C02/T1_L2/LT05_008057_20010129 | 2001 | L5T_8_57_2001 | Tier 1 Level 2 | 30 | | | | | | | 2001 TM REFERENCE, clouds on west and east CVS+center haze, some clouds south CHI |
| **Landsat 7 ETM+** | LE07/C02/T1_L2/LE07_007056_20010130 | 2001 | L7E_7_56_2001 | Tier 1 Level 2 | 30 | | | | | | | 2001 ETM+/TM REFERENCE, completes East IGM ACB and RRB |
| **Landsat 7 ETM+** | LE07/C02/T1_L2/LE07_007057_20010130 | 2001 | L7E_7_57_2001 | Tier 1 Level 2 | 30 | | | | | | | 2001 ETM+/TM REFERENCE, completes East CHI and RBB. Anomalous pixels SouthEast CHI |
| Landsat 7 ETM+ | LE07/C02/T1_L2/LE07_008056_20030111 | 2003 | - | Tier 1 Level 2 | 30 | | | | | | | SEC, Clouds on southwest IGM, RRB, CHI. No Gaps |
| Landsat 7 ETM+ | LE07/C02/T1_L2/LE07_008057_20030111 | 2003 | - | Tier 1 Level 2 | 30 | | | | | | | SEC, Clouds on southwest CHI and east west CVS, fire at south CVS. No Gaps |
| **Landsat 7 ETM+** | LE07/C02/T1_L2/LE07_007056_20070131 | 2007 | L7E_7_56_2007A | Tier 1 Level 2 | 30 | | | | | | | 2007 ETM+ REFERENCE, clouds on RRB, completes East IGM ACB and RRB |
| **Landsat 7 ETM+** | LE07/C02/T1_L2/LE07_007057_20070131 | 2007 | L7E_7_57_2007A | Tier 1 Level 2 | 30 | | | | | | | 2007 ETM+ REFERENCE, clouds on RRB and CHI, completes East CHI and RBB |
| **Landsat 7 ETM+** | LE07/C02/T1_L2/LE07_008056_20070207 | 2007 | L7E_8_56_2007A | Tier 1 Level 2 | 30 | | | | | | | 2007 ETM+ REFERENCE, minimum clouds on west GUE |
| **Landsat 7 ETM+** | LE07/C02/T1_L2/LE07_008057_20070207 | 2007 | L7E_8_57_2007A | Tier 1 Level 2 | 30 | | | | | | | 2007 ETM+ REFERENCE, Peripheral clouds on west CVS |
| **Landsat 7 ETM+** | LE07/C02/T1_L2/LE07_007056_20070216 | 2007 | L7E_7_56_2007B | Tier 1 Level 2 | 30 | | | | | | | 2007 ETM+ REFERENCE, clouds on IGM, completes East IGM ACB and RRB |
| **Landsat 7 ETM+** | LE07/C02/T1_L2/LE07_007057_20070216 | 2007 | L7E_7_57_2007B | Tier 1 Level 2 | 30 | | | | | | | 2007 ETM+ REFERENCE, clouds Southeast RRB and CHI, completes East CHI and RBB |
| **Landsat 7 ETM+** | LE07/C02/T1_L2/LE07_008056_20070223 | 2007 | L7E_8_56_2007B | Tier 1 Level 2 | 30 | | | | | | | 2007 ETM+ REFERENCE, Clouds on northeast CHI |
| **Landsat 7 ETM+** | LE07/C02/T1_L2/LE07_008057_20070223 | 2007 | L7E_8_57_2007B | Tier 1 Level 2 | 30 | | | | | | | 2007 ETM+ REFERENCE, Peripheral clouds on west CVS, clouds+haze southeast CVS |
| Landsat 7 ETM+ | LE07/C02/T1_L2/LE07_008056_20091213 | 2009 | - | Tier 1 Level 2 | 30 | | | | | | | SEC, Clouds on west GUE, haze north CHI |
| Landsat 7 ETM+ | LE07/C02/T1_L2/LE07_008057_20091213 | 2009 | - | Tier 1 Level 2 | 30 | | | | | | | SEC, Peripheral Clouds all over CVS, haze north CHI |
| Landsat 7 ETM+ | LE07/C02/T1_L2/LE07_008056_20091229 | 2009 | - | Tier 1 Level 2 | 30 | | | | | | | SEC, Clouds on west GUE, northwest IGM, north CHI |

| Mission & Inst. | Directory (LANDSAT/…) | Year | Code | Proc. Level | Res. [m] | PÁRAMO COMPLEX | | | | | | Notes |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | GUE | RRB | CHI | IGM | ACB | CVS | |
| Landsat 7 ETM+ | LE07/C02/T1_L2/LE07_008057_20091229 | 2009 | - | Tier 1 Level 2 | 30 | | | | | | | SEC, Peripheral Clouds east and west CVS, south CHI |
| Landsat 7 ETM+ | LE07/C02/T1_L2/LE07_008056_20100114 | 2010 | - | Tier 1 Level 2 | 30 | | | | | | | SEC, Scattered clouds in all complexes |
| Landsat 7 ETM+ | LE07/C02/T1_L2/LE07_008057_20100114 | 2010 | - | Tier 1 Level 2 | 30 | | | | | | | SEC, Scattered clouds in all complexes and clouds on both sides of CVS |
| **Landsat 7 ETM+** | **LE07/C02/T1_L2/LE07_008057_20110101** | **2011** | **L7E_8_57_2011B** | **Tier 1 Level 2** | **30** | | | | | | | **2011 ETM+ REFERENCE, scattered clouds in all Páramos, completes gaps of CVS** |
| **Landsat 7 ETM+** | **LE07/C02/T1_L2/LE07_007056_20110110** | **2011** | **L7E_7_56_2011A** | **Tier 1 Level 2** | **30** | | | | | | | **2011 ETM+ REFERENCE, clouds in RRB, completes East IGM and ACB** |
| **Landsat 7 ETM+** | **LE07/C02/T1_L2/LE07_008056_20110117** | **2011** | **L7E_8_56_2011A** | **Tier 1 Level 2** | **30** | | | | | | | **2011 ETM+ REFERENCE, Clouds on west GUE, northwest IGM** |
| **Landsat 7 ETM+** | **LE07/C02/T1_L2/LE07_008057_20110117** | **2011** | **L7E_8_57_2011A** | **Tier 1 Level 2** | **30** | | | | | | | **2011 ETM+ REFERENCE, Peripheral Clouds east and west CVS, south CHI** |
| **Landsat 7 ETM+** | **LE07/C02/T1_L2/LE07_007056_20110126** | **2011** | **L7E_7_56_2011B** | **Tier 1 Level 2** | **30** | | | | | | | **2011 ETM+ REFERENCE, Clouds Southeast IGM, completes gaps of IGM and ACB** |
| **Landsat 7 ETM+** | **LE07/C02/T1_L2/LE07_007057_20110126** | **2011** | **L7E_7_57_2011** | **Tier 1 Level 2** | **30** | | | | | | | **2011 ETM+ REFERENCE, Clouds East CHI, completes gaps of RRB and CHI** |
| **Landsat 7 ETM+** | **LE07/C02/T1_L2/LE07_008056_20110202** | **2011** | **L7E_8_56_2011B** | **Tier 1 Level 2** | **30** | | | | | | | **2011 ETM+ REFERENCE, Clouds on east and southwest GUE, completes gaps for all rest** |
| Landsat 7 ETM+ | LE07/C02/T1_L2/LE07_008056_20120221 | 2012 | - | Tier 1 Level 2 | 30 | | | | | | | SEC, Scattered clouds in all complexes |
| Landsat 7 ETM+ | LE07/C02/T1_L2/LE07_008057_20120221 | 2012 | - | Tier 1 Level 2 | 30 | | | | | | | SEC, Scattered clouds in all complexes |
| Landsat 7 ETM+ | LE07/C02/T1_L2/LE07_008056_20130106 | 2013 | - | Tier 1 Level 2 | 30 | | | | | | | SEC, clouds in all complexes except IGM |
| Landsat 7 ETM+ | LE07/C02/T1_L2/LE07_008056_20130327 | 2013 | - | Tier 1 Level 2 | 30 | | | | | | | SEC, Clouds on west GUE, northwest IGM, north CHI |
| Landsat 7 ETM+ | LE07/C02/T1_L2/LE07_008057_20130327 | 2013 | - | Tier 1 Level 2 | 30 | | | | | | | SEC, Clouds almost all over CHI, peripheral east and west for CVS |
| Landsat 8 OLI | LC08/C02/T1_L2/LC08_008056_20140101 | 2014 | - | Tier 1 Level 2 | 30 | | | | | | | SEC, some clouds on west GUE, scattered clouds in all others, except for IGM |

| Mission & Inst. | Directory (LANDSAT/…) | Year | Code | Proc. Level | Res. [m] | PÁRAMO COMPLEX | | | | | | Notes |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | GUE | RRB | CHI | IGM | ACB | CVS | |
| Landsat 8 OLI | LC08/C02/T1_L2/LC08_008057_20140101 | 2014 | - | Tier 1 Level 2 | 30 | | | | | | | SEC, Clouds almost all over CHI, clouds east and center west for CVS |
| **Landsat 8 OLI** | **LC08/C02/T1_L2/LC08_007056_20141228** | **2014** | **L8O_7_56_2014** | **Tier 1 Level 2** | **30** | | | | | | | **2015 OLI+TIRS REFERENCE, Scattered clouds all over RRB and ACB, IGM almost cloud free.** |
| **Landsat 8 OLI** | **LC08/C02/T1_L2/LC08_008056_20150104** | **2015** | **L8O_8_56_2015A** | **Tier 1 Level 2** | **30** | | | | | | | **2015 OLI+TIRS REFERENCE, Scattered clouds Southwest GUE and CHI, clouds South RRB** |
| **Landsat 8 OLI** | **LC08/C02/T1_L2/LC08_008057_20150104** | **2015** | **L8O_8_57_2015A** | **Tier 1 Level 2** | **30** | | | | | | | **2015 OLI+TIRS REFERENCE, Scattered clouds all over CHI, peripheral and scattered East and Southwest CVS** |
| **Landsat 8 OLI** | **LC08/C02/T1_L2/LC08_007056_20150113** | **2015** | **L8O_7_56_2015** | **Tier 1 Level 2** | **30** | | | | | | | **2015 OLI+TIRS REFERENCE, Scattered clouds all over RRB and ACB, and south IGM. Better for ACB than Dec 2014** |
| **Landsat 8 OLI** | **LC08/C02/T1_L2/LC08_007057_20150113** | **2015** | **L8O_7_57_2015** | **Tier 1 Level 2** | **30** | | | | | | | **2015 OLI+TIRS REFERENCE, Scattered clouds all over RRB and CHI. Completes East CHI** |
| **Landsat 8 OLI** | **LC08/C02/T1_L2/LC08_008056_20150221** | **2015** | **L8O_8_56_2015B** | **Tier 1 Level 2** | **30** | | | | | | | **2015 OLI+TIRS REFERENCE, Scattered clouds West GUE and IGM, clouds south RRB and north CHI** |
| **Landsat 8 OLI** | **LC08/C02/T1_L2/LC08_008057_20150221** | **2015** | **L8O_8_57_2015B** | **Tier 1 Level 2** | **30** | | | | | | | **2015 OLI+TIRS REFERENCE, Scattered clouds centre southeast CHI, peripheral and scattered East Southwest CVS. Better than Jan for CHI** |
| Landsat 7 ETM+ | LE07/C02/T1_L2/LE07_008056_20160115 | 2016 | | Tier 1 Level 2 | 30 | | | | | | | SEC, Clouds on west GUE, scattered clouds north west IGM, north CHI |
| Landsat 7 ETM+ | LE07/C02/T1_L2/LE07_008057_20160115 | 2016 | | Tier 1 Level 2 | 30 | | | | | | | SEC, Scattered clouds all over CHI, peripheral and scattered east and west CVS |
| Landsat 8 OLI | LC08/C02/T1_L2/LC08_008056_20160123 | 2016 | - | Tier 1 Level 2 | 30 | | | | | | | SEC, clouds in all complexes except ACB |
| Landsat 8 OLI | LC08/C02/T1_L2/LC08_008057_20160123 | 2016 | - | Tier 1 Level 2 | 30 | | | | | | | SEC, Scattered clouds in different location CHI, clouds west and south+peripheral east CVS |
| **Landsat 8 OLI** | **LC08/C02/T1_L2/LC08_008056_20181230** | **2018** | **L8O_8_56_2019** | **Tier 1 Level 2** | **30** | | | | | | | **2019 OLI+TIRS REFERENCE, Scattered clouds southeast and west GUE, clouds east IGM** |
| **Landsat 8 OLI** | **LC08/C02/T1_L2/LC08_008057_20181230** | **2018** | **L8O_8_57_2019** | **Tier 1 Level 2** | **30** | | | | | | | **2019 OLI+TIRS REFERENCE, Cloud southeast CHI, peripheral+internal scattered clouds in CVS** |
| **Landsat 8 OLI** | **LC08/C02/T1_L2/LC08_007056_20190225** | **2019** | **L8O_7_56_2019** | **Tier 1 Level 2** | **30** | | | | | | | **2019 OLI+TIRS REFERENCE, Scattered clouds all over RRB. Completes East IGM and ACB** |

| Mission & Inst. | Directory (LANDSAT/...) | Year | Code | Proc. Level | Res. [m] | PÁRAMO COMPLEX | | | | | | Notes |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | GUE | RRB | CHI | IGM | ACB | CVS | |
| **Landsat 8 OLI** | **LC08/C02/T1_L2/LC08_007057_20190225** | **2019** | **L8O_7_57_2019** | **Tier 1 Level 2** | **30** | | | | | | | **2019 OLI+TIRS REFERENCE, Scattered clouds all over RRB and East CHI. Completes East CHI** |
| Landsat 9 OLI | LC09/C02/T1_L2/LC09_008056_20211113 | 2021 | - | Tier 1 Level 2 | 30 | | | | | | | Cloud+haze southwest GUE, east CHI, center south TBM. Path moved East |
| Landsat 9 OLI | LC09/C02/T1_L2/LC09_008057_20211113 | 2021 | - | Tier 1 Level 2 | 30 | | | | | | | Clouds all over CVS, cloud west and east CHI+ scattered clouds north. Path moved East |
| Landsat 9 OLI | LC09/C02/T1_L2/LC09_008057_20211214 | 2021 | - | Tier 1 Level 2 | 30 | | | | | | | Scattered clouds in all complexes, but better the nov/21 for south GUE and CVS |
| Landsat 9 OLI | LC09/C02/T1_L2/LC09_008056_20220131 | 2022 | - | Tier 1 Level 2 | 30 | | | | | | | Clouds in almost all IGM, northwest GUE, scattered east RRB |
| Landsat 9 OLI | LC09/C02/T1_L2/LC09_008057_20220131 | 2022 | - | Tier 1 Level 2 | 30 | | | | | | | Clouds south CHI, peripheral west+ south southeast in CVS |
| Landsat 9 OLI | LC09/C02/T1_L2/LC09_008056_20230203 | 2023 | - | Tier 1 Level 2 | 30 | | | | | | | Haze on center GUE, north CHI, on ACB, and southeast RRB |
| Landsat 9 OLI | LC09/C02/T1_L2/LC09_008057_20230203 | 2023 | - | Tier 1 Level 2 | 30 | | | | | | | Haze on north CHI, Peripheral clouds on CVS |

# A.4.  Sentinel-2 Data Pool Database

Supplementary table for Sentinel-2 Data Pool. Black for 0% Proportion of Páramo Complex Area Covered in Scene.

Páramo Complex Area Cloud Cover Qualitative Rating

Very Little to None                                                                                                    Mostly Covered

| Mission & Inst. | GEE Directory (COPERNICUS/S2_SR/...) | Tiles | Collection Reducer | Year | Code | Proc. Level | Res. [m] | GUE | RRB | CHI | IGM | ACB | CVS | Notes |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Sentinel-2 MSI | 20181225T152631_20181225T153106_T18N | WJ, WK,WL,WM, XJ,XK,XL,XM, YM | Median | 2018 | **S2A_2019_8A** | Level-2A | 10 | | | | | | | **2019 MSI REFERENCE. Image Collection equivalent to Landsat Path 8. Haze on CVS** |
| Sentinel-2 MSI | 20181230T152639_20181230T152636_T18N | WJ, WK,WL,WM, XJ,XK,XL,XM, YM | Median | 2018 | S2A_2019_8B | Level-2A | 10 | | | | | | | Image Collection equivalent to Landsat Path 8 |
| Sentinel-2 MSI | 20190101T151701_20190101T151658_T18N | XJ, XK,XL,XM, YJ,YK,YL,YM | Median | 2019 | S2A_2019_7C | Level-2A | 10 | | | | | | | Image Collection equivalent to Landsat Path 7 |
| Sentinel-2 MSI | 20190215T151709_20190215T151703_T18N | XJ, XK,XL,XM, YJ,YK,YL,YM | Median | 2019 | **S2A_2019_7A** | Level-2A | 10 | | | | | | | **2019 MSI REFERENCE. Image Collection equivalent to Landsat Path 7** |
| Sentinel-2 MSI | 20190225T151659_20190225T151702_T18N | XJ, XK,XL,XM, YJ,YK,YL,YM | Median | 2019 | S2A_2019_7B | Level-2A | 10 | | | | | | | Image Collection equivalent to Landsat Path 7 |
| Sentinel-2 MSI | 20240123T152649_20240123T152651_T18N | WJ, WK,WL,WM, XJ,XK,XL,XM, YM | Median | 2024 | **S2A_2024_8A** | Level-2A | 10 | | | | | | | **2024 MSI REFERENCE. Image Collection equivalent to Landsat Path 8** |
| Sentinel-2 MSI | 20240125T151711_20240125T152031_T18N | XJ,XK,XL,XM, YJ,YK,YL,YM | Median | 2024 | **S2A_2024_7A** | Level-2A | 10 | | | | | | | **2024 MSI REFERENCE. Image Collection equivalent to Landsat Path 7** |

# A.5. Pre-processing Pipelines

Supplementary table for different sensor pre-processing pipelines.

| | MSS Pre-Processing Pipelines in GEE Script | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Scene Code | Image Shift (x[m],y[m], source) | DN to TOA_SR (source, type) | Co-Registering (resample type) | Compositing (metric) | Cloud Masking Func. (<Threshold; Buffer[px]; Shadow Dist.[m]) | Other Masking | Final Mosaic Order (Last on Top) | Mosaic Code | Notes | Assets | Co-Registering Reference (resample type, Offset[m], Stiffness) | Assets |
| L2_8_56_1977 | - | (Metadata, linear transform) | Overlap 56_57, Reference Image (bicubic) | - | (0,22; 4; 360) | Overlap 56_57, Elevation MERIT>2500 | 2 | L2_1977_TOA | Single date TOA Mosaic (07-01-1977) Reference for epoch **1977** | L2_1977_TOA_ACB_A L2_1977_TOA_CHI L2_1977_TOA_CVS L2_1977_TOA_GUE L2_1977_TOA_IGM L2_1977_TOA_RRB | L4_1988_TOA_ACB_A (bicubic; 300; 6) L4_1988_TOA_CHI (bicubic; 720; 1) L4_1988_TOA_CVS (bicubic; 1440; 1) L4_1988_TOA_GUE (bicubic; 300; 4) L4_1988_TOA_IGM (bicubic; 300; 6) L4_1988_TOA_RRB (bicubic; 300; 6) | L2_1977_TOA_ACB_A_rs L2_1977_TOA_CHI_rs L2_1977_TOA_CVS_rs L2_1977_TOA_GUE_rs L2_1977_TOA_IGM_rs L2_1977_TOA_RRB_rs |
| L2_8_56_57_1977 | - | - | Overlap 56_57, Resulting Image | - | (0,20; 4; 360) | Elevation MERIT>2500 | 3 | | | | | |
| L2_8_57_1977 | (-3240, 3840, control points) | (Metadata, linear transform) | Overlap 56_57, Register Image (bicubic) | Composite_57_58 (min TOA_SR) | (0,20; 4; 360) | Elevation MERIT>2500 | 1 | | | | | |
| L2_8_58_1977 | (-3120, 3480, control points) | (Metadata, linear transform) | - | | (0,20; 4; 360) | Elevation MERIT>2500 | 1 | | | | | |

160

| Scene Code | Image Shift (x[m],y[m], source) | DN to TOA_SR (source, type) | Co-Registering (resample type) | Compositing (metric) | Cloud Masking Func. (<Threshold; Buffer[px]; Shadow Dist.[m]) | Other Masking | Final Mosaic Order (Last on Top) | Mosaic Code | Notes | Assets | Co-Registering Reference (resample type, Offset[m], Stiffness) | Assets |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | MSS Pre-Processing Pipelines in GEE Script | | | | | | | |
| L4_8_56_1988 | - | (Metadata, linear transform) | - | - | (0,18; 3; 360) | Edge Dark Pixel Removal (Scene -180m buffer[3px]), Elevation MERIT>2500 | 5 | | | | | |
| L4_8_57_1988 | (-60, -60, control points) | (Metadata, linear transform) | - | - | (0,19; 3; 360) | Edge Dark Pixel Removal (Scene -180m buffer[3px]), Elevation MERIT>2500 | 4 | | | | | |
| L4_8_58_1988 | (-360, -240, control points) | (Metadata, linear transform) | - | - | (0,19; 3; 360) | Elevation MERIT>2500 | 3 | L4_1988_TOA | Double date TOA Mosaic (22-03-1988, 16-04-1988) Reference for epoch 1988 | L4_1988_TOA_ACB_A L4_1988_TOA_ACB_B L4_1988_TOA_CHI L4_1988_TOA_CVS L4_1988_TOA_GUE L4_1988_TOA_IGM L4_1988_TOA_RRB | | |
| L4_7_56_1988 | - | (Metadata, linear transform) | - | - | (0,18; 3; 300) | Elevation MERIT>2500 | 2 | | | | | |
| L4_7_57_1988 | - | (Metadata, linear transform) | - | - | (0,18; 3; 300) | Elevation MERIT>2500 | 1 | | | | | |

| | TM Pre-Processing Pipelines in GEE Script | | | | | | |
|---|---|---|---|---|---|---|---|
| **Code** | **Cloud Masking Func. (QA_Band; Bits used)** | **Other Masking** | **Colour Correction by Histogram Matching** | **Final Mosaic Order (Last on Top)** | **Mosaic Code** | **Notes** | **Assets** |
| L4T_8_56_1987 | ('QA_PIXEL'; 1,3 and 4) | Elevation MERIT>2500 | Target 1 | 1 | L4T_1988 | Three dates SR Mosaic (17-12-1987, 22-03-1988, 16-04-1988) Reference for epoch 1988T | L4T_1988_ACB_A L4T_1988_ACB_B L4T_1988_CHI L4T_1988_CVS L4T_1988_GUE L4T_1988_IGM L4T_1988_RRB |
| L4T_8_56_1988 | ('QA_PIXEL'; 1,3 and 4) | Elevation MERIT>2500 | Reference 1 & 2 | 6 | | | |
| L4T_8_57_1988 | ('QA_PIXEL'; 1,3 and 4) | Elevation MERIT>2500 | - | 5 | | | |
| L4T_8_58_1988 | ('QA_PIXEL'; 1,3 and 4) | Elevation MERIT>2500 | - | 4 | | | |
| L4T_7_56_1988 | ('QA_PIXEL'; 1,3 and 4) | Elevation MERIT>2500 | Target 2 | 3 | | | |
| L4T_7_57_1988 | ('QA_PIXEL'; 1,3 and 4) | Elevation MERIT>2500 | - | 2 | | | |
| L5T_7_56_1991 | ('QA_PIXEL'; 1,3 and 4) | Elevation MERIT>2500 | Target 1 | 4 | L5T_1991 | Double date SR Mosaic (16-03-1991, 23-03-1991) Reference for epoch 1991T | L5T_1991_ACB_A L5T_1991_ACB_B L5T_1991_CHI L5T_1991_CVS L5T_1991_GUE L5T_1991_IGM L5T_1991_RRB |
| L5T_7_57_1991 | ('QA_PIXEL'; 1,3 and 4) | Elevation MERIT>2500 | - | 3 | | | |
| L5T_8_56_1991 | ('QA_PIXEL'; 1,3 and 4) | Edge Uneven Anomalous Pixel Removal (Scene -360m buffer[6px]), Elevation MERIT>2500 | Reference 1 | 2 | | | |
| L5T_8_57_1991 | ('QA_PIXEL'; 1,3 and 4) | Edge Uneven Anomalous Pixel Removal (Scene -360m buffer[6px]), Elevation MERIT>2500 | - | 1 | | | |
| L5T_8_56_2001 | ('QA_PIXEL'; 1,3 and 4) | Elevation MERIT>2500 | - | 4 | L5T_2001 | Single date SR Mosaic (29-01-2001) Reference for epoch 2001T | L5T_2001_ACB_A L5T_L7E_2001_ACB_B L5T_L7E_2001_CHI L5T_2001_CVS L5T_2001_GUE L5T_L7E_2001_IGM L5T_L7E_2001_RRB |
| L5T_8_57_2001 | ('QA_PIXEL'; 1,3 and 4) | Elevation MERIT>2500 | - | 3 | | | |
| L7E_7_56_2001 | ('QA_PIXEL'; 1,3 and 4) | Elevation MERIT>2500 | - | 2 | L5T_L7E_2001 | Double date Double sensor SR Mosaic (29-01-2001, 30-01-2001) Reference for epoch 2001T | |
| L7E_7_57_2001 | ('QA_PIXEL'; 1,3 and 4) | Elevation MERIT>2500, Sensor Anomalous Pixel Rows Removal (Manually drawn polygons) | - | 1 | | | |

| | ETM+ Pre-Processing Pipelines in GEE Script | | | | | | |
|---|---|---|---|---|---|---|---|
| Code | Cloud Masking Func. (QA_Band; Bits used) | Other Masking | Colour Correction by Histogram Matching | Final Mosaic Order (Last on Top) | Mosaic Code | Notes | Assets |
| L7E_7_56_2007A | ('QA_PIXEL'; 1,3 and 4) | Elevation MERIT>2500 | Reference 2 | 4 | L7E_2007 | Four dates SR Mosaic (2007-01-31, 2007-02-07, 2007-02-16, 2007-02-23) Reference for epoch 2007E | L7E_2007_ACB_A L7E_2007_ACB_B L7E_2007_CHI L7E_2007_CVS L7E_2007_GUE L7E_2007_IGM L7E_2007_RRB |
| L7E_7_57_2007A | ('QA_PIXEL'; 1,3 and 4) | Elevation MERIT>2500 | Reference 1 | 2 | | | |
| L7E_8_56_2007A | ('QA_PIXEL'; 1,3 and 4) | Elevation MERIT>2500 | Reference 3 | 8 | | | |
| L7E_8_57_2007A | ('QA_PIXEL'; 1,3 and 4) | Elevation MERIT>2500 | Reference 4 | 6 | | | |
| L7E_7_56_2007B | ('QA_PIXEL'; 1,3 and 4) | Elevation MERIT>2500 | Target 2 | 3 | | | |
| L7E_7_57_2007B | ('QA_PIXEL'; 1,3 and 4) | Elevation MERIT>2500 | Target 1 | 1 | | | |
| L7E_8_56_2007B | ('QA_PIXEL'; 1,3 and 4) | Elevation MERIT>2500 | Target 3 | 7 | | | |
| L7E_8_57_2007B | ('QA_PIXEL'; 1,3 and 4) | Elevation MERIT>2500 | Target 4 | 5 | | | |
| L7E_8_57_2011B | ('QA_PIXEL'; 1,3 and 4) | Elevation MERIT>2500 | Target 2 | 4 | L7E_2011 | Five dates SR Mosaic (2011-01-01, 2011-01-10, 2011-01-17, 2011-01-26, 2011-02-02) Reference for epoch 2011E | L7E_2011_ACB_A L7E_2011_ACB_B L7E_2011_CHI L7E_2011_CVS L7E_2011_GUE L7E_2011_IGM L7E_2011_RRB |
| L7E_7_56_2011B | ('QA_PIXEL'; 1,3 and 4) | Elevation MERIT>2500 | Target 1 | 2 | | | |
| L7E_8_56_2011A | ('QA_PIXEL'; 1,3 and 4) | Elevation MERIT>2500 | Reference 3 | 7 | | | |
| L7E_8_57_2011A | ('QA_PIXEL'; 1,3 and 4) | Elevation MERIT>2500 | Reference 2 | 5 | | | |
| L7E_7_56_2011A | ('QA_PIXEL'; 1,3 and 4) | Elevation MERIT>2500 | Reference 1 | 3 | | | |
| L7E_7_57_2011 | ('QA_PIXEL'; 1,3 and 4) | Elevation MERIT>2500 | - | 1 | | | |
| L7E_8_56_2011B | ('QA_PIXEL'; 1,3 and 4) | Elevation MERIT>2500 | Target 3 | 6 | | | |

| | OLI+TIRS Pre-Processing Pipelines in GEE Script | | | | | | |
|---|---|---|---|---|---|---|---|
| **Code** | **Cloud Masking Func. (QA_Band; Bits used)** | **Other Masking** | **Colour Correction by Histogram Matching** | **Final Mosaic Order (Last on Top)** | **Mosaic Code** | **Notes** | **Assets** |
| **L8O_7_56_2014** | ('QA_PIXEL'; 1,3 and 4) | Elevation MERIT>2500 | Target 1 | 3 | **L8O_2015** | Four dates SR Mosaic (2014-12-28, 2015-01-04, 2015-01-13, 2015-02-21) Reference for epoch **2015O** | L8O_2015_ACB_A L8O_2015_ACB_B L8O_2015_CHI L8O_2015_CVS L8O_2015_GUE L8O_2015_IGM L8O_2015_RRB |
| **L8O_8_56_2015A** | ('QA_PIXEL'; 1,3 and 4) | Elevation MERIT>2500 | Reference 1, 2 & 3 | 7 | | | |
| **L8O_8_57_2015A** | ('QA_PIXEL'; 1,3 and 4) | Elevation MERIT>2500 | Reference 4 | 5 | | | |
| **L8O_7_56_2015** | ('QA_PIXEL'; 1,3 and 4) | Elevation MERIT>2500 | Target 2 | 2 | | | |
| **L8O_7_57_2015** | ('QA_PIXEL'; 1,3 and 4) | Elevation MERIT>2500 | - | 1 | | | |
| **L8O_8_56_2015B** | ('QA_PIXEL'; 1,3 and 4) | Elevation MERIT>2500 | Target 3 | 6 | | | |
| **L8O_8_57_2015B** | ('QA_PIXEL'; 1,3 and 4) | Elevation MERIT>2500 | Target 4 | 4 | | | |
| **L8O_8_56_2019** | ('QA_PIXEL'; 1,3 and 4) | Elevation MERIT>2500 | Reference 1 | 4 | **L8O_2019** | Double date SR Mosaic (30-12-2018, 25-02-2019) Reference for epoch **2019O** | L8O_2019_ACB_A L8O_2019_ACB_B L8O_2019_CHI L8O_2019_CVS L8O_2019_GUE L8O_2019_IGM L8O_2019_RRB |
| **L8O_8_57_2019** | ('QA_PIXEL'; 1,3 and 4) | Elevation MERIT>2500 | - | 3 | | | |
| **L8O_7_56_2019** | ('QA_PIXEL'; 1,3 and 4) | Elevation MERIT>2500 | Target 1 | 2 | | | |
| **L8O_7_57_2019** | ('QA_PIXEL'; 1,3 and 4) | Elevation MERIT>2500 | - | 1 | | | |

| | MSI Pre-Processing Pipelines in GEE Script | | | | | | |
|---|---|---|---|---|---|---|---|
| Code | Cloud Masking Func. (Dataset; Threshold) | Other Masking | Colour Correction by Histogram Matching | Final Mosaic Order (Last on Top) | Mosaic Code | Notes | Assets |
| S2A_2018_8A | (S2_CLOUD_PROBABILITY;<40%) | Elevation MERIT>2500 | Reference | 2 | S2A_2019 | Double date SR Mosaic (25-12-2018, 15-02-2019) Reference for epoch **2019M** | S2A_2019_ACB_A S2A_2019_ACB_B S2A_2019_CHI S2A_2019_CVS S2A_2019_GUE S2A_2019_IGM S2A_2019_RRB |
| S2A_2019_7A | (S2_CLOUD_PROBABILITY;<30%) | Elevation MERIT>2500 | Target | 1 | | | |
| S2A_2024_8A | (S2_CLOUD_PROBABILITY;<40%) | Elevation MERIT>2500 | Reference | 2 | S2A_2024 | Double date SR Mosaic (23-01-2024, 25-01-2024) Reference for epoch **2024M** | S2A_2024_ACB_A S2A_2024_ACB_B S2A_2024_CHI S2A_2024_CVS S2A_2024_GUE S2A_2024_IGM S2A_2024_RRB |
| S2A_2024_7A | (S2_CLOUD_PROBABILITY;<40%) | Elevation MERIT>2500 | Target | 1 | | | |

# B   Appendix B

This Appendix includes the classification Sample areas interpretation, their spectral signatures and LULC Maps not included in Chapter 4.

## B.1.   Sample Areas Interpretation

Supplementary figures for different class sample areas interpretation.

GUE *Neusa* reservoir close-up sample Areas (1977-1988 NIR composite, 1988T-2024M AGRI composite)

RRB *Páramo de Rabanal* close-up sample Areas (1977-1988 NIR composite, 1988T-2024M AGRI composite)

| Class Name | Class Code |
|---|---|
| Water | 0 / W |
| Bare-Rock-Built | 1 / B |
| Agriculture-Grass | 2 / A |
| Forest | 3 / F |
| Shrubland-Wetland | 4 / S |

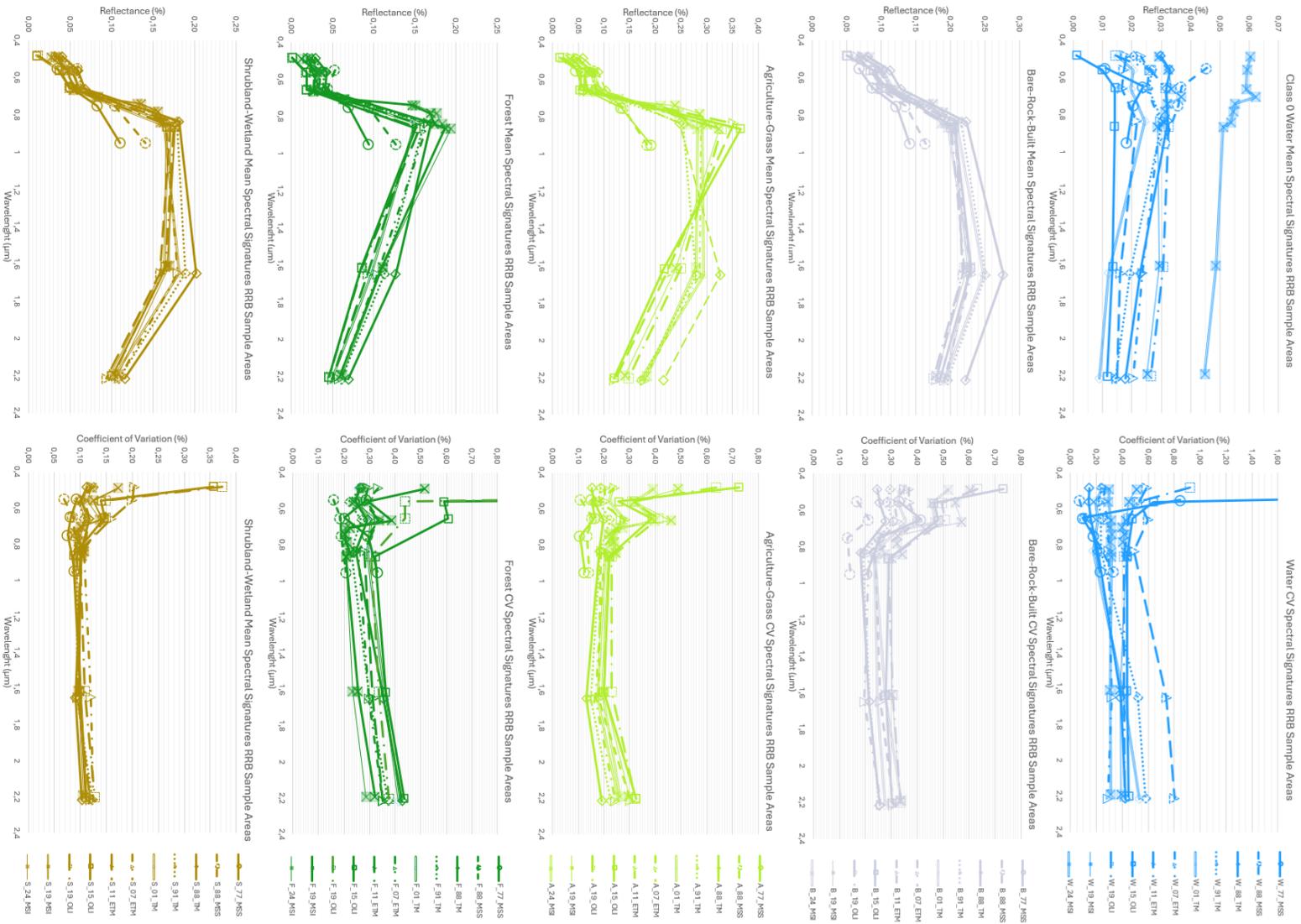| Class Name | Class Code |
|---|---|
| Water | 0 / W |
| Bare-Rock-Built | 1 / B |
| Agriculture-Grass | 2 / A |
| Forest | 3 / F |
| Shrubland-Wetland | 4 / S |

GUE Sample Water class stable area (1977-1988 NIR composite, 1988T-2024M AGRI composite, Google Earth RGB natural):

RRB Sample Bare-Rock-Built class stable area (1977-1988 NIR composite, 1988T-2024M AGRI composite, Google Earth RGB natural):

GUE Sample Agriculture-Grass class stable area (1977-1988 NIR composite, 1988T-2024M AGRI composite, Google Earth RGB natural):

RRB Sample Forest class stable area (1977-1988 NIR composite, 1988T-2024M AGRI composite, Google Earth RGB natural):

GUE Sample Shrubland-Wetland class stable area (1977-1988 NIR composite, 1988T-2024M AGRI composite, Google Earth RGB natural):

# B.2. Sample Areas Spectral Signatures

## B.3.  Epoch Error Matrices & Results Summary

**GUE 1988**

Sample counts error matrix and estimated area proportions error matrix

| | Map | 0 / W | 1 / B | 2 / A | 3 / F | 4 / S | Total |
|---|---|---|---|---|---|---|---|
| | | | | Test dataset | | | |
| Classified | 0 / W | 56 | 0 | 0 | 0 | 0 | 56 |
| | 1 / B | 0 | 90 | 2 | 0 | 0 | 92 |
| | 2 / A | 0 | 13 | 1656 | 29 | 110 | 1808 |
| | 3 / F | 0 | 0 | 16 | 536 | 101 | 653 |
| | 4 / S | 0 | 0 | 18 | 32 | 191 | 241 |
| | Total | 56 | 103 | 1692 | 597 | 402 | 2850 |

| | Map | 0 / W | 1 / B | 2 / A | 3 / F | 4 / S | Total% |
|---|---|---|---|---|---|---|---|
| | | | | Test dataset | | | |
| Classified | 0 / W | **1,1** | 0,0 | 0,0 | 0,0 | 0,0 | 1,1 |
| | 1 / B | 0,0 | **4,7** | 0,1 | 0,0 | 0,0 | 4,8 |
| | 2 / A | 0,0 | 0,4 | **51,4** | 0,9 | 3,4 | 56,1 |
| | 3 / F | 0,0 | 0,0 | 0,5 | **16,2** | 3,1 | 19,8 |
| | 4 / S | 0,0 | 0,0 | 1,4 | 2,4 | **14,4** | 18,1 |
| | Total% | 1,1 | 5,1 | 53,4 | 19,6 | 20,8 | 100 |

Result summary:

| 20 Seed Mode | | 0 / W | 1 / B | 2 / A | 3 / F | 4 / S | | | |
|---|---|---|---|---|---|---|---|---|---|
| **GUE 1988** | Total% | 1,12 | 4,47 | 53,50 | 20,15 | 20,76 | | Ô | **88,23%** |
| | (SE%) | 0,00 | 0,10 | 0,49 | 0,52 | 0,63 | | (SE) | **0,664%** |
| | Prod% | 100,00 | 95,81 | 96,67 | 82,88 | 69,38 | | | |
| | (SE%) | 0,00 | 1,75 | 0,62 | 3,07 | 6,98 | | | |
| | User% | 100,00 | 97,83 | 91,65 | 83,15 | 80,08 | | | |
| | (SE%) | 0,00 | 1,53 | 0,65 | 1,47 | 2,58 | | | |
| | F-score | 100,00 | 95,74 | 94,71 | 86,05 | 60,41 | | | |
| Mode Pixel count | Area [Ha] | 2215,63 | 8630,81 | 111380,82 | 39643,63 | 35494 | | Total | 197364,89 |
| Accuracy Derived | Âⱼ [Ha] | 2215,63 | 8812,81 | 105593,98 | 39772,98 | 40969,49 | | Total | 197364,89 |
| | (SE)[Ha] | 0,00 | 199,79 | 961,23 | 1032,32 | 1235,22 | | | |
| Mean Pixel count | Area [Ha] | 2207,42 | 8776,01 | 110837,50 | _39512,95_ | 36031,02 | | Total | 197364,90 |
| | (SE)[Ha]* | 5,50 | 69,54 | 174,87 | 97,47 | 158,27 | | | |
| **ACB_A 1988** | | | | | | | | | |
| Mode Pixel count | Area [Ha] | 265,05 | 1744,40 | 15190,54 | 1689,64 | 3201,03 | | Total | 22090,66 |
| | Area% | 1,20 | 7,90 | 68,76 | 7,65 | 14,49 | | | |
| Mean Pixel count | Area [Ha] | 264,46 | _1796,87_ | 15060,65 | 1683,47 | 3285,22 | | Total | 22090,66 |
| | (SE)[Ha]* | 1,46 | 15,57 | 23,99 | 6,78 | 22,58 | | | |
| **GUE 1988 PA** | | | | | | | | | |
| Mode Pixel count | Area [Ha] | 63,71 | 28,18 | 12137,33 | 15486,76 | 15506,46 | | Total | 43222,44 |
| | Area% | 0,15 | 0,07 | 28,08 | 35,83 | 35,88 | | | |
| Mean Pixel count | Area [Ha] | _60,88_ | 30,20 | 12161,71 | 15438,34 | 15531,31 | | Total | 43222,45 |
| | (SE)[Ha]* | 2,90 | 0,83 | 43,37 | 36,34 | 40,80 | | | |
| **ACB_A 1988 PA** | | | | | | | | | |
| Mode Pixel count | Area [Ha] | 0,00 | 3,49 | 897,34 | 405,06 | 547,05 | | Total | 1852,94 |
| | Area% | 0,00 | 0,19 | 48,43 | 21,86 | 29,52 | | | |
| Mean Pixel count | Area [Ha] | 0,00 | 4,12 | 892,18 | 405,89 | 550,74 | | Total | 1852,93 |
| | (SE)[Ha]* | 0,00 | 0,28 | 3,05 | 1,00 | 3,09 | | | |

*Pixel count SE was calculated as $SE = \sigma/\sqrt{n}$, where $\sigma$ is the standard deviation and $n$, the number of iterations.
Underlined results in italics indicate that the 20 iteration results were not normally distributed.

**GUE 1988T**

Sample counts error matrix and estimated area proportions error matrix

|  |  | Test dataset | | | | | |
|---|---|---|---|---|---|---|---|
| **Map** | | **0 / W** | **1 / B** | **2 / A** | **3 / F** | **4 / S** | **Total** |
| **0 / W** | | 259 | 0 | 0 | 0 | 0 | 259 |
| **1 / B** | | 0 | 338 | 1 | 0 | 0 | 339 |
| **2 / A** | | 0 | 84 | 7171 | 54 | 90 | 7399 |
| **3 / F** | | 0 | 0 | 42 | 2378 | 183 | 2603 |
| **4 / S** | | 0 | 0 | 16 | 140 | 804 | 960 |
| **Total** | | 259 | 422 | 7230 | 2572 | 1077 | 11560 |

|  |  | Test dataset | | | | | |
|---|---|---|---|---|---|---|---|
| **Map** | | **0 / W** | **1 / B** | **2 / A** | **3 / F** | **4 / S** | **Total%** |
| **0 / W** | | **1,2** | 0,0 | 0,0 | 0,0 | 0,0 | 1,2 |
| **1 / B** | | 0,0 | **6,1** | 0,0 | 0,0 | 0,0 | 6,1 |
| **2 / A** | | 0,0 | 0,7 | **58,4** | 0,4 | 0,7 | 60,3 |
| **3 / F** | | 0,0 | 0,0 | 0,4 | **19,8** | 1,5 | 21,7 |
| **4 / S** | | 0,0 | 0,0 | 0,2 | 1,6 | **9,0** | 10,8 |
| **Total%** | | 1,2 | 6,8 | 58,9 | 21,9 | 11,3 | 100,0 |

Result summary:

| 20 Seed Mode |  | 0 / W | 1 / B | 2 / A | 3 / F | 4 / S |  |  |  |
|---|---|---|---|---|---|---|---|---|---|
| **GUE 1988T** | Total% | 1,17 | 6,76 | 58,94 | 21,85 | 11,28 |  | **Ô** | **94,50%** |
|  | (SE%) | 0,00 | 0,08 | 0,14 | 0,18 | 0,18 |  | **(SE)** | **0,214%** |
|  | Prod% | 100,00 | 89,88 | 99,07 | 90,80 | 79,97 |  |  |  |
|  | (SE%) | 0,00 | 1,02 | 0,12 | 0,76 | 2,47 |  |  |  |
|  | User% | 100,00 | 99,71 | 96,92 | 91,36 | 83,75 |  |  |  |
|  | (SE%) | 0,00 | 0,29 | 0,20 | 0,55 | 1,19 |  |  |  |
|  | F-score | 100,00 | 88,83 | 98,04 | 91,90 | 78,94 |  |  |  |
| Mode Pixel count | Area [Ha] | 2318,36 | 12084,55 | 119524,8 | 43087,2 | 21365,51 |  | Total | 198380,42 |
| Accuracy Derived | Âj [Ha] | 2318,36 | 13405,85 | 116928,61 | 43350,93 | 22376,67 |  | Total | 198380,42 |
|  | (SE)[Ha] | 0,00 | 151,45 | 279,37 | 359,92 | 366,78 |  |  |  |
| Mean Pixel count | Area [Ha] | 2316,54 | 12090,26 | _119374,80_ | 43041,76 | 21557,06 |  | Total | 198380,42 |
|  | (SE)[Ha]* | 5,02 | 48,50 | 85,01 | 47,77 | 65,73 |  |  |  |
| **ACB_A 1988T** |  |  |  |  |  |  |  |  |  |
| Mode Pixel count | Area [Ha] | 289,81 | 2354,43 | 16262,74 | 2055,70 | 1182,26 |  | Total | 22144,94 |
|  | Area% | 1,31 | 10,63 | 73,44 | 9,28 | 5,34 |  |  |  |
| Mean Pixel count | Area [Ha] | _288,92_ | _2409,41_ | 16188,86 | 2052,40 | 1205,36 |  | Total | 22144,94 |
|  | (SE)[Ha]* | 0,99 | 18,64 | 20,08 | 5,74 | 7,26 |  |  |  |
| **GUE 1988T PA** |  |  |  |  |  |  |  |  |  |
| Mode Pixel count | Area [Ha] | 31,34 | _90,99_ | 16323,80 | 15088,63 | 11691,88 |  | Total | 43226,64 |
|  | Area% | 0,07 | 0,21 | 37,76 | 34,91 | 27,05 |  |  |  |
| Mean Pixel count | Area [Ha] | 32,72 | 96,60 | 16288,94 | 15071,92 | 11736,46 |  | Total | 43226,63 |
|  | (SE)[Ha]* | 2,14 | 1,32 | 17,07 | 21,79 | 22,42 |  |  |  |
| **ACB_A 1988T PA** |  |  |  |  |  |  |  |  |  |
| Mode Pixel count | Area [Ha] | 0,00 | 2,08 | 1351,07 | 438,75 | 61,32 |  | Total | 1853,22 |
|  | Area% | 0,00 | 0,11 | 72,90 | 23,68 | 3,31 |  |  |  |
| Mean Pixel count | Area [Ha] | 0,00 | 2,42 | 1347,15 | 439,13 | 64,52 |  | Total | 1853,22 |
|  | (SE)[Ha]* | 0,00 | 0,09 | 1,11 | 0,94 | 0,67 |  |  |  |

*Pixel count SE was calculated as $SE = \sigma/\sqrt{n}$, where $\sigma$ is the standard deviation and $n$, the number of iterations.

Underlined results in italics indicate that the 20 iteration results were not normally distributed.

## GUE 2019O

Sample counts error matrix and estimated area proportions error matrix

|  | | Test dataset | | | | | |
|---|---|---|---|---|---|---|---|
| **Map** | | **0 / W** | **1 / B** | **2 / A** | **3 / F** | **4 / S** | **Total** |
| Classified | **0 / W** | 169 | 0 | 0 | 0 | 0 | 169 |
| | **1 / B** | 0 | 310 | 26 | 0 | 3 | 339 |
| | **2 / A** | 0 | 134 | 7160 | 91 | 14 | 7399 |
| | **3 / F** | 0 | 2 | 27 | 2481 | 93 | 2603 |
| | **4 / S** | 0 | 0 | 19 | 165 | 776 | 960 |
| **Total** | | 169 | 446 | 7232 | 2737 | 886 | 11470 |

|  | | Test dataset | | | | | |
|---|---|---|---|---|---|---|---|
| **Map** | | **0 / W** | **1 / B** | **2 / A** | **3 / F** | **4 / S** | **Total%** |
| Classified | **0 / W** | **1,4** | 0,0 | 0,0 | 0,0 | 0,0 | 1,4 |
| | **1 / B** | 0,0 | **3,8** | 0,3 | 0,0 | 0,0 | 4,2 |
| | **2 / A** | 0,0 | 1,1 | **57,0** | 0,7 | 0,1 | 58,9 |
| | **3 / F** | 0,0 | 0,0 | 0,3 | **24,1** | 0,9 | 25,3 |
| | **4 / S** | 0,0 | 0,0 | 0,2 | 1,8 | **8,3** | 10,2 |
| **Total%** | | 1,4 | 4,9 | 57,8 | 26,6 | 9,3 | 100,0 |

Result summary:

| 20 Seed Mode | | 0 / W | 1 / B | 2 / A | 3 / F | 4 / S | | | |
|---|---|---|---|---|---|---|---|---|---|
| **GUE 2019O** | Total% | 1,41 | 4,89 | 57,81 | 26,58 | 9,30 | | **Ô** | **94,60%** |
| | (SE%) | 0,00 | 0,11 | 0,15 | 0,18 | 0,16 | | **(SE)** | **0,216%** |
| | Prod% | 100,00 | 77,77 | 98,65 | 90,67 | 88,70 | | | |
| | (SE%) | 0,00 | 3,23 | 0,16 | 0,62 | 1,84 | | | |
| | User% | 100,00 | 91,45 | 96,77 | 95,31 | 80,83 | | | |
| | (SE%) | 0,00 | 1,52 | 0,21 | 0,41 | 1,27 | | | |
| | F-score | 100,00 | 78,98 | 97,87 | 92,92 | 84,07 | | | |
| Mode Pixel count | Area [Ha] | 2608,46 | 7666,23 | 108663,97 | 46620,54 | 18821,68 | | Total | 184380,88 |
| Accuracy Derived | Âj [Ha] | 2608,46 | 9014,20 | 106598,00 | 49006,91 | 17153,30 | | Total | 184380,88 |
| | (SE)[Ha] | 0,00 | 206,35 | 279,09 | 330,47 | 300,79 | | | |
| Mean Pixel count | Area [Ha] | 2603,89 | 7664,32 | 108697,13 | 46315,81 | _19099,74_ | | Total | 184380,88 |
| | (SE)[Ha]* | 6,13 | 38,17 | 65,24 | 77,19 | 64,65 | | | |
| **ACB_A 2019O** | | | | | | | | | |
| Mode Pixel count | Area [Ha] | 204,01 | 703,48 | 16626,29 | 2870,38 | 1465,93 | | Total | 21870,09 |
| | Area% | 0,93 | 3,22 | 76,02 | 13,12 | 6,70 | | | |
| Mean Pixel count | Area [Ha] | 203,61 | 739,37 | 16587,57 | 2849,17 | 1490,38 | | Total | 21870,10 |
| | (SE)[Ha]* | 0,51 | 8,15 | 7,74 | 6,33 | 6,58 | | | |
| **GUE 2019O PA** | | | | | | | | | |
| Mode Pixel count | Area [Ha] | 39,59 | 282,36 | 16312,26 | 15410,66 | 10202,45 | | Total | 42247,32 |
| | Area% | 0,09 | 0,67 | 38,61 | 36,48 | 24,15 | | | |
| Mean Pixel count | Area [Ha] | _40,09_ | 293,34 | 16279,94 | 15381,56 | 10252,40 | | Total | 42247,32 |
| | (SE)[Ha]* | 2,52 | 3,12 | 15,63 | 28,26 | 23,22 | | | |
| **ACB_A 2019O PA** | | | | | | | | | |
| Mode Pixel count | Area [Ha] | 0,00 | 18,16 | 1344,59 | 379,89 | 108,18 | | Total | 1850,82 |
| | Area% | 0,00 | 0,98 | 72,65 | 20,53 | 5,84 | | | |
| Mean Pixel count | Area [Ha] | _0,01_ | 18,77 | 1342,64 | 378,71 | 110,70 | | Total | 1850,83 |
| | (SE)[Ha]* | 0,01 | 0,36 | 0,67 | 0,75 | 0,75 | | | |

*Pixel count SE was calculated as $SE = \sigma/\sqrt{n}$, where $\sigma$ is the standard deviation and $n$, the number of iterations.

Underlined results in italics indicate that the 20 iteration results were not normally distributed.

## GUE 2019M

Sample counts error matrix and estimated area proportions error matrix

| | | Test dataset | | | | | |
|---|---|---|---|---|---|---|---|
| | **Map** | **0 / W** | **1 / B** | **2 / A** | **3 / F** | **4 / S** | **Total** |
| Classified | **0 / W** | 259 | 0 | 0 | 0 | 0 | 259 |
| | **1 / B** | 0 | 301 | 44 | 0 | 2 | 347 |
| | **2 / A** | 0 | 131 | 7126 | 64 | 29 | 7350 |
| | **3 / F** | 0 | 4 | 181 | 2324 | 129 | 2638 |
| | **4 / S** | 0 | 3 | 20 | 212 | 755 | 990 |
| | **Total** | 259 | 439 | 7371 | 2600 | 915 | 11584 |

| | | Test dataset | | | | | |
|---|---|---|---|---|---|---|---|
| | **Map** | **0 / W** | **1 / B** | **2 / A** | **3 / F** | **4 / S** | **Total%** |
| Classified | **0 / W** | **1,5** | 0,0 | 0,0 | 0,0 | 0,0 | 1,5 |
| | **1 / B** | 0,0 | **3,8** | 0,6 | 0,0 | 0,0 | 4,4 |
| | **2 / A** | 0,0 | 1,1 | **61,5** | 0,6 | 0,3 | 63,4 |
| | **3 / F** | 0,0 | 0,0 | 1,5 | **19,1** | 1,1 | 21,6 |
| | **4 / S** | 0,0 | 0,0 | 0,2 | 1,9 | **6,9** | 9,1 |
| | **Total%** | 1,5 | 5,0 | 63,7 | 21,6 | 8,3 | 100,0 |

Result summary:

| 20 Seed Mode | | 0 / W | 1 / B | 2 / A | 3 / F | 4 / S | | | |
|---|---|---|---|---|---|---|---|---|---|
| **GUE 2019M** | Total% | 1,52 | 4,97 | 63,69 | 21,57 | 8,25 | | **Ô** | **92,76%** |
| | (SE%) | 0,00 | 0,13 | 0,19 | 0,19 | 0,16 | | **(SE)** | **0,237%** |
| | Prod% | 100,00 | 76,06 | 96,51 | 88,43 | 83,84 | | | |
| | (SE%) | 0,00 | 4,12 | 0,22 | 0,92 | 2,62 | | | |
| | User% | 100,00 | 86,74 | 96,95 | 88,10 | 76,26 | | | |
| | (SE%) | 0,00 | 1,82 | 0,20 | 0,63 | 1,35 | | | |
| | F-score | 100,00 | 76,59 | 96,81 | 88,74 | 79,27 | | | |
| Mode Pixel count | Area [Ha] | 2998,86 | 8606,32 | 125137,59 | 42731,12 | 17907,16 | | Total | 197381,05 |
| Accuracy Derived | Âⱼ [Ha] | 2998,86 | 9814,83 | 125708,82 | 42569,15 | 16289,40 | | Total | 197381,05 |
| | (SE)[Ha] | 0,00 | 252,72 | 370,45 | 381,42 | 316,90 | | | |
| Mean Pixel count | Area [Ha] | 2992,02 | 8682,06 | 124625,04 | 43007,21 | 18074,74 | | Total | 197381,06 |
| | (SE)[Ha]* | 4,75 | 36,08 | 103,25 | 102,41 | 58,75 | | | |
| **ACB_A 2019M** | | | | | | | | | |
| Mode Pixel count | Area [Ha] | 204,24 | 834,29 | 17771,75 | 2290,52 | 977,04 | | Total | 22077,84 |
| | Area% | 0,93 | 3,78 | 80,50 | 10,37 | 4,43 | | | |
| Mean Pixel count | Area [Ha] | 203,82 | 868,10 | 17703,31 | 2310,48 | 992,12 | | Total | 22077,84 |
| | (SE)[Ha]* | 0,45 | 6,30 | 10,29 | 9,28 | 5,73 | | | |
| **GUE 2019M PA** | | | | | | | | | |
| Mode Pixel count | Area [Ha] | 169,26 | 440,32 | 18394,94 | 14830,55 | 9392,94 | | Total | 43228,01 |
| | Area% | 0,39 | 1,02 | 42,55 | 34,31 | 21,73 | | | |
| Mean Pixel count | Area [Ha] | 167,57 | 474,75 | 18286,53 | 14868,18 | 9430,99 | | Total | 43228,02 |
| | (SE)[Ha]* | 2,45 | 4,13 | 19,36 | 21,15 | 14,12 | | | |
| **ACB_A 2019M PA** | | | | | | | | | |
| Mode Pixel count | Area [Ha] | 0,50 | 25,07 | 1403,17 | 358,30 | 66,29 | | Total | 1853,33 |
| | Area% | 0,03 | 1,35 | 75,71 | 19,33 | 3,58 | | | |
| Mean Pixel count | Area [Ha] | 0,50 | 26,90 | 1395,94 | 361,85 | 68,14 | | Total | 1853,33 |
| | (SE)[Ha]* | 0,03 | 0,28 | 0,93 | 0,91 | 0,50 | | | |

*Pixel count SE was calculated as $SE = \sigma/\sqrt{n}$, where $\sigma$ is the standard deviation and $n$, the number of iterations.

Underlined results in italics indicate that the 20 iteration results were not normally distributed.

**GUE 2024M**

Sample counts error matrix and estimated area proportions error matrix

| | Map | Test dataset | | | | | Total |
|---|---|---|---|---|---|---|---|
| | | 0 / W | 1 / B | 2 / A | 3 / F | 4 / S | |
| Classified | 0 / W | 259 | 0 | 0 | 0 | 0 | 259 |
| | 1 / B | 0 | 291 | 51 | 1 | 4 | 347 |
| | 2 / A | 0 | 325 | 6935 | 42 | 46 | 7348 |
| | 3 / F | 0 | 3 | 115 | 2415 | 105 | 2638 |
| | 4 / S | 0 | 1 | 5 | 123 | 861 | 990 |
| | Total | 259 | 620 | 7106 | 2581 | 1016 | 11582 |

| | Map | Test dataset | | | | | Total% |
|---|---|---|---|---|---|---|---|
| | | 0 / W | 1 / B | 2 / A | 3 / F | 4 / S | |
| Classified | 0 / W | 1,7 | 0,0 | 0,0 | 0,0 | 0,0 | 1,7 |
| | 1 / B | 0,0 | 4,3 | 0,8 | 0,0 | 0,1 | 5,1 |
| | 2 / A | 0,0 | 2,8 | 59,7 | 0,4 | 0,4 | 63,3 |
| | 3 / F | 0,0 | 0,0 | 0,9 | 18,8 | 0,8 | 20,5 |
| | 4 / S | 0,0 | 0,0 | 0,0 | 1,2 | 8,2 | 9,4 |
| | Total% | 1,7 | 7,1 | 61,4 | 20,3 | 9,4 | 100,0 |

Result summary:

| 20 Seed Mode | | 0 / W | 1 / B | 2 / A | 3 / F | 4 / S | | | |
|---|---|---|---|---|---|---|---|---|---|
| **GUE 2024M** | Total% | 1,71 | 7,12 | 61,42 | 20,32 | 9,45 | | Ô | 92,66% |
| | (SE%) | 0,00 | 0,18 | 0,21 | 0,16 | 0,14 | | (SE) | 0,248% |
| | Prod% | 100,00 | 60,21 | 97,24 | 92,40 | 86,54 | | | |
| | (SE%) | 0,00 | 5,80 | 0,22 | 0,67 | 1,71 | | | |
| | User% | 100,00 | 83,86 | 94,38 | 91,55 | 86,97 | | | |
| | (SE%) | 0,00 | 1,98 | 0,27 | 0,54 | 1,07 | | | |
| | F-score | 100,00 | 60,19 | 95,96 | 92,55 | 85,84 | | | |
| Mode Pixel count | Area [Ha] | 3379 | 10119,1 | 125349,57 | 40621,02 | 18618,98 | | Total | 198087,67 |
| Accuracy Derived | Âj [Ha] | 3379,00 | 14095,23 | 121656,30 | 40246,08 | 18711,07 | | Total | 198087,67 |
| | (SE)[Ha] | 0,00 | 362,49 | 422,22 | 315,45 | 283,29 | | | |
| Mean Pixel count | Area [Ha] | 3357,80 | 10233,49 | 124925,50 | 40790,30 | 18780,58 | | Total | 198087,67 |
| | (SE)[Ha]* | 10,62 | 42,97 | 87,27 | 94,19 | 52,21 | | | |
| **ACB_A 2024M** | | | | | | | | | |
| Mode Pixel count | Area [Ha] | 194,96 | 1750,14 | 16995,51 | 2043,64 | 1091,94 | | Total | 22076,19 |
| | Area% | 0,88 | 7,93 | 76,99 | 9,26 | 4,95 | | | |
| Mean Pixel count | Area [Ha] | 194,61 | 1782,27 | 16942,25 | 2059,76 | 1097,28 | | Total | 22076,18 |
| | (SE)[Ha]* | 0,39 | 12,27 | 12,40 | 6,25 | 4,79 | | | |
| **GUE 2024M PA** | | | | | | | | | |
| Mode Pixel count | Area [Ha] | 133,96 | 774,86 | 18334,84 | 14076,14 | 9908,21 | | Total | 43228,01 |
| | Area% | 0,31 | 1,79 | 42,41 | 32,56 | 22,92 | | | |
| Mean Pixel count | Area [Ha] | 128,28 | 823,54 | 18237,77 | 14101,80 | 9936,63 | | Total | 43228,02 |
| | (SE)[Ha]* | 4,04 | 12,42 | 20,19 | 23,53 | 16,39 | | | |
| **ACB_A 2024M PA** | | | | | | | | | |
| Mode Pixel count | Area [Ha] | 0,22 | 62,00 | 1403,24 | 308,97 | 78,90 | | Total | 1853,33 |
| | Area% | 0,01 | 3,35 | 75,71 | 16,67 | 4,26 | | | |
| Mean Pixel count | Area [Ha] | 0,23 | 66,57 | 1396,46 | 310,15 | 79,93 | | Total | 1853,34 |
| | (SE)[Ha]* | 0,02 | 1,11 | 1,10 | 0,64 | 0,41 | | | |

*Pixel count SE was calculated as $SE = \sigma/\sqrt{n}$, where $\sigma$ is the standard deviation and $n$, the number of iterations.

Underlined results in italics indicate that the 20 iteration results were not normally distributed.

**RRB 1988**

Sample counts error matrix and estimated area proportions error matrix

| | Map | 0 / W | 1 / B | 2 / A | 3 / F | 4 / S | Total |
|---|---|---|---|---|---|---|---|
| | | | | Test dataset | | | |
| Classified | 0 / W | 115 | 0 | 0 | 0 | 0 | 115 |
| | 1 / B | 0 | 6 | 9 | 0 | 0 | 15 |
| | 2 / A | 0 | 0 | 140 | 0 | 2 | 142 |
| | 3 / F | 0 | 0 | 5 | 126 | 13 | 144 |
| | 4 / S | 0 | 0 | 11 | 21 | 212 | 244 |
| | Total | 115 | 6 | 165 | 147 | 227 | 660 |

| | Map | 0 / W | 1 / B | 2 / A | 3 / F | 4 / S | Total% |
|---|---|---|---|---|---|---|---|
| | | | | Test dataset | | | |
| Classified | 0 / W | 1,1 | 0,0 | 0,0 | 0,0 | 0,0 | 1,1 |
| | 1 / B | 0,0 | 0,3 | 0,5 | 0,0 | 0,0 | 0,9 |
| | 2 / A | 0,0 | 0,0 | 57,3 | 0,0 | 0,8 | 58,1 |
| | 3 / F | 0,0 | 0,0 | 0,9 | 21,8 | 2,2 | 24,9 |
| | 4 / S | 0,0 | 0,0 | 0,7 | 1,3 | 13,0 | 15,0 |
| | Total% | 1,1 | 0,3 | 59,4 | 23,1 | 16,1 | 100,0 |

Result summary:

| 20 Seed Mode | | 0 / W | 1 / B | 2 / A | 3 / F | 4 / S | | | |
|---|---|---|---|---|---|---|---|---|---|
| **RRB 1988** | Total% | 1,11 | 0,34 | 59,35 | 23,09 | 16,10 | | Ô | 93,58% |
| | (SE%) | 0,00 | 0,11 | 0,73 | 0,74 | 0,89 | | (SE) | 0,962% |
| | Prod% | 100,00 | 100,00 | 96,54 | 94,41 | 80,95 | | | |
| | (SE%) | 0,00 | 0,00 | 0,80 | 2,00 | 5,67 | | | |
| | User% | 100,00 | 40,00 | 98,59 | 87,50 | 86,89 | | | |
| | (SE%) | 0,00 | 13,09 | 0,99 | 2,77 | 2,17 | | | |
| | F-score | 100,00 | 57,14 | 91,21 | 86,60 | 90,02 | | | |
| Mode Pixel count | Area [Ha] | 1018,99 | 786,7 | 53260,1 | 22828,9 | 13751,63 | | Total | 91646,32 |
| Accuracy Derived | Âⱼ [Ha] | 1018,99 | 314,68 | 54394,60 | 21158,83 | 14759,22 | | Total | 91646,32 |
| | (SE)[Ha] | 0,00 | 99,51 | 664,80 | 675,88 | 814,21 | | | |
| Mean Pixel count | Area [Ha] | 1019,42 | 818,33 | 53054,42 | 22797,72 | 13956,43 | | Total | 91646,32 |
| | (SE)[Ha]* | 13,17 | 16,18 | 53,97 | 74,78 | 62,59 | | | |
| **ACB_B 1988** | | | | | | | | | |
| Mode Pixel count | Area [Ha] | 5,04 | 314,65 | 5284,64 | 455,06 | 259,57 | | Total | 6318,96 |
| | Area% | 0,08 | 4,98 | 83,63 | 7,20 | 4,11 | | | |
| Mean Pixel count | Area [Ha] | 4,64 | 316,67 | 5275,68 | 456,14 | 265,83 | | Total | 6318,96 |
| | (SE)[Ha]* | 0,19 | 4,25 | 4,98 | 2,26 | 1,79 | | | |
| **RRB 1988 PA** | | | | | | | | | |
| Mode Pixel count | Area [Ha] | 838,28 | 111,72 | 2783,46 | 11073,03 | 6786,42 | | Total | 21592,91 |
| | Area% | 3,88 | 0,52 | 12,89 | 51,28 | 31,43 | | | |
| Mean Pixel count | Area [Ha] | 833,78 | 117,82 | 2765,53 | 11024,21 | 6851,57 | | Total | 21592,90 |
| | (SE)[Ha]* | 9,13 | 1,60 | 11,18 | 26,13 | 27,48 | | | |
| **ACB_B 1988 PA** | | | | | | | | | |
| Mode Pixel count | Area [Ha] | 2,48 | 4,31 | 1122,75 | 246,02 | 86,48 | | Total | 1462,04 |
| | Area% | 0,17 | 0,29 | 76,79 | 16,83 | 5,92 | | | |
| Mean Pixel count | Area [Ha] | 2,14 | 4,77 | 1118,35 | 246,54 | 90,23 | | Total | 1462,04 |
| | (SE)[Ha]* | 0,06 | 0,14 | 0,86 | 0,68 | 0,80 | | | |

*Pixel count SE was calculated as $SE = \sigma/\sqrt{n}$, where $\sigma$ is the standard deviation and $n$, the number of iterations.

Underlined results in italics indicate that the 20 iteration results were not normally distributed.

## RRB 1988T

Sample counts error matrix and estimated area proportions error matrix

|  |  | Test dataset | | | | | |
|---|---|---|---|---|---|---|---|
| | **Map** | **0 / W** | **1 / B** | **2 / A** | **3 / F** | **4 / S** | **Total** |
| **Classified** | **0 / W** | 451 | 0 | 0 | 0 | 0 | 451 |
| | **1 / B** | 0 | 57 | 0 | 0 | 0 | 57 |
| | **2 / A** | 0 | 0 | 734 | 2 | 1 | 737 |
| | **3 / F** | 0 | 6 | 19 | 534 | 23 | 582 |
| | **4 / S** | 0 | 0 | 17 | 57 | 935 | 1009 |
| | **Total** | 451 | 63 | 770 | 593 | 959 | 2836 |

|  |  | Test dataset | | | | | |
|---|---|---|---|---|---|---|---|
| | **Map** | **0 / W** | **1 / B** | **2 / A** | **3 / F** | **4 / S** | **Total%** |
| **Classified** | **0 / W** | **0,5** | 0,0 | 0,0 | 0,0 | 0,0 | 0,5 |
| | **1 / B** | 0,0 | **2,2** | 0,0 | 0,0 | 0,0 | 2,2 |
| | **2 / A** | 0,0 | 0,0 | **57,2** | 0,2 | 0,1 | 57,4 |
| | **3 / F** | 0,0 | 0,3 | 1,0 | **29,5** | 1,3 | 32,1 |
| | **4 / S** | 0,0 | 0,0 | 0,1 | 0,4 | **7,2** | 7,7 |
| | **Total%** | 0,5 | 2,5 | 58,4 | 30,1 | 8,5 | 100,0 |

Result summary:

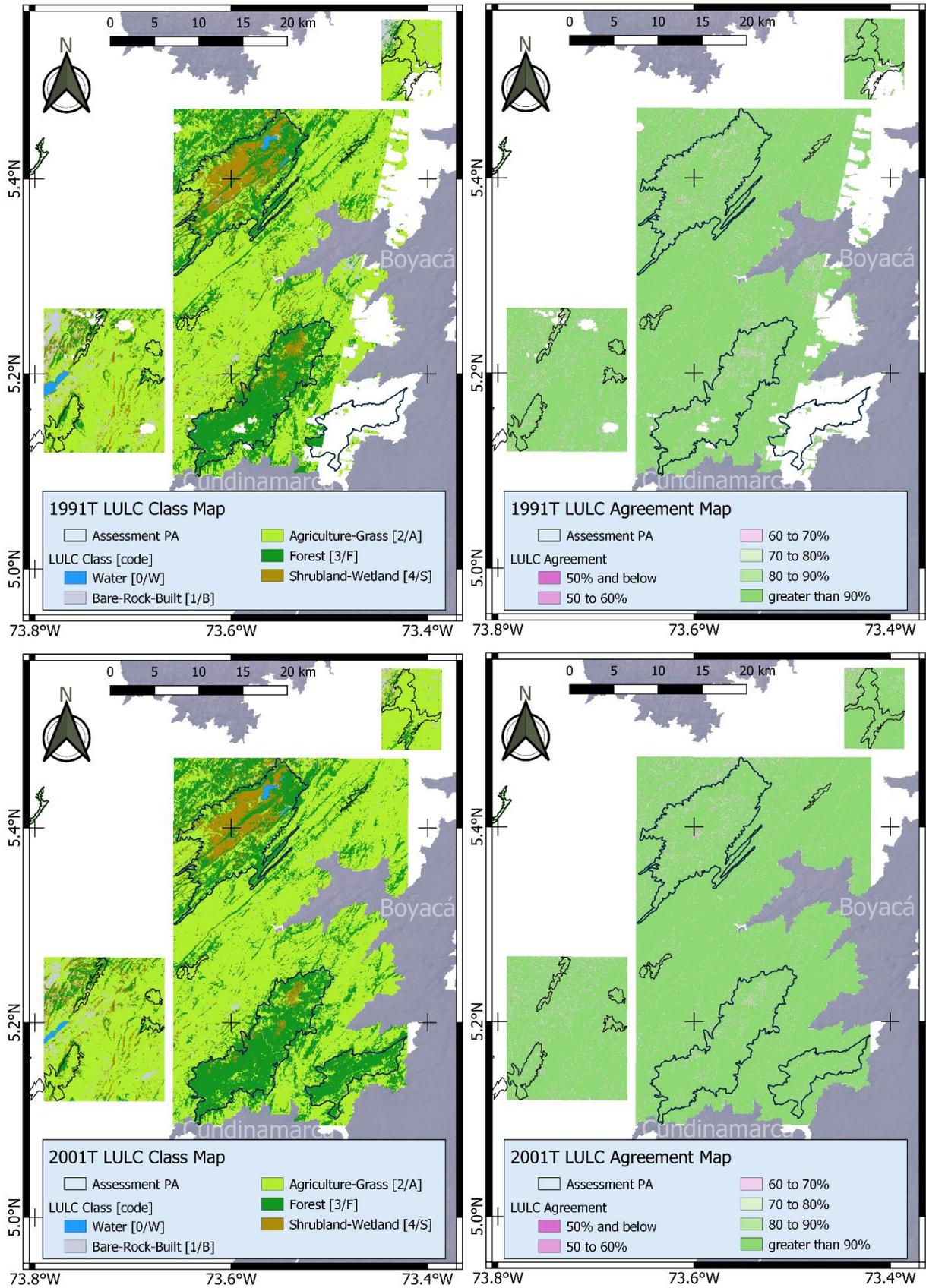| 20 Seed Mode | | 0 / W | 1 / B | 2 / A | 3 / F | 4 / S | | | |
|---|---|---|---|---|---|---|---|---|---|
| **RRB 1988T** | Total% | 0,53 | 2,52 | 58,37 | 30,07 | 8,51 | | **Ô** | _96,55%_ |
| | (SE%) | 0,00 | 0,13 | 0,27 | 0,39 | 0,28 | | **(SE)** | **0,396%** |
| | Prod% | 100,00 | 86,86 | 97,98 | 98,03 | 84,16 | | | |
| | (SE%) | 0,00 | 4,64 | 0,40 | 0,47 | 2,93 | | | |
| | User% | 100,00 | 100,00 | 99,59 | 91,75 | 92,67 | | | |
| | (SE%) | 0,00 | 0,00 | 0,23 | 1,14 | 0,82 | | | |
| | F-score | 100,00 | 95,00 | 97,41 | 90,89 | 95,02 | | | |
| Mode Pixel count | Area [Ha] | 510,22 | 2115,29 | 55461,26 | 31033,24 | 7464,15 | | Total | 96584,16 |
| Accuracy Derived | Âj [Ha] | 510,22 | 2435,22 | 56374,37 | 29045,96 | 8218,38 | | Total | 96584,16 |
| | (SE)[Ha] | 0,00 | 129,94 | 264,75 | 373,44 | 268,73 | | | |
| Mean Pixel count | Area [Ha] | 518,39 | 2203,25 | 55087,03 | 31037,16 | 7738,33 | | Total | 96584,16 |
| | (SE)[Ha]* | 17,34 | 39,10 | 89,59 | 79,94 | 58,56 | | | |
| **ACB_B 1988T** | | | | | | | | | |
| Mode Pixel count | Area [Ha] | 1,89 | 508,25 | 5013,63 | 562,80 | 137,71 | | Total | 6224,28 |
| | Area% | 0,03 | 8,17 | 80,55 | 9,04 | 2,21 | | | |
| Mean Pixel count | Area [Ha] | 2,06 | 510,32 | 4997,97 | 570,66 | 143,27 | | Total | 6224,28 |
| | (SE)[Ha]* | 0,08 | 6,50 | 7,17 | 2,88 | 2,18 | | | |
| **RRB 1988T PA** | | | | | | | | | |
| Mode Pixel count | Area [Ha] | 471,68 | 122,09 | 3775,53 | 15273,21 | 5004,73 | | Total | 24647,24 |
| | Area% | 1,91 | 0,50 | 15,32 | 61,97 | 20,31 | | | |
| Mean Pixel count | Area [Ha] | 474,03 | 134,16 | 3714,59 | 15230,66 | 5093,80 | | Total | 24647,24 |
| | (SE)[Ha]* | 17,70 | 4,84 | 25,02 | 30,09 | 25,12 | | | |
| **ACB_B 1988T PA** | | | | | | | | | |
| Mode Pixel count | Area [Ha] | 1,09 | 12,51 | 1132,57 | 290,31 | 25,73 | | Total | 1462,21 |
| | Area% | 0,07 | 0,86 | 77,46 | 19,85 | 1,76 | | | |
| Mean Pixel count | Area [Ha] | _1,20_ | 13,68 | 1127,13 | 293,03 | 27,17 | | Total | 1462,21 |
| | (SE)[Ha]* | 0,04 | 0,41 | 1,11 | 1,02 | 0,44 | | | |

*Pixel count SE was calculated as $SE = \sigma/\sqrt{n}$, where $\sigma$ is the standard deviation and $n$, the number of iterations.
Underlined results in italics indicate that the 20 iteration results were not normally distributed.
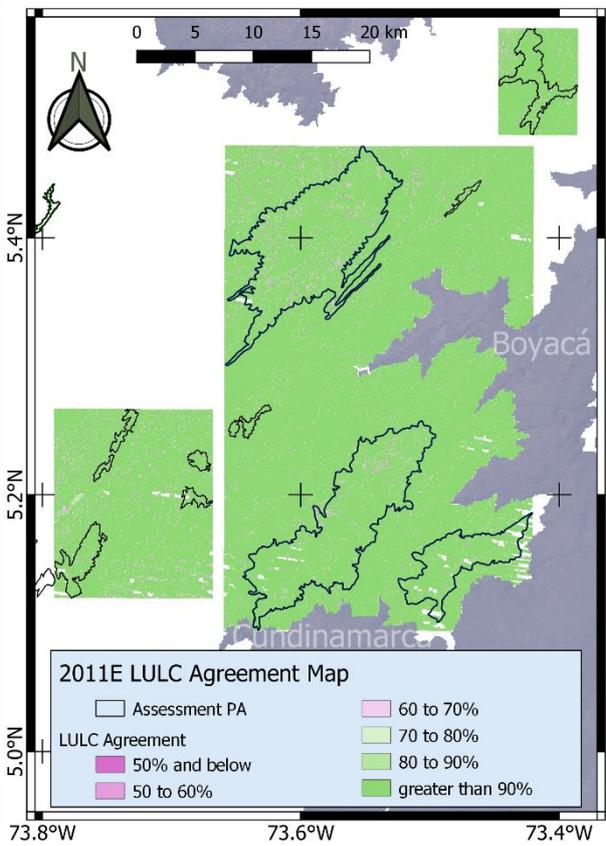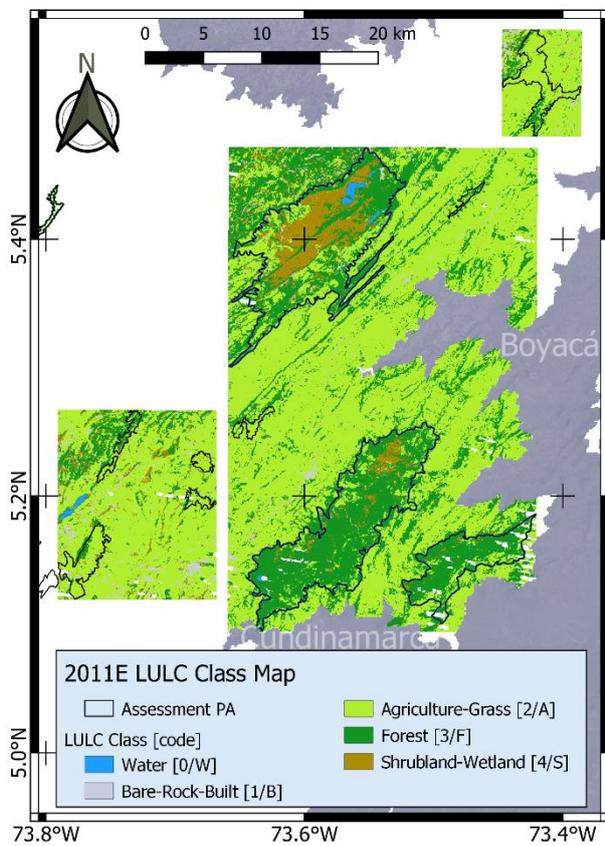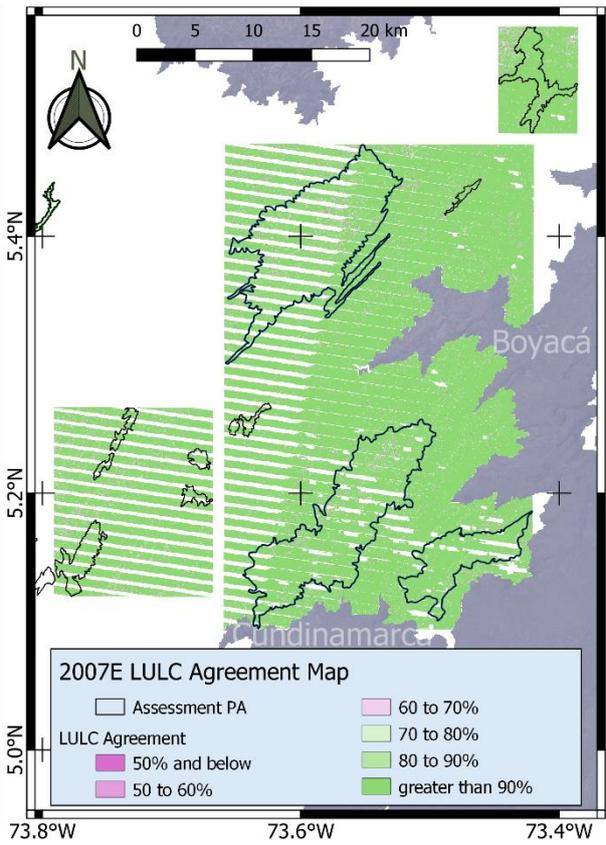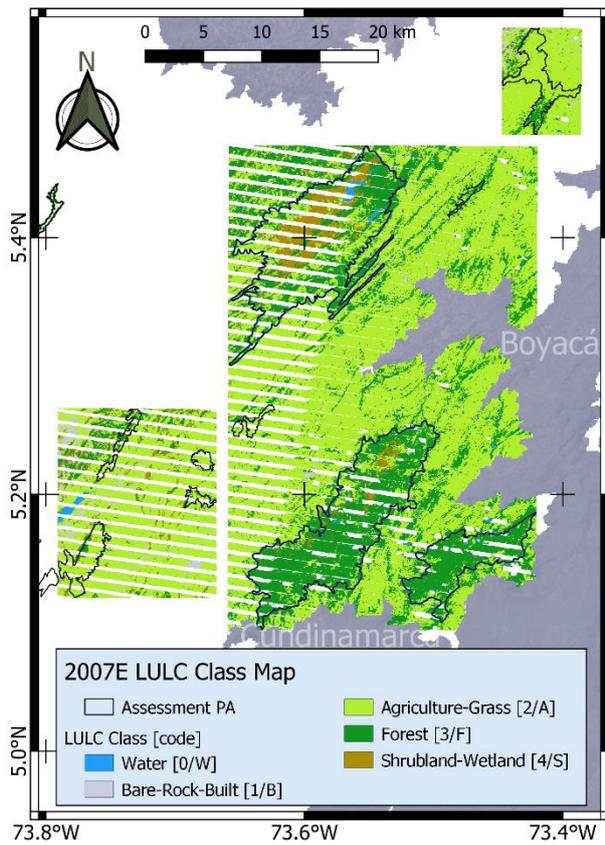
## RRB 2019O

### Sample counts error matrix and estimated area proportions error matrix

| | | Test dataset | | | | | |
|---|---|---|---|---|---|---|---|
| | **Map** | **0 / W** | **1 / B** | **2 / A** | **3 / F** | **4 / S** | **Total** |
| Classified | **0 / W** | 41 | 0 | 410 | 0 | 0 | 451 |
| | **1 / B** | 0 | 55 | 2 | 0 | 0 | 57 |
| | **2 / A** | 0 | 0 | 659 | 1 | 0 | 660 |
| | **3 / F** | 2 | 3 | 3 | 571 | 3 | 582 |
| | **4 / S** | 0 | 1 | 0 | 62 | 946 | 1009 |
| | **Total** | 43 | 59 | 1074 | 634 | 949 | 2759 |

| | | Test dataset | | | | | |
|---|---|---|---|---|---|---|---|
| | **Map** | **0 / W** | **1 / B** | **2 / A** | **3 / F** | **4 / S** | **Total%** |
| Classified | **0 / W** | **0,0** | 0,0 | 0,3 | 0,0 | 0,0 | 0,3 |
| | **1 / B** | 0,0 | **0,9** | 0,0 | 0,0 | 0,0 | 0,9 |
| | **2 / A** | 0,0 | 0,0 | **59,8** | 0,1 | 0,0 | 59,9 |
| | **3 / F** | 0,1 | 0,2 | 0,2 | **32,0** | 0,2 | 32,6 |
| | **4 / S** | 0,0 | 0,0 | 0,0 | 0,4 | **5,9** | 6,3 |
| | **Total%** | 0,1 | 1,0 | 60,3 | 32,5 | 6,1 | 100,0 |

### Result summary:

| 20 Seed Mode | | 0 / W | 1 / B | 2 / A | 3 / F | 4 / S | | | |
|---|---|---|---|---|---|---|---|---|---|
| **RRB 2019O** | Total% | 0,14 | 1,05 | 60,26 | 32,47 | 6,08 | | **Ô** | **98,61%** |
| | (SE%) | 0,08 | 0,10 | 0,13 | 0,21 | 0,11 | | **(SE)** | **0,212%** |
| | Prod% | 18,64 | 83,32 | 99,24 | 98,53 | 97,24 | | | |
| | (SE%) | 226,31 | 8,51 | 0,16 | 0,32 | 1,56 | | | |
| | User% | 9,09 | 96,49 | 99,85 | 98,11 | 93,76 | | | |
| | (SE%) | 1,36 | 2,46 | 0,15 | 0,56 | 0,76 | | | |
| | F-score | 16,60 | 94,83 | 76,01 | 93,91 | 96,63 | | | |
| Mode Pixel count | Area [Ha] | 261,76 | 836,79 | 55527,67 | 30229,82 | 5847,98 | | Total | 92704,02 |
| Accuracy Derived | $\hat{A}_j$ [Ha] | 127,68 | 969,05 | 55866,69 | 30101,94 | 5638,67 | | Total | 92704,02 |
| | (SE)[Ha] | 73,42 | 92,20 | 124,69 | 195,29 | 100,18 | | | |
| Mean Pixel count | Area [Ha] | *256,56* | 906,97 | 55432,11 | *30145,36* | 5963,03 | | Total | 92704,03 |
| | (SE)[Ha]* | 6,09 | 11,75 | 48,74 | 44,26 | 20,61 | | | |
| **ACB_B 2019O** | | | | | | | | | |
| Mode Pixel count | Area [Ha] | 3,69 | 175,78 | 5130,19 | 619,86 | 280,99 | | Total | 6210,51 |
| | Area% | 0,06 | 2,83 | 82,60 | 9,98 | 4,52 | | | |
| Mean Pixel count | Area [Ha] | *3,67* | 180,24 | 5119,39 | 620,85 | 286,35 | | Total | 6210,50 |
| | (SE)[Ha]* | 0,10 | 3,48 | 4,47 | 1,88 | 3,06 | | | |
| **RRB 2019O PA** | | | | | | | | | |
| Mode Pixel count | Area [Ha] | 173,52 | 125,48 | 4670,72 | 15142,55 | 4237,10 | | Total | 24349,37 |
| | Area% | 0,71 | 0,52 | 19,18 | 62,19 | 17,40 | | | |
| Mean Pixel count | Area [Ha] | 170,53 | 173,97 | 4656,75 | 15058,44 | 4289,68 | | Total | 24349,37 |
| | (SE)[Ha]* | 3,68 | 8,24 | 16,47 | 18,92 | 10,25 | | | |
| **ACB_B 2019O PA** | | | | | | | | | |
| Mode Pixel count | Area [Ha] | 2,06 | 14,72 | 1226,03 | 170,21 | 48,83 | | Total | 1461,85 |
| | Area% | 0,14 | 1,01 | 83,87 | 11,64 | 3,34 | | | |
| Mean Pixel count | Area [Ha] | *2,08* | 16,79 | 1221,37 | 171,69 | 49,92 | | Total | 1461,86 |
| | (SE)[Ha]* | 0,06 | 0,54 | 1,14 | 0,65 | 0,70 | | | |

*Pixel count SE was calculated as $SE = \sigma/\sqrt{n}$, where $\sigma$ is the standard deviation and $n$, the number of iterations.

Underlined results in italics indicate that the 20 iteration results were not normally distributed.

## RRB 2019M

Sample counts error matrix and estimated area proportions error matrix

| | Map | | | Test dataset | | | |
|---|---|---|---|---|---|---|---|
| | | 0 / W | 1 / B | 2 / A | 3 / F | 4 / S | Total |
| Classified | 0 / W | 216 | 213 | 0 | 21 | 0 | 450 |
| | 1 / B | 0 | 56 | 8 | 0 | 0 | 64 |
| | 2 / A | 0 | 0 | 730 | 15 | 3 | 748 |
| | 3 / F | 0 | 0 | 58 | 550 | 10 | 618 |
| | 4 / S | 0 | 0 | 4 | 60 | 949 | 1013 |
| | Total | 216 | 269 | 800 | 646 | 962 | 2893 |

| | Map | | | Test dataset | | | |
|---|---|---|---|---|---|---|---|
| | | 0 / W | 1 / B | 2 / A | 3 / F | 4 / S | Total% |
| Classified | 0 / W | 0,1 | 0,1 | 0,0 | 0,0 | 0,0 | 0,2 |
| | 1 / B | 0,0 | 0,7 | 0,1 | 0,0 | 0,0 | 0,8 |
| | 2 / A | 0,0 | 0,0 | 58,7 | 1,2 | 0,2 | 60,1 |
| | 3 / F | 0,0 | 0,0 | 3,1 | 29,4 | 0,5 | 33,1 |
| | 4 / S | 0,0 | 0,0 | 0,0 | 0,3 | 5,5 | 5,8 |
| | Total% | 0,1 | 0,8 | 61,9 | 31,0 | 6,2 | 100,0 |

Result summary:

| 20 Seed Mode | | 0 / W | 1 / B | 2 / A | 3 / F | 4 / S | | | |
|---|---|---|---|---|---|---|---|---|---|
| **RRB 2019M** | Total% | 0,10 | 0,79 | 61,90 | 30,99 | 6,23 | | Ô | 94,34% |
| | (SE%) | 0,00 | 0,03 | 0,51 | 0,52 | 0,22 | | (SE) | 0,539% |
| | Prod% | 100,00 | 88,16 | 94,79 | 94,97 | 87,54 | | | |
| | (SE%) | 0,00 | 5,00 | 0,66 | 1,17 | 3,19 | | | |
| | User% | 48,00 | 87,50 | 97,59 | 89,00 | 93,68 | | | |
| | (SE%) | 2,36 | 4,17 | 0,56 | 1,26 | 0,76 | | | |
| | F-score | 64,86 | 33,63 | 94,32 | 87,03 | 96,10 | | | |
| Mode Pixel count | Area [Ha] | 191,53 | 771,15 | 58037,85 | 31921,97 | 5617,65 | | Total | 96540,15 |
| Accuracy Derived | Âj [Ha] | 91,93 | 765,41 | 59755,71 | 29915,05 | 6012,04 | | Total | 96540,15 |
| | (SE)[Ha] | 4,51 | 32,20 | 497,11 | 501,70 | 214,67 | | | |
| Mean Pixel count | Area [Ha] | 191,45 | 774,88 | 57887,06 | 31938,06 | 5748,70 | | Total | 96540,15 |
| | (SE)[Ha]* | 3,41 | 8,98 | 32,18 | 36,17 | 28,79 | | | |
| **ACB_B 2019M** | | | | | | | | | |
| Mode Pixel count | Area [Ha] | 1,59 | 158,12 | 5132,10 | 771,17 | 161,30 | | Total | 6224,28 |
| | Area% | 0,03 | 2,54 | 82,45 | 12,39 | 2,59 | | | |
| Mean Pixel count | Area [Ha] | 1,57 | *160,29* | 5109,57 | 784,71 | 168,14 | | Total | 6224,28 |
| | (SE)[Ha]* | 0,02 | 1,49 | 2,43 | 2,61 | 2,05 | | | |
| **RRB 2019M PA** | | | | | | | | | |
| Mode Pixel count | Area [Ha] | 185,14 | 124,18 | 4084,66 | 15998,60 | 4255,21 | | Total | 24647,79 |
| | Area% | 0,75 | 0,50 | 16,57 | 64,91 | 17,26 | | | |
| Mean Pixel count | Area [Ha] | 184,64 | 116,91 | 4083,03 | 15958,81 | 4304,40 | | Total | 24647,80 |
| | (SE)[Ha]* | 3,39 | 3,68 | 7,66 | 14,89 | 17,00 | | | |
| **ACB_B 2019M PA** | | | | | | | | | |
| Mode Pixel count | Area [Ha] | 0,00 | 12,41 | 1132,43 | 292,58 | 24,89 | | Total | 1462,31 |
| | Area% | 0,00 | 0,85 | 77,44 | 20,01 | 1,70 | | | |
| Mean Pixel count | Area [Ha] | 0,00 | 13,39 | 1126,17 | *296,74* | 26,03 | | Total | 1462,32 |
| | (SE)[Ha]* | 0,00 | 0,19 | 0,61 | 0,61 | 0,41 | | | |

*Pixel count SE was calculated as $SE = \sigma/\sqrt{n}$, where $\sigma$ is the standard deviation and $n$, the number of iterations.

Underlined results in italics indicate that the 20 iteration results were not normally distributed.

## RRB 2024M

Sample counts error matrix and estimated area proportions error matrix

| | | Test dataset | | | | | |
|---|---|---|---|---|---|---|---|
| | **Map** | **0 / W** | **1 / B** | **2 / A** | **3 / F** | **4 / S** | **Total** |
| Classified | **0 / W** | 450 | 0 | 0 | 0 | 0 | 450 |
| | **1 / B** | 0 | 55 | 8 | 0 | 1 | 64 |
| | **2 / A** | 0 | 0 | 735 | 9 | 4 | 748 |
| | **3 / F** | 0 | 0 | 43 | 567 | 8 | 618 |
| | **4 / S** | 0 | 1 | 11 | 40 | 961 | 1013 |
| | **Total** | 450 | 56 | 797 | 616 | 974 | 2893 |

| | | Test dataset | | | | | |
|---|---|---|---|---|---|---|---|
| | **Map** | **0 / W** | **1 / B** | **2 / A** | **3 / F** | **4 / S** | **Total%** |
| Classified | **0 / W** | **0,3** | 0,0 | 0,0 | 0,0 | 0,0 | 0,3 |
| | **1 / B** | 0,0 | **0,9** | 0,1 | 0,0 | 0,0 | 1,0 |
| | **2 / A** | 0,0 | 0,0 | **60,8** | 0,7 | 0,3 | 61,9 |
| | **3 / F** | 0,0 | 0,0 | 2,1 | **27,8** | 0,4 | 30,3 |
| | **4 / S** | 0,0 | 0,0 | 0,1 | 0,3 | **6,2** | 6,5 |
| | **Total%** | 0,3 | 0,9 | 63,1 | 28,8 | 6,9 | 100,0 |

Result summary:

| 20 Seed Mode | | 0 / W | 1 / B | 2 / A | 3 / F | 4 / S | | | |
|---|---|---|---|---|---|---|---|---|---|
| **RRB 2024M** | Total% | 0,27 | 0,87 | 63,11 | 28,83 | 6,91 | | **Ô** | **95,95%** |
| | (SE%) | 0,00 | 0,04 | 0,43 | 0,42 | 0,22 | | **(SE)** | **0,452%** |
| | Prod% | 100,00 | 99,26 | 96,35 | 96,53 | 89,31 | | | |
| | (SE%) | 0,00 | 0,82 | 0,51 | 0,93 | 2,87 | | | |
| | User% | 100,00 | 85,94 | 98,26 | 91,75 | 94,87 | | | |
| | (SE%) | 0,00 | 4,38 | 0,48 | 1,11 | 0,69 | | | |
| | F-score | 100,00 | 91,67 | 95,15 | 91,90 | 96,73 | | | |
| Mode Pixel count | Area [Ha] | 264,67 | 968,18 | 59737,12 | 29283,96 | 6284,53 | | Total | 96538,46 |
| Accuracy Derived | Âj [Ha] | 264,67 | 838,23 | 60925,73 | 27834,24 | 6675,59 | | Total | 96538,46 |
| | (SE)[Ha] | 0,00 | 42,53 | 416,32 | 404,05 | 212,67 | | | |
| Mean Pixel count | Area [Ha] | 263,69 | 980,35 | 59579,99 | 29367,70 | 6346,71 | | Total | 96538,45 |
| | (SE)[Ha]* | 1,59 | 9,68 | 54,33 | 54,41 | 30,36 | | | |
| **ACB_B 2024M** | | | | | | | | | |
| Mode Pixel count | Area [Ha] | 6,85 | 216,42 | 5188,83 | 582,42 | 229,73 | | Total | 6224,25 |
| | Area% | 0,11 | 3,48 | 83,36 | 9,36 | 3,69 | | | |
| Mean Pixel count | Area [Ha] | 6,38 | 216,96 | 5179,93 | 589,64 | 231,34 | | Total | 6224,25 |
| | (SE)[Ha]* | 0,57 | 2,23 | 2,63 | 2,18 | 1,74 | | | |
| **RRB 2024M PA** | | | | | | | | | |
| Mode Pixel count | Area [Ha] | 226,16 | 137,78 | 4246,30 | 15462,81 | 4574,74 | | Total | 24647,79 |
| | Area% | 0,92 | 0,56 | 17,23 | 62,74 | 18,56 | | | |
| Mean Pixel count | Area [Ha] | 225,02 | 145,39 | 4247,10 | 15445,79 | 4584,50 | | Total | 24647,80 |
| | (SE)[Ha]* | 0,38 | 3,41 | 13,84 | 17,92 | 14,56 | | | |
| **ACB_B 2024M PA** | | | | | | | | | |
| Mode Pixel count | Area [Ha] | 0,06 | 18,98 | 1179,93 | 230,11 | 33,24 | | Total | 1462,32 |
| | Area% | 0,00 | 1,30 | 80,69 | 15,74 | 2,27 | | | |
| Mean Pixel count | Area [Ha] | _0,04_ | 19,71 | 1176,61 | 232,41 | 33,56 | | Total | 1462,32 |
| | (SE)[Ha]* | 0,01 | 0,30 | 0,79 | 0,78 | 0,37 | | | |

*Pixel count SE was calculated as $SE = \sigma/\sqrt{n}$, where $\sigma$ is the standard deviation and $n$, the number of iterations.

Underlined results in italics indicate that the 20 iteration results were not normally distributed.

# B.4. Epoch LULC & Change Maps



1991T LULC Class Map

Assessment PA

LULC Class [code]
- Water [0/W]
- Bare-Rock-Built [1/B]
- Agriculture-Grass [2/A]
- Forest [3/F]
- Shrubland-Wetland [4/S]

1991T LULC Agreement Map

Assessment PA

LULC Agreement
- 50% and below
- 50 to 60%
- 60 to 70%
- 70 to 80%
- 80 to 90%
- greater than 90%

2001T LULC Class Map

Assessment PA

LULC Class [code]
- Water [0/W]
- Bare-Rock-Built [1/B]
- Agriculture-Grass [2/A]
- Forest [3/F]
- Shrubland-Wetland [4/S]

2001T LULC Agreement Map

Assessment PA

LULC Agreement
- 50% and below
- 50 to 60%
- 60 to 70%
- 70 to 80%
- 80 to 90%
- greater than 90%

**2015O LULC Class Map**

Assessment PA

LULC Class [code]
- Water [0/W]
- Bare-Rock-Built [1/B]
- Agriculture-Grass [2/A]
- Forest [3/F]
- Shrubland-Wetland [4/S]

**2015O LULC Agreement Map**

Assessment PA

LULC Agreement
- 50% and below
- 50 to 60%
- 60 to 70%
- 70 to 80%
- 80 to 90%
- greater than 90%

**2019M Binary Class High Confidence Map**

- Assessment PA
- Veredas (Districts)
- Watersheds

LULC Binary Class [code]
- Other [1/O]
- Agriculture-Grass [2/A]

**2019-24M Binary Class Bi-temporal Change Map**

- Assessment PA
- Veredas (Districts)
- Watersheds

Change Typology [code]
- Stable Other [101/OO]
- Stepped Gain [102/OA]
- Stepped Loss [201/AO]
- Stable Agriculture-Grass [202/AA]

1991T LULC Class Map

1991T LULC Agreement Map

2001T LULC Class Map

2001T LULC Agreement Map

**2015O LULC Class Map**

Assessment PA

LULC Class [code]
- Water [0/W]
- Bare-Rock-Built [1/B]
- Agriculture-Grass [2/A]
- Forest [3/F]
- Shrubland-Wetland [4/S]

**2015O LULC Agreement Map**

Assessment PA

LULC Agreement
- 50% and below
- 50 to 60%
- 60 to 70%
- 70 to 80%
- 80 to 90%
- greater than 90%

**2019M Binary Class High Confidence Map**

Assessment PA
Veredas (District)
Watersheds

LULC Binary Class [code]
- Other [1/O]
- Agriculture-Grass [2/A]

**2019-24M Binary Class Bi-temporal Change Map**

Assessment PA
Veredas (Districts)
Watersheds

Change Typology [code]
- Stable Other [101/OO]
- Stepped Gain [102/OA]
- Stepped Loss [201/AO]
- Stable Agriculture-Grass [202/AA]

# List of Figures

¡Error! Utilice la pestaña Inicio para aplicar Heading 1 al texto que desea que aparezca aquí.

192

| ¡Error! Utilice la pestaña Inicio para aplicar Heading 1 al texto que desea que aparezca aquí.

193

194

| ¡Error! Utilice la pestaña Inicio para aplicar Heading 1 al texto que desea que aparezca aquí.

# List of Tables

198

¡Error! Utilice la pestaña Inicio para aplicar Heading 1 al texto que desea que aparezca aquí.

# List of symbols

| Variable | Description | SI unit / value |
|:---:|:---:|:---:|
| $\mu$ | *Micro* prefix | [-] / $x10^{-3}$ |
| $k$ | *Kilo* prefix | [-] / $x10^{3}$ |
| $\lambda$ | Wavelength | [m] |
| $\rho$ | Reflectance | [-] |

# Acknowledgments

I would like to acknowledge the significant contribution of Dr France Gerard, of her expertise to support the methodological development of the Thesis, and her constant feedback, encouragement and supervision.

I would also like to recognise the support I received from Prof. Giovana Venuti and her disposition to assist the remote development of this work.

I would also like to thank Prof. David Edwards for his valuable insight in the early development of the proposals and objectives for this Thesis

I dedicate this work to my partner Kiren and the countless hours we dedicated to the revision of this work, for her patience and unconditional support. To my parents and all my supporting family in Colombia, and my Geoinformatics friends Dori, Ernesa, Juan and Niki. We made it!