Novel methods and technologies to improve monitoring and understanding of land use land cover dynamics based on satellite Earth Observation. The case of forest monitoring in Colombia

Autor

Carlos Alberto Pedraza Peñaloza

Director

Nicola Clerici

Doctorado en Ciencias Biomédicas y Biológicas

Facultad de Ciencias Naturales

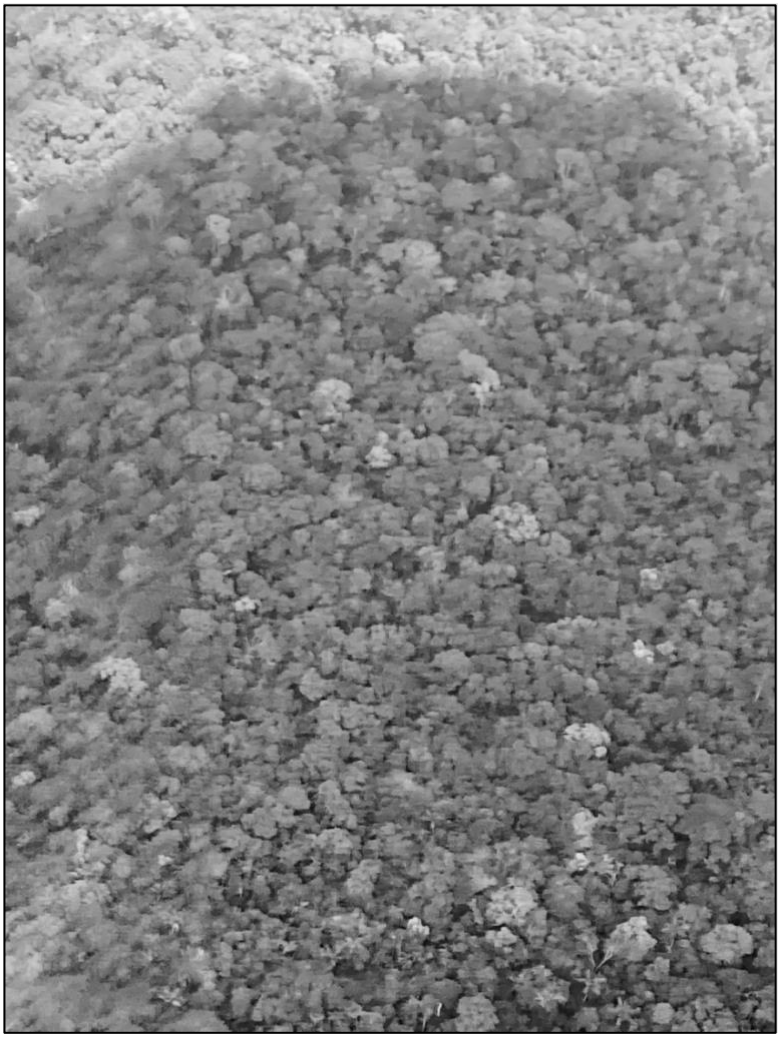
Programa Biología

Universidad del Rosario

Bogotá - Colombia

2024

Table of contents



Chapter 1: Introduction

1. Introduction

a. Approach and objective of the study

The general objective of this project is to develop and apply novel approaches related to satellite Earth Observation (EO) sensors, analytical methods, and technologies to manage and process Big Data providing consistent, robust, and cost-efficient monitoring systems for projects aiming to reduce deforestation and forest degradation.

To achieve the objective of this study we propose innovative methods associated to the space segment (how efficiently manage and process of EO big data) as well as the ground segment (apply analytical methods to provide users with EO product value-added products (Figure 1).

In recent decades, international environmental policies and the potential application/benefit of remote sensing data to multiple sectors, has promoted the development of unprecedented new satellite missions to generate EO data, due their unique position to observe, measure and monitor how, when and where land is changing across the globe (Townshend, Latham, & Arino, 2008).

One first approach of innovation related to the space segment is to apply EO data based on Synthetic Aperture Radar (SAR) systems to forest monitoring and assess the fulfillment of zero deforestation agreements in Colombia (Figure 1, phase 1 represents the use of Advanced Land Observing Satellites (ALOS) for zero deforestation agreements assessment at farm level). SAR systems, all-weather technologies, represent an innovation in EO especially in tropical regions compared to optical / passive sensors, by dealing with two factors: frequent cloud cover and rapid vegetation regrowth (Deutscher et al., 2017). The presence of frequent clouds in tropical regions generates temporal and spatial gaps in optical imagery for some regions, which in conjunction with rapid vegetation regrowth in the same areas and sensor biomass saturation (Joshi et al., 2017) can lead to underestimation of deforestation change detection. SAR systems have important advantages due the capabilities of providing a cloud-free view of Earth's surface, with data acquisition during day/night. Additionally, recent improvements in the implementation of systematic observation strategies based on SAR ensure high dense time series with consistent EO data provision to generate value-added products. Recent missions provided public SAR data, like Sentinel-1 (between 6-12 days of revisit time), or commercial missions such as TerraSar-X (between 11-22 days of revisit time), ALOS-2 (revisit time of 14 days). To provide forest monitoring systems based on SAR imagery with the adequate quality of noise reduction (speckle reduction), temporal filtering, co-registration, radiometric and terrain correction, specific pre-processing routines were implemented to project specific needs that ensure high quality SAR products for subsequent analysis.

The high rate of data provision and access to SAR imagery due to systematic observation strategies improvements in conjunction with the emergence of cloud computing, facilitated the development of highly efficient processing environments to access, pre-process, structure, analyze and distribute large remote sensing datasets. To achieve the impact of improvement of deforestation monitoring systems to ingest, process remote sensing big datasets and generate near real time value-added products, a cloud-based environment are necessary to store and process large volumes of SAR

time series datasets, allowing extensive resources optimization to scale the application of multiple analytical routines (Figure 1, phase 2 represents cloud processing and storage environment for the project).

Pre-processing routines for SAR data and cloud-based environment for data storage/processing and, analytical routines for forest dynamics monitoring were parametrized to fulfill the specific needs of the multiple actors associated to deforestation and forest degradation reduction strategies (e.g. gain and loss detection and mapping, deforestation early warning alerts, deforestation area estimation). Based on multiple users' typologies with specific needs and applications, specific analytical approaches need to be developed and parametrized, e.g.: forest and deforested area estimations, early warning systems, reporting, law enforcement. The cloud-based storage and processing and the use of open source applications will contribute to develop a highly efficient and elastic system that can have an impact on how forest monitoring systems operate and will provide innovative approaches to generate robust, consistent and cost-efficient value-added products (Figure 1, phase 3 show the Python open source programming language to be used to develop the analytical procedures documented through Jupyter Notebooks and GIT HUB).

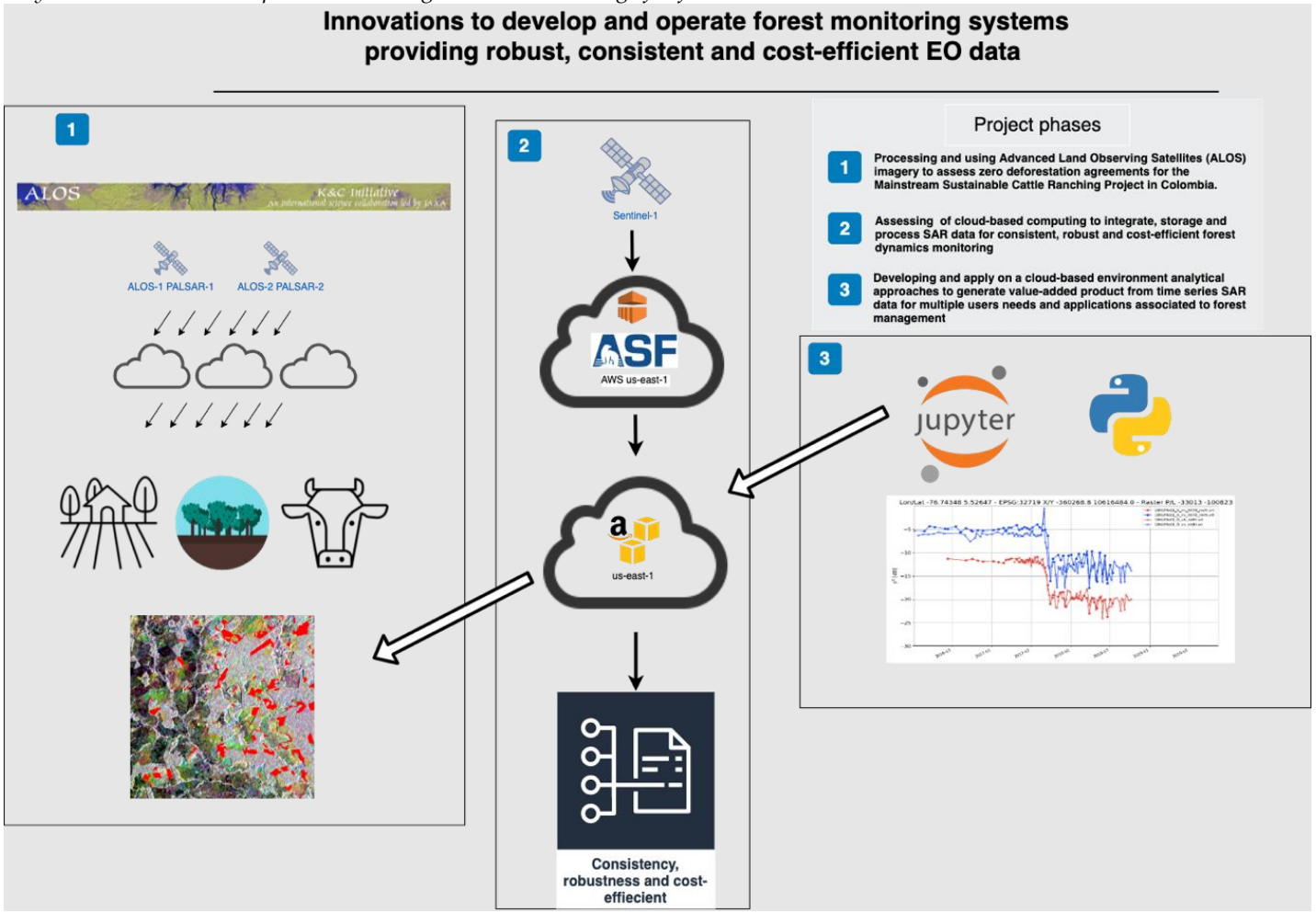


Figure 1. Project phases, involving the development and application of analytical approaches for specific user's needs for forest monitoring in cloud-based computing environment.

Satellite observation strategies, cloud-based computing and the advance in analytics can now provide highly efficient, scalable, and elastic platforms for the implementation of monitoring systems. *Efficient* refers to the capacity of a system to rapidly process, analyze and generate products from large volumes of data. Moreover, when the concept of *scalability* is integrated, i.e. the capacity of the system to replicate its functions to analyze any place on the planet, the expected data storage and processing resources needed can exponentially increase in time. Additionally, forest dynamics research based on remote sensing big data needs to be supported by *elasticity,* to tactically resource allocation and manage different applications, e. g. above ground biomass estimation, forest/no forest/deforestation/regeneration area estimators, time series change detection, etc. To explore the scalability and elasticity concepts for a cloud-based computing system integrating remote sensing data and multiple analytical methods, the study areas, and scales (temporal and spatial) for this project will vary specifically to each of the applications addressed (each study area will be described in the methods section for each chapter). For example, in this work the zero deforestation agreements assessment was carried out at a farm scale for a six (6) year's period in multiple regions of Colombia; to assess cloud-based computing to generate near real-time early warning system to monitor large areas, multiples deforestation hotspots in the Amazon region of Colombia will be analyzed.

1.1. Justification

About 42.9% of global forests are concentrated in the tropics (1,770,156 K Ha in 2015, Keenan et al., 2015), which relates also to the vast majority of forest loss: about 6.4 M ha year−1 between 2010 and 2015 (Keenan et al., 2015), contributing to 32% of the total global forest loss from 2010 to 2012, based on Hansen et al. (2013). Nearly half of tropical forest loss reported occurs in South America (Hansen et al., 2013). In Colombia, the current land cover and land use patterns are result of multiple, heterogeneous historical processes (Andrés Etter, McAlpine, Wilson, Phinn, & Possingham, 2006). Historically, initial colonization and urbanization started from lowland landscapes, followed by the introduction of cattle practices (Andres Etter, McAlpine, Pullar, & Possingham, 2006). At the beginning of the 20th century the agricultural footprint rapidly increased due to population growth, where cattle ranching played an especially important role in landscape change dynamics within the country (Andrés Etter et al., 2006). It has been estimated that approximately one third of forest cover has been cleared since year 1700 (Andrés Etter et al., 2006), with a notable acceleration of deforestation rates in recent years (Armenteras, Espelta, Rodríguez, & Retana, 2017), (Clerici et al., 2020).

As a response to the rapid advance of global forest loss and degradation, the UN Framework Convention on Climate Change (UNFCCC) launched the Reducing Emissions from Deforestation and Forest Degradation program (REDD+). The general aim of REDD+ is to contribute to the mitigation of climate change by reducing greenhouse gas (GHG) emissions by decreasing and reversing forest loss and degradation, and by increasing the removal of GHGs through conservation and the expansion of forests (Muradian, Corbera, Pascual, Kosoy, & May, 2010). REDD+ potentially represents a valuable incentive to reduce GHG emissions and simultaneously promote sustainable forest management (Salvini et al., 2014). In 2008, the national government of Colombia in collaboration with the UN launched the UN-REDD program in the country; since then, multiple collaboration initiatives, promoted especially by NGOs and multilateral organizations, implemented environmental programs based on the REDD+ approach and presented the Readiness Preparation Proposal for Colombia in 2013 (Gutiérrez et al., 2018). Since then, several ONG´s, national, regional, and local governments, civil society and international collaboration agencies have been promoting programs and initiatives based on policy, research and investments to reduce forest degradation and deforestation motivated by REDD+ framework. To effectively mitigate climate change, the countries participating in the REDD+ initiative need to implement actions addressed to drivers of forest degradation and deforestation. Since conference COP 13 (Bali 2007), all participants, acquired the

compromise to implement mitigation actions that could be measured / monitored, reported and verified, so to generate results that can be demonstrable, transparent and verifiable (Gutiérrez et al., 2018).

As part of the implementation of a national REDD+ program the government of Colombia in 2010 launched the National Forest and Carbon Monitoring System (SMByC) to monitor the extent of forest and carbon stocks, generate the national greenhouse gases inventory, build the national forest inventory, and understand the drivers of forest degradation and deforestation, among others. Since his launch, the SMByC has been documenting the extent of deforested area in Colombia. From 1990 to 2020 an estimation of 7,463,162 ha where deforested in Colombia, equivalent to the 6.44% of the continental territory of the country; where the variation of total deforested area is: 2,654,459 ha (38.2%) from 1990 to 2000, 2,987,884 ha (42.9%) between 2000-2010, and 1,820,819ha (28.9%) during the period 2010-2021 (Figure 2).

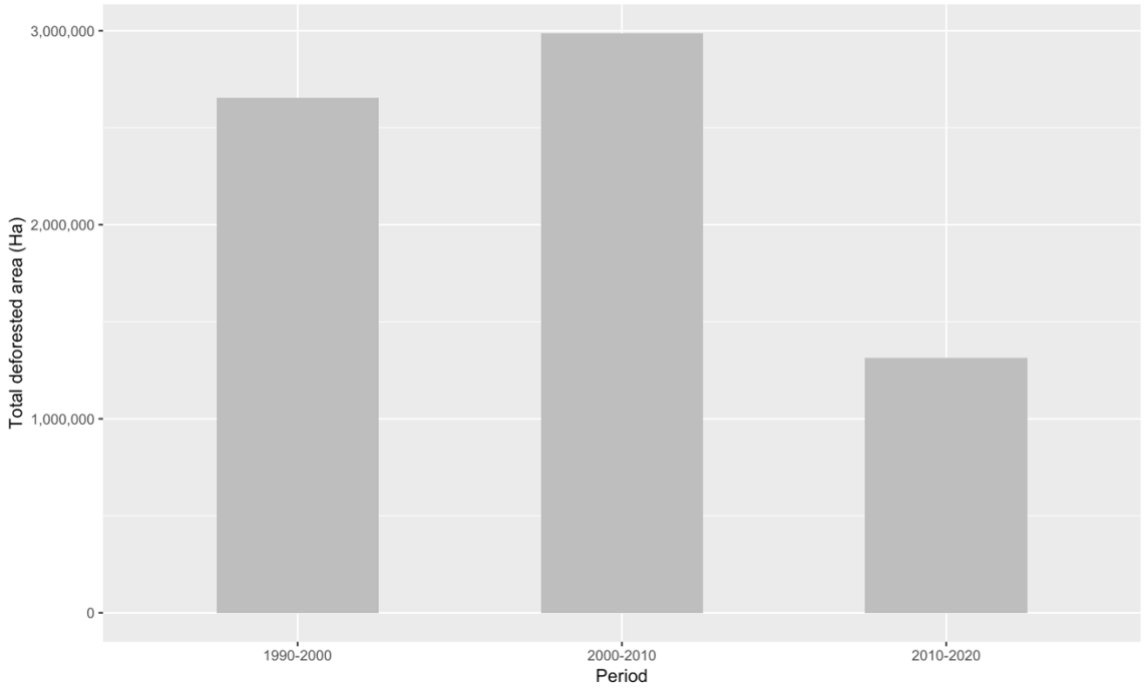


Figure 2. Total deforested areas in Colombia during the last three decades for the periods 1990-2000, 2000-2010 and 2010-2021. Data provided by SMByC (Cabrera, Vargas, Galindo, García, & Ordoñez, 2011)

During the last 8-year period, between 2013 and 2021, the rates of deforestation (ha year-1) were estimated in Colombia by the SMByC, with the highest mean extension between 2016-2017 (219,975 ha year-1, Figure 3). The lowest historical rate of deforestation in 2013, 2014, and 2015 period (120,939, 140,353, and 124,034 ha year-1 respectively) and a recent stability on the last two years reported (2020 and 2021) with 1747,19 and 174,102 hectares deforested respectively (Figure 3). Considering the rates of deforestation at the department level, current rates of deforestation presented a recent (± from 2012) reduction compared with past rates (1990-2010), Figure 4. Nevertheless, over the last 6 years some departments increased their rates of deforestation; during 2016-2017 Caquetá, showed the highest rate of deforestation: 60,373 ha (27.44 % of total deforested area for this period); Meta during the 2017-2018 period 44,712 ha (22.68%), Guaviare during the 2016-2017 period 38,221 ha year-1 (17.37%) and Choco during 2015-2016 with 24,025 ha (13.45%), Figure 4 (Cabrera et al., 2011).

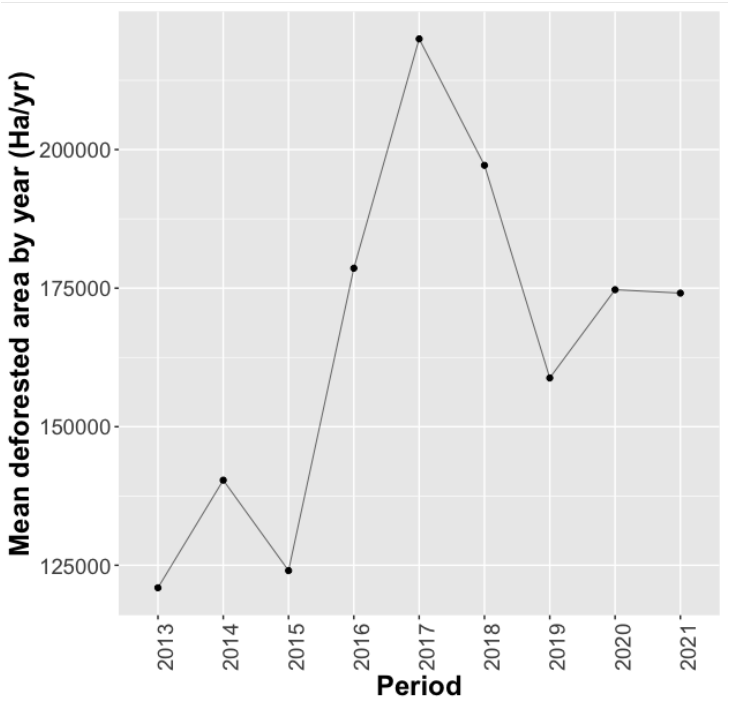


Figure 3. Mean deforested area (hectares year-1) in Colombia 2013-2021. Data provided by SMByC (Cabrera et al., 2011).

Different deforestation trends across geographies are the response of multiple complex social, ecological, economic, and political contexts. The causes behind deforestation are classified as proximate and distal or underlying, where the governance of forest and decision-making are very dynamic fields that requires the integration of ecological and social sciences (Nielsen, 2016)(Robinson, Fuller, Stedman, Siemer, & Decker, 2019). Natural resources management focused to reduce forest degradation and deforestation emissions, rarely integrates social and ecological sciences to achieve its goals. Information related to forest trends are critical to multiple actors and different phases involved in the decision-making of policies and investments promoting the conservation of forests and their ecosystem services. Understanding the dynamics of land-use and land-cover change is a key research topic in environmental science globally (Geist & Lambin, 2001). Knowing location, quantity and the time of deforestation events are crucial elements in many phases of the decision-making process at local, regional, and national scales.

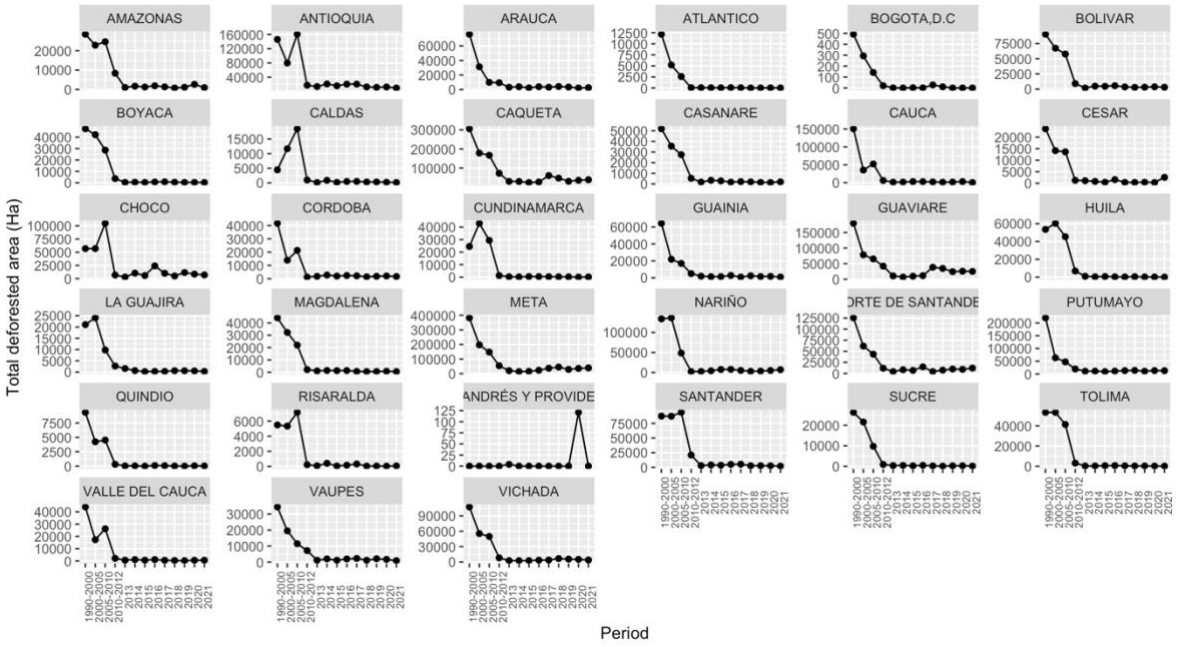


Figure 4. Rates of deforestation (hectares year-1) for all 32 departments of the country. The estimation of mean deforested area for 1990-2000, 2000-2005, 2005-2010, 2010-2012, 2012-2013, 2013-2014, 2014-2015, 2015-2016, 2016-2107, 2017-2018, 2018-2019, 2019-2020, and 2020-2021. is a result of the total deforested area of the period divided by the number of years of the period (Cabrera et al., 2011).

REDD+ implementations must be coordinated by national governments with activities at sub-national scales developed in collaboration with government agencies and promoted by local private or public actors (Corbera & Schroeder, 2011). To achieve an impact on a global scale from local implementation, unprecedented financial and technical support from multilateral and bilateral funds have been implemented. The advances in remote sensing science during the 21st century have provided unprecedented opportunities for EO (e.g. from Landsat to the Sentinel constellation), the advance of algorithms with numerous and diverse applications (e. g. from Maximum Likelihood and unsupervised clustering to Random Forest through Object-oriented classifiers to Machine Learning and Deep Learning), and the access to high performance scalable computing resources to handle, process and distribute remote sensing big data; all these advances can potentially contribute not just to the understanding of land systems, but also to assess the consequences of decisions connected with normative processes (Dong, Metternicht, Hostert, Fensholt, & Chowdhury, 2019).

Such advances in EO science and technology can entail a real paradigm's shift, introducing *the fourth paradigm*, as Hey, Tansley, & Tolle (2009) called it. Nativi et al. (2015) suggest that, to make it reality, it requires innovative approaches to implement technologies for the management, analysis, delivery, and presentation of large amounts of data, providing decision makers an extraordinary range of data and information. Nevertheless, to have an impact in actions associated to the decision-making process, it requires more than providing data and information at high-speed rates to users. Platforms where users interact with data and information, are necessary to ingest remote sensing big data and analytics, but also designed with functionalities and data visualization that fits users' needs. In the case of REDD+ and similar initiatives, multiple actors are involved in the decision-making process (e.g. local communities, private sector, regional and national agencies); every actor has different responsibilities and needs related with forest dynamics of monitoring (early deforestation change detection, deforested area estimations, local participatory ground truth data collection, dates of forest change).

1.2. Goal

The general objective of this project is to develop and apply innovative approaches related to satellite Earth Observation sensors, develop analytical methods and technologies to manage and process Big Data, providing consistent, transparent, robust, and cost-efficient monitoring systems for projects aimed to reduce deforestation and forest degradation. This project will also design a robust, consistent, and cost-efficient forest monitoring system, that will integrate multiple analytical approaches to provide spatially explicit value-added products to actors and users involved in initiatives to reduce forest degradation and deforestation.

Specific objectives of the project are:

1. Develop and apply a methodology to assess the compliance to zero deforestation agreements, applied to 2615 livestock farms associated with the Sustainable Colombian Livestock project through the processing of Advanced Land Observations Satellites (ALOS) Phased Array Type L-band Synthetic Aperture Radar (PALSAR) imagery.

2. Design and evaluate a cloud-based computing infrastructure to ingest remote sensing big data (big volume, big velocity, big variety; Nativi et al., 2015) for forest monitoring. The assessment of the digital infrastructure or cloud will be based on technical specifications and financial resources needed to be fully operational.

3. Develop and integrate in the cloud-based digital infrastructure, analytical approaches and algorithms to generate specific value-added forest products for multiple actor and users involved in initiatives associated to reduction of forest degradation and deforestation.

1.3. Structure of the thesis

The first chapter is an introduction to deforestation trends in Colombia. The second chapter focus in the development and application of a methodology to integrate remote sensing SAR EO to assess 2615 zero deforestation agreements implemented by the Mainstream Sustainable Cattle Ranching Project in Colombia; the third chapter, is focus to configure a digital infrastructure with capabilities to integrate, structure and process EO Big Data, providing analytical products that can have an impact on projects providing consistency, robustness and cost-efficient within a cloud-based environment. The fourth chapter's is to develop specific processing routines using highly dense time series EO data, and to develop products for specific needs of users, specifically in this case study forest dynamics and evaluate restoration implementations.

Aerial view Caquetá, Florencia. Photo By Carlos Pedraza (2020)

Chapter 2: Zero deforestation agreements assessment

This chapter was designed to integrate EO data generated by pathfinder radar mission ALOS-PALSAR to assess zero deforestation agreements in multiple regions of Colombia where different atmospheric and topographic conditions are present. Specific workflows were developed to pre-process radar imagery and generate thematic products for zero deforestation assessment at farm level.

This chapter was published the 13th of September 2018 in the Remote Sensing journal as part of the Special Issue "The Kyoto and Carbon Initiative-Environmental Applications by ALOS-2 PALSAR-2". Available at: https://www.mdpi.com/2072-4292/10/9/1464.

Also was presented at the International Geoscience and Remote Sensing Symposium (IGARSS) for the Forests and Biomass from Space Session of Land Applications, July 16th 2021, https://igarss2021.com/view_paper.php?PaperNum=3736#top.

These zero deforestation agreements assessment work is a collaboration with the Colombian Mainstream Sustainable Cattle Ranching project (MSCR), where Carlos Pedraza worked as a spatial and monitoring specialist for seven years (2012-2018) as the spatial and monitoring expert Carlos defined the methods to assess the project zero deforestation agreements based on Advanced Land Observing Satellites (ALOS) imagery and ground truth data and developed ground truth data collection protocols. Carlos leaded multiple workshops on all five regions to train local technical researchers in protocols for ground truth data collection, in addition to process and analyze ALOS-1 PALSAR-1 data.

The MSCR aims are to improve the ecosystem functioning of degraded pastures lands through the implementation of sustainable silvopastoral practices, contributing to national goals to reduce the total cattle ranching land, contribute to climate change mitigation, as well as to generate socioeconomic benefits. The MSCR project integrates small holder cattle ranching farmers in a payment for the environmental services scheme (PES), where farmers have compromised through the signing of a contract to a zero-deforestation agreement inside the farms during the project's life. Farmers receive materials, technical assistance, and PES associated with the establishment of silvopastoral systems and the restoration/conservation of areas that include forest. During the technical assistance phase project's staff monitored the fulfillment of the zero-deforestation agreements by field inspection of the farm's forest areas. The integration of a further assessment based on remote sensing imagery provides the required independent key performance indicator of estimation of deforestation extent at the project level, complementing field deforestation monitoring at the individual farm level. The MSCR project is supported by several international institutions including the Global Environment Facility (GEF), the UK's Department of Energy and Climate Change (DECC), the World Bank (financial support and supervision), national Colombian agencies like the Center for Research in Sustainable Systems of Agricultural Production (CIPAV), the National Federation of Cattle Ranching, (FEDEGAN), the Action Fund for Environment and Children, and The Nature Conservancy.

The present research involved the following collaborations:

Sistema de Monitoreo de Bosques y Carbono de Colombia (SMByC). Through the collaboration with the SMByC specific knowledge related to national standards related to forest definition and digital image processing were implemented in the methodological approach for forest dynamics mapping and area estimation.

The Nature Conservancy. Information related to ground truth and farm delimitation was provided by the MSCR through collaboration with The Nature Conservancy.

Article

Zero Deforestation Agreement Assessment at Farm Level in Colombia Using ALOS PALSAR

Carlos Pedraza [1,2,*], Nicola Clerici [1], Cristian Fabián Forero [3], América Melo [2], Diego Navarrete [2], Diego Lizcano [2], Andrés Felipe Zuluaga [2], Juliana Delgado [2] and Gustavo Galindo [3]

[1] Universidad del Rosario; Biology Program, Faculty of Natural Sciences and Mathematics, Carrera 26 # 63B-48, Bogotá, 111221 , Colombia; nicola.clerici@urosario.edu.co

[2] The Nature Conservancy; Northern Andes and Central Southern America program. Calle 67 # 7-94, Bogotá, 110231, Colombia; america.melo@tnc.org (A.M.); diego.navarrete@tnc.org (D.N.); d.lizcano@tnc.org (D.L.); andres.zuluaga@tnc.org (A.F.Z.); jdelgado@tnc.org (J.D.)

[3] Instituto de Hidrología, Meteorología, y Estudios Ambientales-IDEAM, Sistema de Monitoreo de Bosques y Carbono, Calle 25 D No. 96 B-70, Bogotá, 110911, Colombia; cforero@ideam.gov.co (C.F.F.); ggalindo@ideam.gov.co (G.G.)

* Correspondence: cpedraz@gmail.com; carlosa.pedraza@urosario.edu.co; Tel.: +57-1606-5837

Received: 28 July 2018; Accepted: 11 September 2018; Published: 13 September 2018

Abstract: Due to the fast deforestation rates in the tropics, multiple international efforts have been launched to reduce deforestation and develop consistent methodologies to assess forest extension and change. Since 2010 Colombia implemented the Mainstream Sustainable Cattle Ranching project with the participation of small farmers in a payment for environmental services (PES) scheme where zero deforestation agreements are signed. To assess the fulfillment of such agreements at farm level, ALOS-1 and ALOS-2 PALSAR fine beam dual imagery for years 2010 and 2016 was processed with ad-hoc routines to estimate stable forest, deforestation, and stable nonforest extension for 2615 participant farms in five heterogeneous regions of Colombia. Landsat VNIR imagery was integrated in the processing chain to reduce classification uncertainties due to radar limitations. Farms associated with Meta Foothills regions showed zero deforestation during the period analyzed (2010–2016), while other regions showed low deforestation rates with the exception of the Cesar River Valley (75 ha). Results, suggests that topography and dry weather conditions have an effect on radar-based mapping accuracy, i.e., deforestation and forest classes showed lower user accuracy values on mountainous and dry regions revealing overestimations in these environments. Nevertheless, overall ALOS Phased Array L-band SAR (PALSAR) data provided overall accurate, relevant, and consistent information for forest change analysis for local zero deforestation agreements assessment. Improvements to preprocessing routines and integration of high dense radar time series should be further investigated to reduce classification errors from complex topography conditions.

Keywords: carbon cycle; deforestation; Colombia; sustainable cattle ranching; Synthetic Aperture Radar, ALOS PALSAR

1. Introduction

About 44% of global forests are concentrated in the tropics (1,770,156 thousand ha in 2015 [1]), and is also where the vast majority of forest loss occurs, with reported rates of loss of 6.4 M ha year−1 between 2010 and 2015 [1]. In Colombia, approximately one third of forest cover has been cleared since the year 1700, as a result of multiple, heterogeneous historical processes [2]. At the beginning of the 20th century the agricultural footprint rapidly increased due to population growth; cattle ranching played an especially important role in landscape change dynamics within the country [2]. Currently, ranching represents one of the key economic subsectors in Colombia, contributing to approximately 3.5% of the overall Gross Domestic Product (GDP) and 27% of the agricultural and livestock GDP [3]. Cattle ranching exploited more than 38 million hectares over the last 50 years, holding approximately 23.5 million heads, supporting 7% and 28% of national and rural employment, respectively.

Information related to forest trends are critical to different actors involved in the decision-making of policies and investments promoting the conservation of forests and their ecosystem services. Globally, several efforts have been put in place to develop consistent and robust methodologies to assess forest extension and change [4–10]. As a response to the rapid advance of global forest loss and degradation, the UN Framework Convention on Climate Change (UNFCCC) launched the Reducing Emissions from Deforestation and Forest Degradation program (REDD+). The general aim of REDD+ is to contribute to the mitigation of climate change by reducing greenhouse gas (GHG) emissions by decreasing and reversing forest loss and degradation, and by increasing the removal of GHGs through conservation and the expansion of forests [11]. In 2008, the national government of Colombia in collaboration with UN launched the UN-REDD program in Colombia; since then, multiple collaboration initiatives, promoted especially by NGOs and multilateral organizations, implemented environmental programs based on the REDD+ approach that presented the Readiness Preparation Proposal for Colombia in 2013.

Dominating the forestry-based climate mitigation programs in Latin America and the Caribbean [12], REDD+ programs face multiple challenges for their operational implementation and the achievement of multiple goals involving climate change, biodiversity conservation, and sustainable development [13]. For effective implementation and assessment of such programs it is often necessary to obtain systematic Earth Observation-based forest data, together with specific methods and protocols integrating ground truth, geospatial information, and capacity building to ensure the project's monitoring, reporting, and verification (MRV) [14,15]. Consequently, several international partnerships, like the Global Forest Information Initiative (GFOI), have been established to provide national forest monitoring systems with guidelines to exploit Earth Observation data for REDD+, in order to foster robust, reliable, and achievable forest monitoring and assessment [16].

The Kyoto & Carbon (K&C) initiative, an international collaborative project led by the JAXA Earth Observation Research Center (EORC), was designed to contribute data and information to the UNFCCC Kyoto Protocol and international actors for the development of a Terrestrial Carbon Observing system, together with giving continuation to the previous initiatives, such as the Global Rain Forest Mapping (GRFM) and the Global Boreal Observing Satellite (GBFM) [17,18]. The K&C research is based on the Advanced Land Observing Satellites (ALOS). ALOS carries on board the active sensor Phased Array L-band SAR (PALSAR). ALOS PALSAR is considered a pathfinder global monitoring mission due the improvement of sensor performance and its systematic data-observation strategy, providing reliable wall-to-wall observations at fine resolution with consistency in spatial and temporal resolutions [18,19]. This ensures land observation acquisition through multiple

missions, i.e., ALOS-1 (2006–2011) and ALOS-2 (2014–present). ALOS PALSAR information has been extensively used in forest applications, such as forest mapping [9], deforestation monitoring [20,21], aboveground biomass estimation [22], and mangrove monitoring [23]. In addition to the advantage of cloud-free imagery provided by the SAR sensors, ALOS L-band provides key information related to forest canopy and surface features [24]. With the ability to penetrate vegetation canopy, ALOS PALSAR L-band sensors, compared to other SAR instruments (e.g. C-band based), are more sensitive to trees' aboveground structural characteristics, providing very suitable Earth Observations data for forest monitoring [4] with a systematic acquisition strategy.

Recent small-scale deforestation patterns found in Amazonian countries [25] have been found to be increasingly related to land cover conversion from small landowners, e.g., Brasil [26]. Current methodological and technical advances in remote sensing [19,20,21] allow the inclusion of robust small-scale deforestation detection in the assessment phase of deforestation monitoring programs. Previous projects exploit ALOS PALSAR data to quantify deforestation patterns at small-scale farm level to detect deforestation events below the hectare [27], use Landsat data to estimate deforestation with a minimum area detection of 6.25 ha [28] to assess farmers' 'no deforestation' compliance agreements.

In this work we exploit ALOS PALSAR data to assess zero deforestation agreements of the Colombian Mainstream Sustainable Cattle Ranching project (MSCR) in five different regions of Colombia. The MSCR aims are to improve the ecosystem functioning of degraded pastures lands through the implementation of sustainable silvopastoral practices, contributing to national goals to reduce the total cattle ranching land, contribute to climate change mitigation, as well as to generate socioeconomic benefits. The MSCR project integrates small holder cattle ranching farmers in a payment for the environmental services scheme (PES), where farmers have compromised through the signing of a contract to a zero-deforestation agreement inside the farms during the project´s life. Farmers receive materials, technical assistance, and PES associated with the establishment of silvopastoral systems and the restoration/conservation of areas that include forest. During the technical assistance phase project's staff monitored the fulfillment of the zero-deforestation agreements by field inspection of the farm's forest areas. The integration of a further assessment based on remote sensing imagery provides the required independent key performance indicator of estimation of deforestation extent at the project level, complementing field deforestation monitoring at the individual farm level. The MSCR project is supported by several international institutions including the Global Environment Facility (GEF), the UK's Department of Energy and Climate Change (DECC), the World Bank (financial support and supervision), national Colombian agencies like the Center for Research in Sustainable Systems of Agricultural Production (CIPAV), the National Federation of Cattle Ranching, (FEDEGAN), the Action Fund for Environment and Children, and The Nature Conservancy. Specific research objectives of this work are:

- To develop an ALOS PALSAR processing workflow to classify forest and generate forest change products at local scales.
- To assess zero deforestation agreements implementation in 2615 farms participating to the Colombian Mainstream Sustainable Cattle Ranching project by exploiting ALOS PALSAR forest-change products.

2. Materials and Methods

2.1. Study Area

Colombia is located in the northwest of the South American continent, with an area of 1.1 million Km2 (the fourth largest country in South America), with coast on both the Caribbean Sea and the Pacific Ocean. Due to its heterogeneous geographic and topographic characteristics the country holds seven different biogeographic regions. With variations in mean annual precipitation (300–10,000 mm)

and altitude (0–5765 m) this environmental heterogeneity is expressed in a large variety of ecosystems [29]. The study area includes five different regions that cover 83 municipalities (Figure 1), which were selected for their high levels of biodiversity and their proximity to strategic ecosystems and protected areas. The Cesar River Valley and the Magdalena River regions contain the last remaining fragments of dry tropical forest, considered one of the most endangered Neotropical ecosystems, currently 8% of its original extent remains in Colombia [30]. Other strategic ecosystems are mountain forests, where oak-dominated forests are characterized by a high level of degradation, the threat of habitat loss, and under-representation in protected areas [31] and wetland systems associated with the Magdalena River. The river system concentrates approximately 80% of the population of Colombia

and represents a key contribution to the country's fisheries [32].

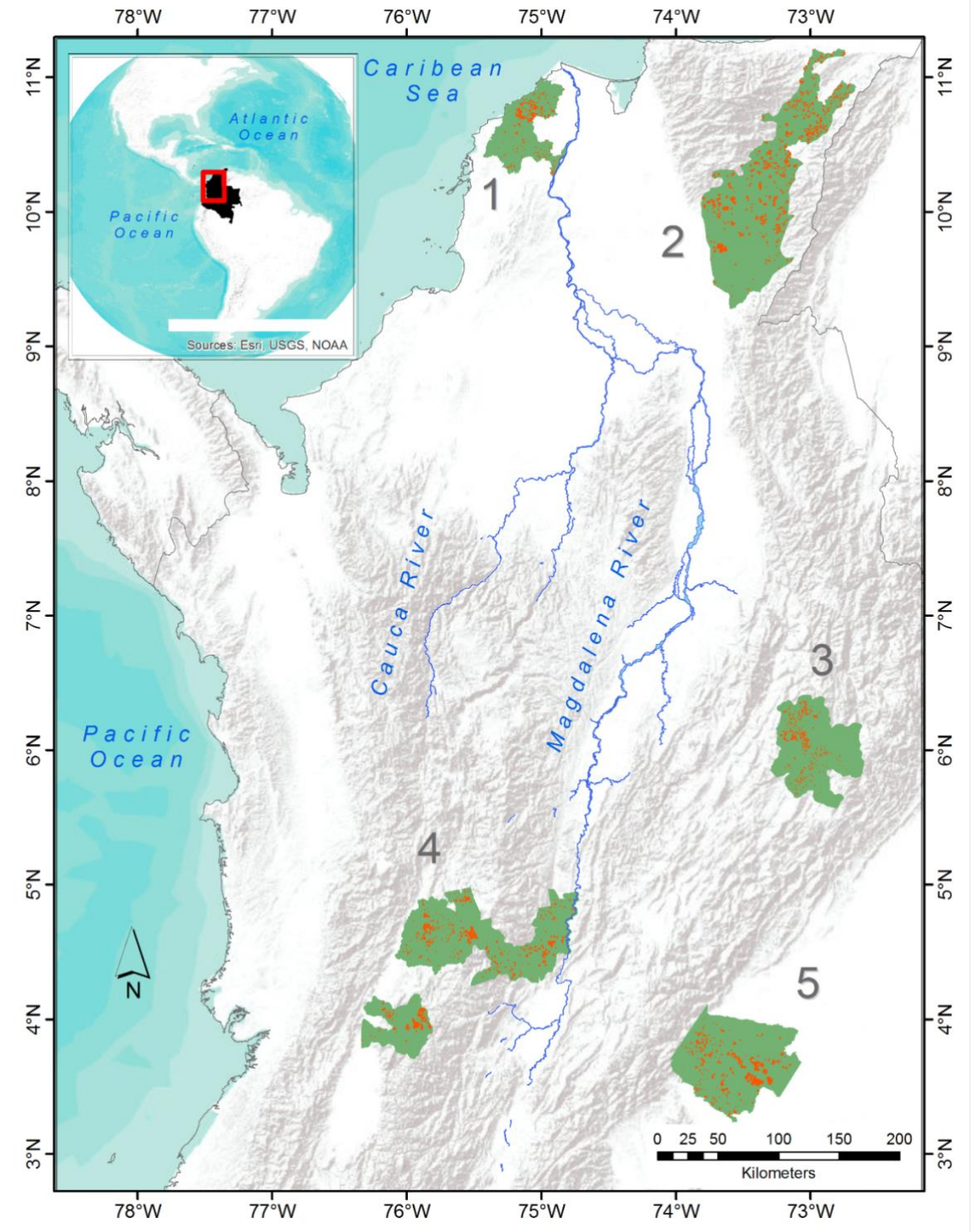


Figure 5. Location of the five regions studied in Colombia (green areas): 1: Bajo Magdalena; 2: Cesar River Valley; 3: Boyacá-Santander; 4: Coffee Ecoregion; and 5: Meta Foothills. Orange dots represent the 2615 farms participating to the Mainstream Sustainable Cattle Ranching project. Geographical Coordinate System WGS 84.

2.2. Data Processing

2.2.1. Project Impact Areas: Farms Delineation

Project impact areas where the zero deforestation agreements were implemented between land owners and the MSCR project (year 2010), correspond to 2615 small and medium-sized farms (Table 1). To determine the farm area to assess agreements in the absence of precise property boundaries, two different protocols for boundaries delimitation were implemented. For 802 farms an accurate boundary delimitation was carried out in situ using GPS technology; boundaries where generated through walks along the farms perimeter and the farm total extent was corroborated with legal documentation and owner feedback. For the remaining 1813 farms, approximate circular boundaries were generated based on a circle center determined using GPS, and on a radius estimated from the farm area reported by legal cadastral documents. Figure 2 shows an example of property boundary generation based on both protocols. Table 1 provides a description of the 2615 farms adhering to the MSCR project for the five regions under study.

Table 1. Descriptive statistics of the 2615 farms associated with the Mainstream Sustainable Cattle Ranching Project.

Region Name	Farms Total Size (ha) [1]	Farms (n)	Farms Mean Size (ha)	Farms Min Size (ha)	Farms Max Size (ha)	Administrative Departments	Forest Type
Bajo Magdalena (1)	7354	421	17.5	2	217	Atlántico, Bolívar	Dry forest
Cesar River Valley (2)	46,587	692	67.1	3.7	2582	Cesar, La Guajira	Dry forest
Boyacá-Santander (3)	6593	462	14.3	2	250	Boyacá, Santander	Mountain forest
Coffee Ecoregion (4)	29,521	667	44.2	2	1055	Quindío, Caldas, Risaralda, Tolima, Valle del Cesar	Mountain forest
Meta foothills (5)	23,747	373	64.8	3.9	1581	Meta	Foothill forest

[1] Farms size were estimated based on GPS farm boundaries delimitation (802 farms), and by legal documentation (1813 farms).

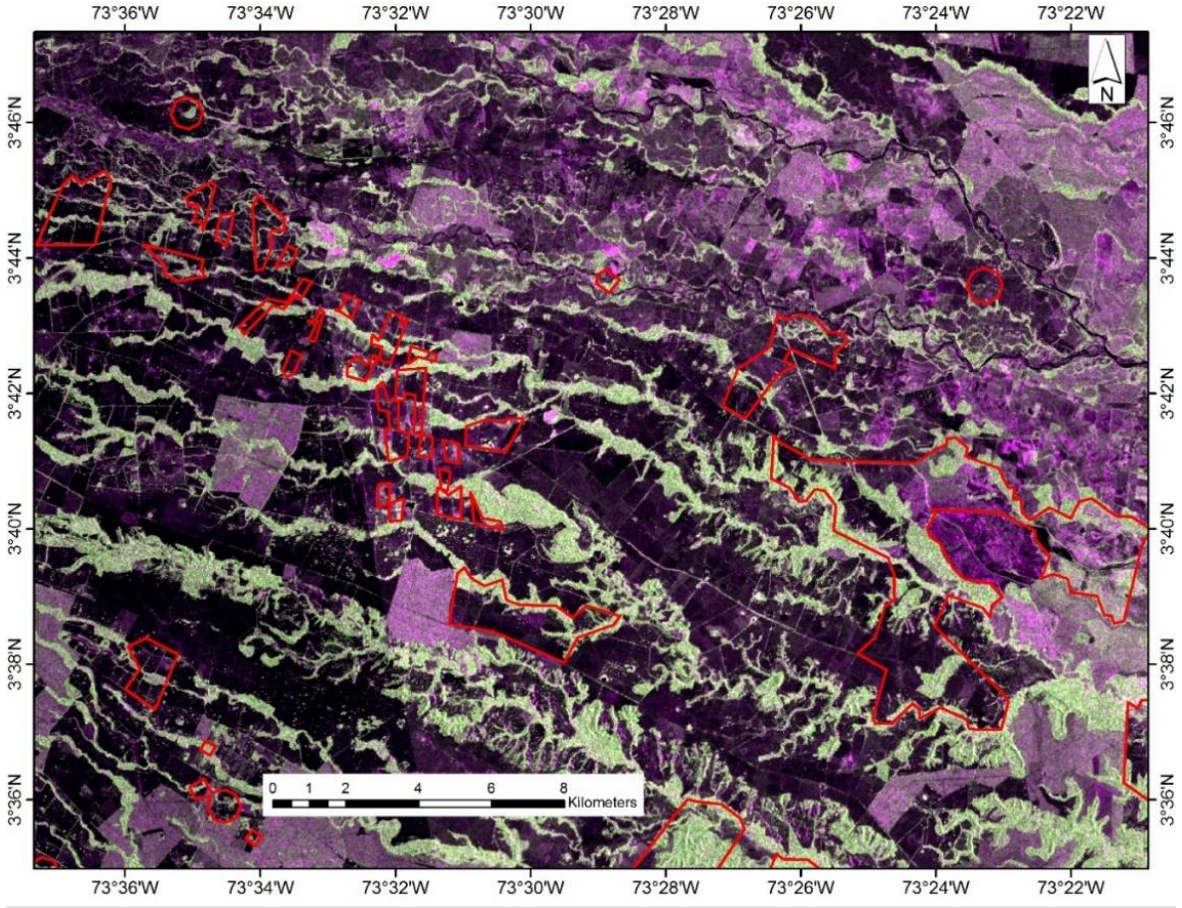


Figure 6. Examples of property delimitation of farms (yellow polylines) in the Meta foothills regions. Polylines were produced using GPS-based precise boundary delineation (802 farms) or by a circle buffer generated using a center GPS location and farm area reported by legal documentation (1813

farms). Background image: ALOS-2 Fine Beam (FCC: Red: HH, Green: HV, and Blue: |HH-HV| polarizations), March 2016. Geographical Coordinate System WGS 84.

2.2.2. Fine Beam Dual ALOS PALSAR Imagery Preprocessing

Zero deforestation agreement assessments require a change analysis to estimate forest extent at the time of the MSCR agreement (2010) with respect to the target year of assessment (2016). ALOS PALSAR imagery was used to classify forest and nonforest cover, and to estimate deforestation extent at farm level for the five regions under analysis. ALOSPALSAR Fine Beam Dual (FBD) imagery was obtained through the Kyoto & Carbon Initiative as part of the collaborative research between The Nature Conservancy and the Japanese Aerospace Exploration Agency (JAXA). Data was accessed through the ALOS User Interface Gateway. A total of 125 SLC (Level 1.1) ALOS-1 and ALOS-2 FBD images were used, covering years 2010 (ALOS-1) and 2016 (ALOS-2) (Supplementary Material, Table S1).

Multiple routines were implemented to convert single look complex (SLC) images into geocoded gamma naught ($\gamma 0$) backscatter intensity images through: multilooking, radiometric, geocoding, and geometric/radiometric correction procedures (Figure 3). Multilooking was performed to SLC imagery complex data to produce multilook intensity imagery, using four range and one azimuth looks as parameters. Radiometric calibration of multilook imagery was applied using different calibration factor for ALOS-1 (−115 dB) and ALOS-2 (−83 dB) [33] for the normalization reference area correction to multilooking gamma naught imagery at a 15 m pixel resolution. Shuttle Radar Topography Terrain Mission (SRTM) 30 m digital elevation model was used for terrain correction and geocoding; additionally, a fine registration is implemented where SAR image is simulated based on the digital elevation model and used to determine the fine registration using a cross correlation analysis. Geocoded, radiometric, and geometric corrected products for ALOS PALSAR imagery products were obtained after preprocessing procedures performed using Gamma software® [34], which provides Synthetic Aperture Radar (SAR) preprocessing routines. Functions and parameters implemented through the software are detailed in Supplementary Material Figure S1.

All imagery was projected to UTM 18N WGS84. An enhanced lee speckle filter was applied to HH and HV polarizations to reduce speckle (Figure 4). The filter was applied by means of least squares of the signal intensity in a kernel area of 3*3 pixels [34]. Mosaicking was finally performed to generate HH and HV polarization mosaics for both years 2010 and 2016. A detailed description of the preprocessing workflow is provided in Supplementary Material Figure S1.

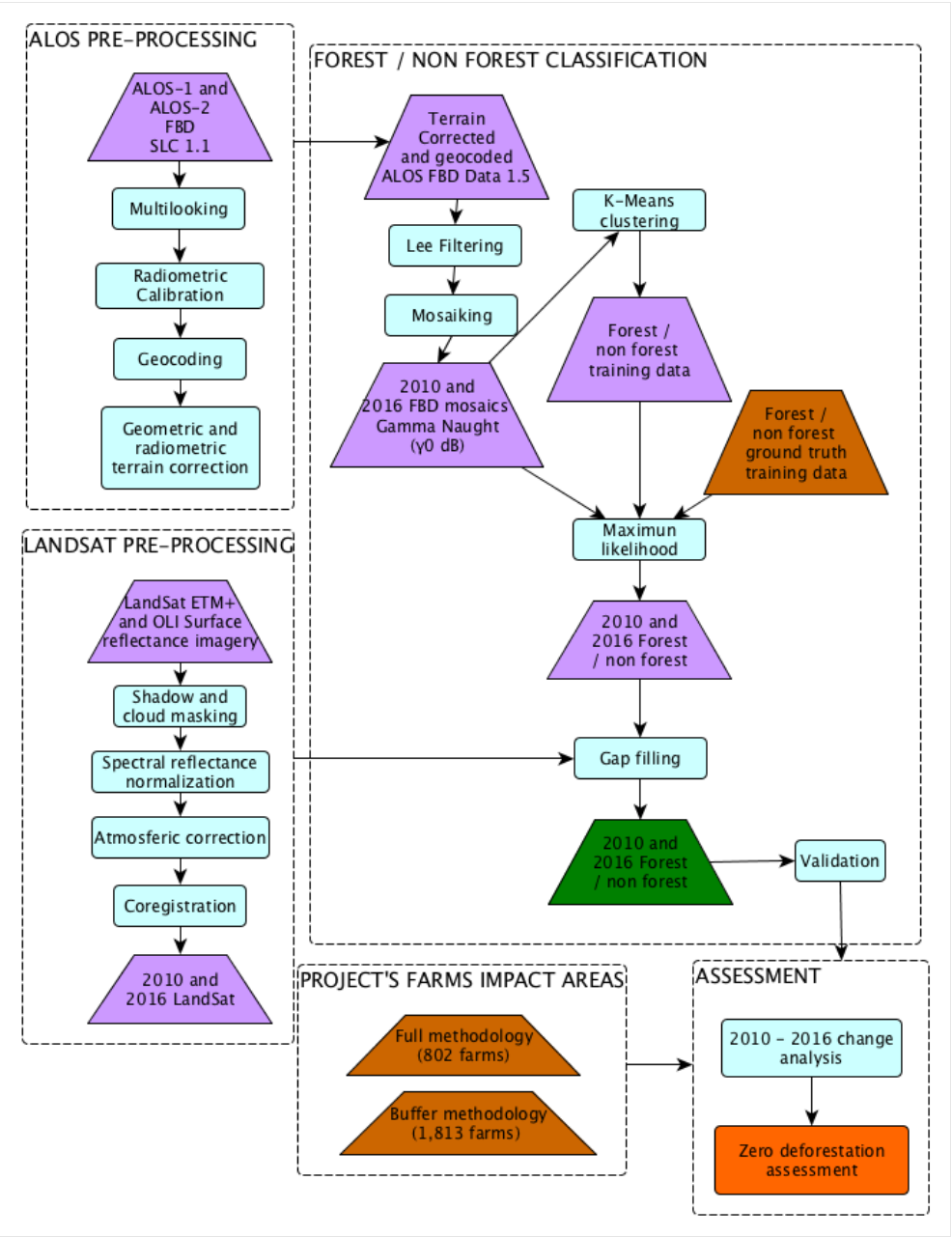


Figure 7. Zero deforestation agreements assessment data processing workflow for the MSCR project, including: ALOS Phased Array L-band SAR (PALSAR) Fine Beam Dual (FBD) preprocessing, Landsat preprocessing, thematic classification, and accuracy assessment.

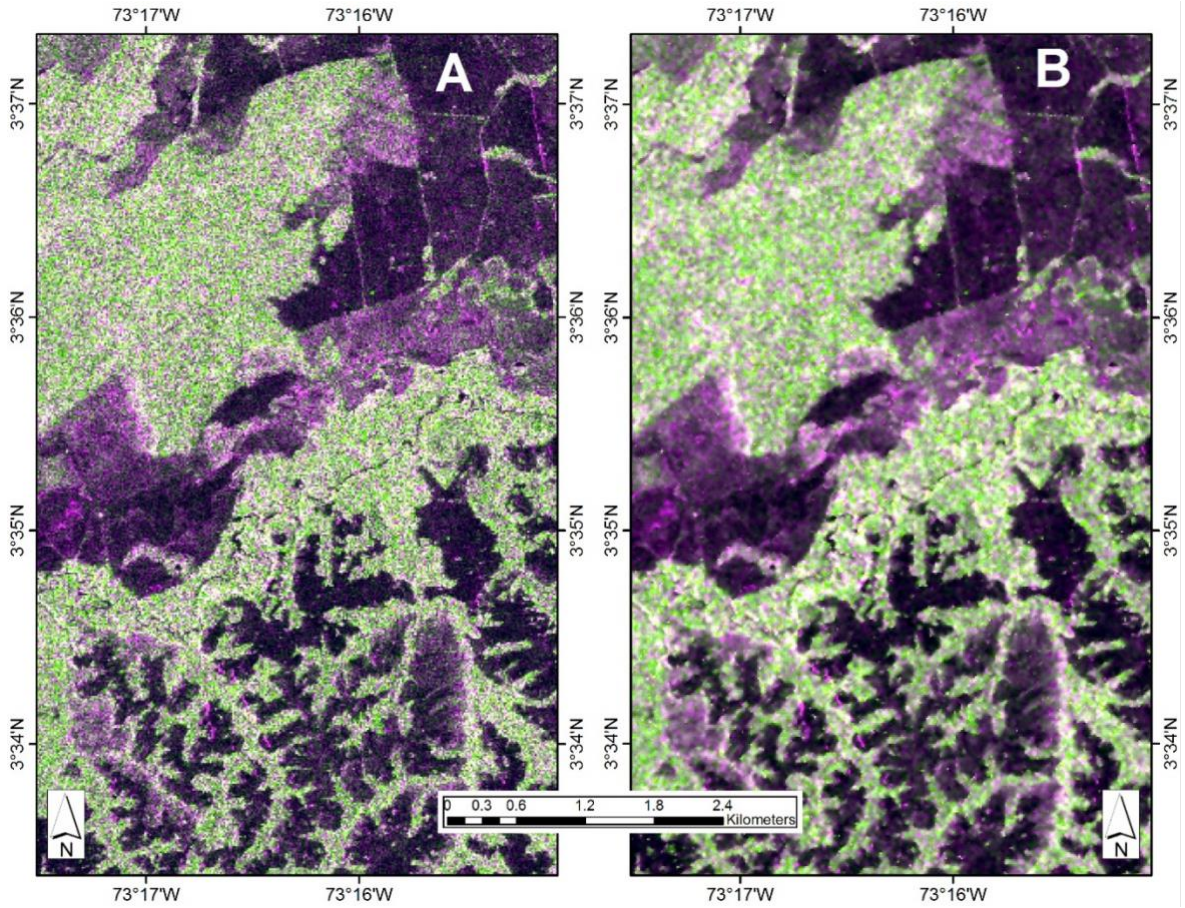


Figure 8. (**A**) ALOS-2 Fine Beam image (FCC: Red: HH, Green: HV and Blue: |HH-HV| polarizations); (**B**) the same image with Lee filter speckle reduction.

2.2.3. Forest Mapping

To estimate the extent of stable forest, stable nonforest, and deforestation (2010–2016) for the five MSCR regions, supervised classification procedures were applied to the FBD ALOS-1 PALSAR-1 (2010) and ALOS-2 PALSAR-2 (2016) geocoded, radiometric, and geometric corrected images. To generate training datasets for the maximum likelihood classifier [36], we followed two different strategies: (i) forest/nonforest GPS-based ground truth data were obtained during the technical assistance field visits of 802 farms, obtaining 1076 forest and 144 nonforest polygons (Figure 5E). (ii) Forest/nonforest regions were derived by a k-means unsupervised classification [37] of gamma naught ($\gamma 0$) dB backscatter intensity values from HH and HV polarizations of FBD imagery (2010 and 2016). Forty classes were initially considered which were then visually merged to derive forest and nonforest regions (Figure 5B–D). Sample regions from the unsupervised classification were selected randomly and homogeneously distributed with respect of the ground truth data training samples previously acquired. Only forest/nonforest regions with no significant temporal variation based on visual assessment of ALOS PALSAR and Landsat images, and Google Earth® (2010–2016) was selected as training data. Ground truth and unsupervised data were merged to train the maximum likelihood classifier [36], in order to produce forest/nonforest maps for years 2010 and 2016.

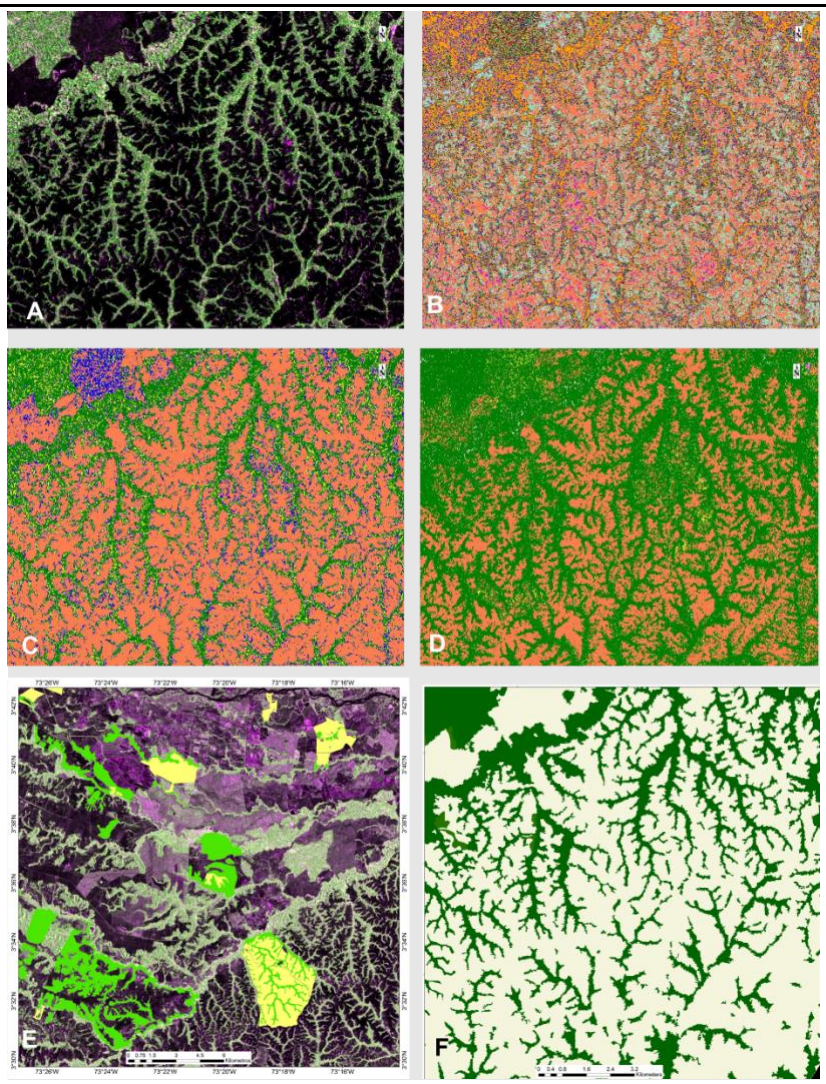


Figure 9. Forest classification Meta foothills example: (**A**) FBD ALOS-2 PALSAR-2 2016 (FCC: Red: HH, Green: HV and Blue: |HH-HV| polarizations); (**B–D**) training data of forest (green) and nonforest based on unsupervised k-means classification; (**E**) ground truth data of forest (green) and nonforest (yellow) obtained during field trip. Training data and ALOS PALSAR imagery were integrated to generate forest nonforest products (**F**) based on a maximum likelihood supervised classification.

Landsat VNIR imagery was integrated in the workflow to reduce misclassifications of classified forest/nonforest areas based on ALOS PALSAR imagery. Landsat scenes were provided by the Colombian National Forest and Carbon Monitoring System. Shadow/cloud masking spectral reflectance normalization, atmospheric correction, and surface reflectance conversion standard procedures [37–42] were implemented to generate mosaic products (Figure 3). Integration was implemented in mountain areas where the geometric correction of ALOS imagery was not effective due the presence of steep topography (Figure 6). ALOS-1 PALSAR-1 (2010) and ALOS-2 PALSAR-2 (2016) forest classifications products were explored visually, and misclassified forest/nonforest classes of steep slopes were reclassified to the correct class. This procedure was performed mainly in Boyacá-Santander, in the Coffee ecoregion, and partially in Meta foothills and Cesar river valley. Four Landsat VNIR spectral bands were used, avoiding bands in the visible blue and green intervals due to their sensitivity to atmospheric effects, especially in mountainous areas [43]. Forest change products (2010–2016) were then derived using the consolidated products, and zero deforestation assessment carried out at farm level.

The use of difference techniques using polarization values (especially HV) between image pairs can also lead to robust and cost-effective change detection results [28,44–46]. This research, however, needed thematic forest/nonforest extent products to estimate the carbon balance of the project (tons of equivalent CO2 at the baseline versus the assessment year). Additionally, the thematic maps were also necessary for forest management interventions in the project, e.g., defining conservation areas using fence isolation to avoid cattle access to forest.

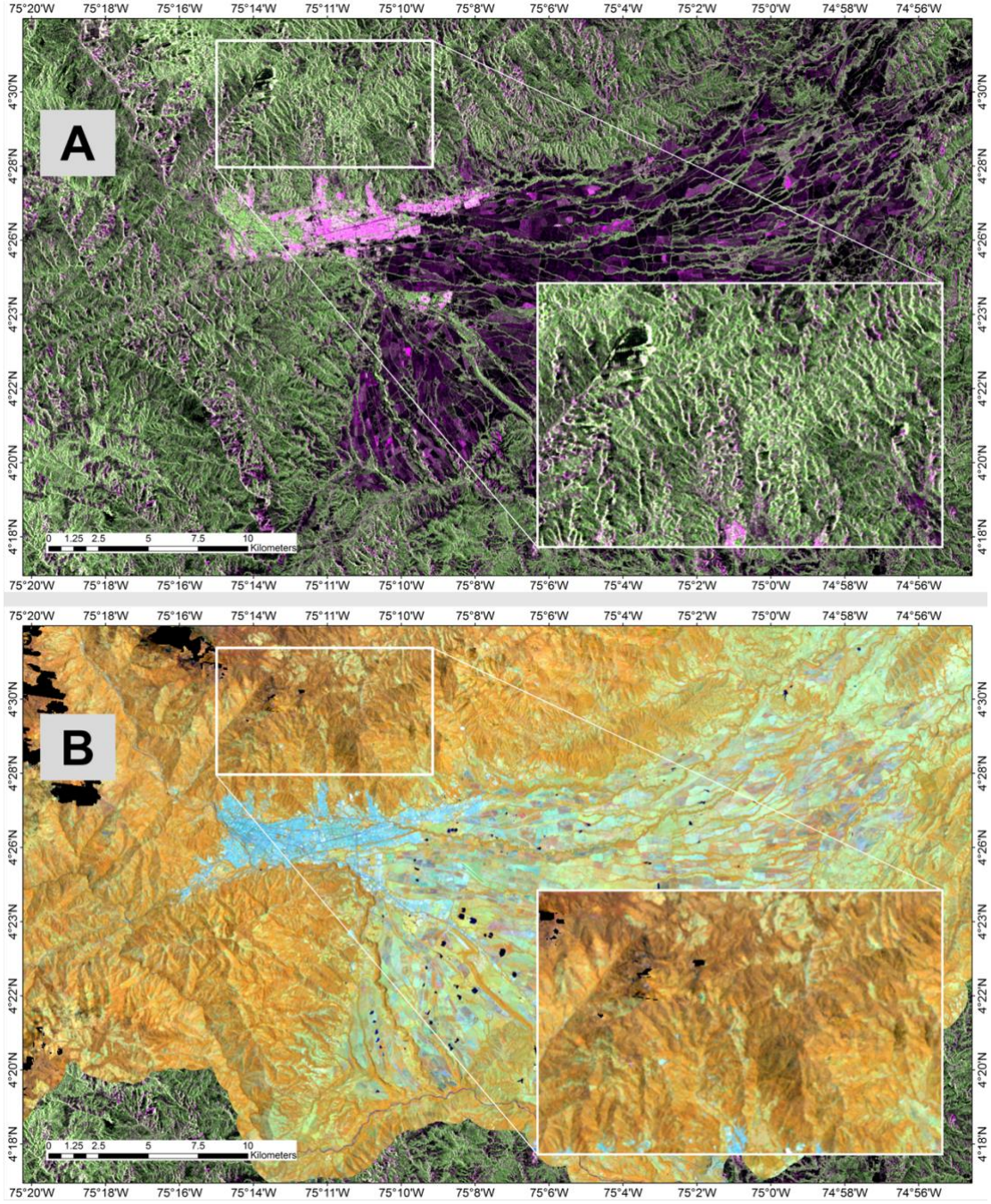


Figure 10. 2010 imagery (**A**) ALOS-1 fine beam imagery (FCC: Red: HH, Green: HV, and Blue: |HH-HV| polarizations). Steep slopes affect ALOS backscatter, where signal is intensified by terrain aspect (brighter pixel on mountain areas) or is lost by terrain aspect (black areas in mountain areas) in reference of the beam angle; (B) Landsat 8 OLI (FCC: Red: NIR, Green: SWIR, and Blue: red). Geographical Coordinate System WGS 84.

A quantitative accuracy assessment of stable forest, stable nonforest and forest change thematic products were carried out based on Olofsson et al. validation approach [47,48], to derive accuracy

indices at each of the five regions. The application of the validation at farm level was discarded due to (i) the impossibility to reach the minimum validation sample size per farm (n = 50 [47]), and (ii) the presence of highly concentrated points within small farms, thus not ensuring the sample independence [47]. The approach produced thematic accuracy indices and adjusted area estimations for each thematic class reported, with estimates of lower and upper limits (confidence intervals) at regional scale. The mapped areas were adjusted to take into account bias attributable to omission and commission classification errors [47]. The total validation sample size (n) per region was determined setting the expected standard error Si = 0.01 and validation accuracy Ui = 0.9 for each class, as suggested in this approach. In the case of a region rare class stratum, the minimum sample size was set at n = 50 following the reported validation approach [47]. The validation points were spatially randomly distributed within each region. The validation process was independent from the classification process: three project researchers visually interpreted the class of each validation point using ALOS PALSAR, Landsat, and Google Earth® imagery available for 2010 and 2016, and had no information on the previously classified maps. When at least two researchers coincided on the class assigned to the point, that point was taken as the validation reference, while when there was no agreement on the class assigned the three researchers defined the class by discussion and consensus. Results derived from the classification and validation phases were integrated to produce a confusion matrix and to derive the accuracy indices.

3. Results

3.1. Forest Extents and Change

Based on the adjusted areas estimations, Bajo Magdalena stable forest (2010–2016) represents 1.21% of the total area of the region with 3086 ha, deforestation 0.15% (373 ha), and nonforest 98.64% (251,032 ha) (Table 2). For the Cesar River Valley region stable forest represents 15.17% of the total area of the region (144,737 ha), deforestation 0.41% (3899 ha), and nonforest 84.42% (805,342 ha). In Boyacá-Santander the estimations showed that forest represent 20.78% of the total extension of the region with 94,499 ha, deforestation 0.13% with 592 ha, and nonforest 79.09% with 359,594 ha. Regarding to the Coffee Ecoregion, stable forests represent 25.93% of the total area of the region with 224,534 ha, deforestation 0.02% with 210 ha, and nonforest 7440% with 641,093 ha. For the Meta region Foothills stable forest represents 16.13% of the total extension of the region with 99,275 ha, deforestation 0.06% with 377 ha, and nonforest 83.81% with 515,879 ha (Table 2).

Table 2. Stable forest, stable nonforest, and deforestation area estimations for Bajo Magdalena, Cesar River Valley, Boyacá-Santander, Coffee Ecoregion, and Meta Foothills regions.

	Class	Mapped Area (ha)	Adjusted Area (ha)	Margin of Error (ha; 95% Confidence Interval)
Bajo Magdalena	F	4528	3086	302
	DF	336	373	92
	NF	249,627	251,032	297
Cesar River Valley	F	143,618	144,737	12,433
	DF	1999	3899	2663
	NF	808,361	805,342	12,266
Boyacá-Santander	F	94,469	94,499	4074
	DF	800	592	50
	NF	359,417	359,594	4075
Coffee Ecoregion	F	231,785	224,534	13,761
	DF	250	210	13
	NF	633,802	641,093	13,761
Meta Foothills	F	113,404	99,275	6605
	DF	484	377	29
	NF	501,645	515,879	6605

Considering the total MSCR farms area (2615 farms), we estimated 17,162.1 ha of forest extent for year 2010 and 17,066.6 ha for 2016 (Table 3). The Meta foothills is the region had the largest forest cover of all farms with 6057.6 ha (2016), followed by the Coffee ecoregion with a total forest cover of 4931.8 ha and the Cesar river valley 4430.8 ha. The total forest area of these three regions represents 90% of the total forest cover associated with the farms of the MSCR project. Forest cover was estimated at 80.6 ha for Bajo Magdalena and 1565.8 ha for Boyacá-Santander. Bajo Magdalena has less than one percent of the total forest cover. Forest cover proportion inside the MSCR farms are low for all the regions: Bajo Magdalena and Cesar Valley river showed the lowest proportions with 1.1% and 9.5%, respectively, followed by the Coffee ecoregion (16.7%), Boyacá-Santander (23.7%), and Meta foothills (25.5%).

Forest change analysis indicated that the overall estimate of 2010–2016 deforestation, considering all MSCR farms, is approximately 95 ha, 0.5% percent of the total forest cover within farms (Table 3). Meta foothills farms showed zero deforestation hectares during the period of the assessment, followed by Bajo Magdalena (1.6 ha), Boyacá-Santander (3.7 ha), and Coffee ecoregion (15.2 ha). Cesar river valley showed the largest estimated deforestation extent (75 ha).

Table 3. Forest, nonforest and deforestation area estimations at farm level for 2615 farms associated to the MSCR project (2010–2016).

Region	Stable Forest (ha)	Deforestation (ha)	Stable No Forest (ha)
Bajo Magdalena	80.6	1.6	5727
Cesar river Valley	4430.8	75	34,118.6
Boyacá-Santander	1565.8	3.7	3655.8
Coffee ecoregion	4931.8	15.2	21,834
Meta foothills	6057.6	0	17,677.9

3.2. Forest Mapping Accuracy

Accuracy assessment was performed for the 2010–2016 forest change map. Thematic overall accuracy estimated for each study region varied from 92% to 99%, with a 90% confidence level. Cesar river valley (93%) and Coffee ecoregion (92%) presented the lowest overall thematic accuracy, followed by Meta foothills (95%), Boyacá-Santander (97%), and Bajo Magdalena (95%) (Table 4). User accuracy (UA) and producer accuracy (PA) values related to the stable forest, stable nonforest, and deforestation classes varied significantly among regions, i.e., between 62% and 100% (UA), and between 68% and 1% (PA) (Table 4).

The lowest UA values associated with forest class were associated to the Bajo Magdalena, Cesar River Valley, and Coffee Ecoregions, with 68%, 78%, and 80%, respectively; higher values were calculated for Boyacá-Santander and Meta foothills (93%). Stable forest PA showed generally high values (≥ 82%), with the exception of Cesar River Valley (77%). Stable nonforest classes showed high values for UA and PA consistently for all five regions (≥ 93%). Related to the deforestation class, Cesar River Valley showed the lowest UA and PA values with 32% and 68%, respectively. The Meta Foothills region consistently showed the highest values for both UA and PA (100%). Other regions showed intermediate user and producer accuracy values (Table 4).

Table 4. Thematic accuracy indices and number of validation points per region assigned to stable forest (F), deforestation (DF) and stable nonforest (NF) classes. Accuracy measures are calculated with 90% confidence interval, following the method of a previous paper [47].

Region	Class	F	DF	NF	Total	W_i	User Accuracy	Producer Accuracy	Overall Accuracy
Bajo Magdalena	F	34	1	15	50	0.0178	0.68	1	0.99
	DF	1	42	7	50	0.0013	0.84	0.76	
	NF	0	0	296	296	0.9809	1	0.99	
Total		35	43	318	396				
Boyacá-Santander	F	67	0	5	72	0.208	0.93	0.93	0.97
	DF	2	37	11	50	0.002	0.74	1	

	NF	5	0	269	274	0.79	0.98	0.98	
Total		74	37	285	396				
Meta foothills	F	52	0	12	64	0.184	0.93	0.93	0.95
	DF	2	39	9	50	0.001	1	1	
	NF	4	0	278	282	0.815	0.96	0.96	
Total		58	39	299	396				
Coffee ecoregion	F	74	0	19	93	0.268	0.8	0.82	0.92
	DF	4	42	4	50	0.000	0.84	1	
	NF	16	0	237	253	0.732	0.94	0.93	
Total		94	42	260	396				
Cesar river valley	F	42	1	11	54	0.151	0.78	0.77	0.93
	DF	1	31	18	50	0.002	0.32	0.68	
	NF	12	0	282	294	0.847	0.96	0.96	
Total		55	32	311	398				

4. Discussion

Forest and forest change products generated using ALOS PALSAR showed satisfactory overall accuracy (OA) values for all the regions analyzed (Table 4), presenting no significant differences among regions. High overall accuracy can be partially explained by the high accuracies obtained in the nonforest class, which is the class with largest area proportion (82%) compared to the forest and deforestation classes. This is a well-known limitation of this accuracy parameter.

Regions with highly complex topography (Boyacá-Santander and partially Cesar River Valley) presented lower accuracy levels for the forest and deforestation classes, respectively. Our results suggest that local slope orientation present in mountains with respect to the incidence angle of SAR sensors had a relevant effect on image distortion [49] or hampered the surface visibility by the sensor [50,51]. Here the integration with optical sensor information was necessary to reduce the misclassified areas. Higher accuracy performance was obtained in the Meta foothills region, which is characterized mostly by flat areas compared with mountain regions (i.e., Boyacá-Santander and Coffee Ecoregion). The Cesar river Valley region showed lower classification accuracy values, especially user's accuracy, indicating an overestimation in the deforestation extension.

Deforestation products were characterized by misclassification errors especially within regions with dry conditions (Cesar River Valley and Bajo Magdalena), compared to the humid regions (Coffee ecoregion, Boyacá-Santander, and Meta foothills). During forest classification, detection of dry forests was more challenging, due to both (i) the high variation in vegetation structure observed in the ALOS PALSAR Fine Beam Dual imagery, and (ii) the seasonal deciduous behavior (phenology), well-observed using Landsat optical imagery. Our results suggest that dry forests mapping could need comprehensive ground truth data surveys to integrate the remote sensing-based mapping workflow. This is especially necessary for Colombia, where limited detailed and spatially explicit information is available for this forest type [30].

The low proportion of forests in the project's farms reflects historical processes of deforestation, prior to the implementation of the MSCR project. The study regions characterized by greater proportion of forest cover within farms are located in mountainous or partially mountainous areas, possibly due to the limited accessibility [52], while those with less forest cover proportion are associated with dry regions which are historically characterized by high agricultural pressure [30].

In the Meta foothills region, zero hectares of deforestation detected between 2010 and 2016 (Table 3) advocates the fulfillment of zero deforestation. For the flat Bajo Magdalena region, deforestation estimate is 1.6 ha and with low variation due to uncertainties (Table 2), while in the mountain region of Boyacá-Santander, the deforestation estimated area (3.7 ha) is overestimated based on relatively low UA accuracy results. This was somehow expected due to the presence of steep slopes characteristic of the region, coupled with the lower efficiency of radar sensing in these topographic conditions [52].

In the case of the Cesar River Valley region, where major deforestation areas were detected and the lowest UA calculated, a postvalidation visual assessment was performed, revealing that the

largest deforested area corresponds to a single event of approximately 33 ha; for this single event the area of the corresponding farm was estimated using the circular buffer method, so the direct attribution of noncompliance had an additional uncertainty due to the simplified geometric protocol used.

Nevertheless, other multi-temporal postclassification studies reported change classes as showing generally higher commission errors compared to errors associated to stable classes (forest/nonforest) [48,53–56]. Multiple research has discussed general minimum accuracy standards for remote sensing-based thematic mapping [57], although there is no global consensus for thematic accuracy of deforestation products. We stress that projects associated with zero deforestation agreements monitoring and assessment should always report omission and commission accuracy indices of each thematic class coupled with error-adjusted areas and their confidence intervals. Minimum accuracies should be included in the project specific requirements, together with additional procedures to corroborate agreements compliance/noncompliance, e.g., field corroboration of change detection procedures based on remote sensing analysis.

The application of ALOS PALSAR FBD imagery was found to provide significant and consistent information for the detection of forest and nonforest cover; the results were especially relevant for highly clouded tropical conditions [4]. Nonetheless, we found that its application in mountainous areas presents some limitations, since the signal of ALOS PALSAR resulted rather affected by the topography characteristics of some regions. The integration of highly detailed digital elevation models, dense temporal image series and improved preprocessing routines should be used to generate more accurate products for forest, nonforest, and deforestation detection and quantification in these specific conditions. The integration of optical sensors information improved the detection of forests and deforestation in topographic areas characterized by steep slopes, however, at the cost of the time of processing. Operational integration of Synthetic Aperture Radar imagery with optical imagery can significantly improve the consistency and robustness of forest monitoring in the tropics [58], to achieve efficient forest monitoring procedures for interoperability and classifications needs.

5. Conclusions

The Kyoto and Carbon Initiative has contributed significantly to the development and application of methodologies for forest monitoring in Colombia based on ALOS PALSAR products, allowing us to investigate their application for forest monitoring and deforestation at the local scale; complementing national efforts to quantify forest extent and forest change.

The present work highlighted some of the operational issues to be considered in the implementation and replicate of SAR-based systems for forest monitoring. These include the influence of topography features in SAR backscatters and the effect of structural and phenological characteristics of tropical dry forests. Our findings suggest that additional research should focus on more efficient geometric correction procedures to reduce errors associated with topography features. Future improvements to the current approach can include the use of radar dense time series to improve detection change analysis, together with ground truth data for dry forests to calibrate ALOS PALSAR sensors and reduce classification uncertainties. Large scale and medium resolution advances have been implemented on the generation of ALOS PALSAR products, i.e., JAXAS's worldwide PALSAR mosaics [59] for forest monitoring, however additional research needs to be done on how to generate robust, consistent, and high accuracy forest products at local scales. ALOS PALSAR and SAR sensors provide multiple operation modes and processing approaches that need to be further explored to understand their real potential contribution to local scale forest monitoring.

Supplementary Materials: Table S1: ALOS-1 and ALOS-2 Fine Beam Dual imagery used for forest change mapping; Figure S1: ALOS-1 and ALOS-2 Fine Beam Dual imagery preprocessing workflow implemented in Gamma ® software.

Author Contributions: C.P., N.C., C.F., and G.G. conceived and designed the experiments; C.P. and C.F. performed the experiments; C.P. and C.F. analyzed the data; A.M. contributed with ground truth data; C.P., N.C., A.M., D.N., D.L., A.F.Z., J.D., C.F., and G.G wrote the paper.

Acknowledgments: The World Bank provided the funding for this study. The Global Environmental Facility (GEF) and the United Kingdom Government funded the Mainstream Sustainable Cattle Ranching Project, the Foundation Center for Research in Sustainable Systems of Agricultural Production (CIPAV), the Colombian Cattle Ranching Federation (FEDEGAN), The Nature Conservancy NASCA program, the Environmental action Fund (Fondo para la Acción Ambiental), and the World Bank are the institutions part of the Sustainable Cattle Ranching Project partnership. ALOS PALSAR data was provided by JAXA through the Fourth phase of the Kyoto and Carbon Initiative SIPM. No. 0216001. We acknowledge Julian Gonzalo Jimenez, Stavros Papageorgiou, and Luz Berania Díaz for their comments and recommendations on the methods applied in this research. Claudia Huertas and Ecometrica staff: Richard Tipper, Ryan Elfman, Jil Bournazel, and Véronique Morel are acknowledged for their helpful technical advice. Thanks are also given to all the field team members and researchers from this project's partner institutions that provided the ground truth data.

Conflicts of Interest: The authors declare no conflicts of interest. The founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

6. References

1. Keenan, R.J.; Reams, G.A.; Achard, F.; de Freitas, J.V.; Grainger, A.; Lindquist, E. Dynamics of global forest area: Results from the FAO Global Forest Resources Assessment 2015. *For. Ecol. Manag.* **2015**, *352*, 9–20.
2. Etter, A.; McAlpine, C.; Possingham, H. Historical Patterns and Drivers of Landscape Change in Colombia Since 1500: A Regionalized Spatial Approach. *Ann. Assoc. Am. Geogr.* **2008**, *98*, 2–23.
3. Ministerio de Ambiente y Desarrollo Sostenible. *Quinto Informe Nacional de Biodiversidad de Colombia ante el Convenio de Diversidad Biológica*; Ministerio de Ambiente y Desarrollo Sostenible: Bogotá, Colombia, 2014; pp. 1–101.
4. Rosenqvist, A.; Shimada, M.; Chapman, B.; Freeman, A.; De Grandi, G.; Saatchi, S.; Rauste, Y. The Global Rain Forest Mapping project—A review. *Int. J. Remote Sens.* **2000**, *21*, 1375–1387.
5. Achard, F.; Eva, H.D.; Stibig, H.-J.; Mayaux, P.; Gallego, J.; Richards, T.; Malingreau, J.-P. Determination of Deforestation Rates of the World's Humid Tropical Forests. *Science* **2002**, *297*, 999–1002.
6. Hansen, M.C.; Stehman, S.V.; Potapov, P.V. Quantification of global gross forest cover loss. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 8650–8655.
7. Gong, P.; Wang, J.; Yu, L.; Zhao, Y.; Zhao, Y.; Liang, L.; Niu, Z.; Huang, X.; Fu, H.; Liu, S.; et al. Finer resolution observation and monitoring of global land cover: First mapping results with Landsat TM and ETM+ data. *Int. J. Remote Sens.* **2013**, *34*, 2607–2654.
8. Hansen, M.C.; Potapov, P.V.; Moore, R.; Hancher, M.; Turubanova, S.A.; Tyukavina, A.; Thau, D.; Stehman, S.V.; Goetz, S.J.; Loveland, T.R.; et al. High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science* **2013**, *342*, 850–853.
9. Shimada, M.; Itoh, T.; Motooka, T.; Watanabe, M.; Shiraishi, T.; Thapa, R.; Lucas, R. New global forest/non-forest maps from ALOS PALSAR data (2007–2010). *Remote Sens. Environ.* **2014**, *155*, 13–31.
10. Kim, D.-H.; Sexton, J.O.; Townshend, J.R. Accelerated deforestation in the humid tropics from the 1990s to the 2000s. *Geophys. Res. Lett.* **2015**, *42*, 3495–3501.
11. Corbera, E.; Schroeder, H. Governing and implementing REDD+. *Environ. Sci. Policy* **2011**, *14*, 89–99.
12. Edenhoffer, O.; Pichs-Madruga, R.; Sokona, Y.; Minx, J.C.; Farahani, E.; Kadner, S.; Seyboth, K.; Adler, A.; Baum, I.; Brunner, S.; et al. *Climate Change 2014 Mitigation of Climate Change*; Cambridge University Press: Cambridge, UK, 2014; pp. 1–1454.
13. Visseren-Hamakers, I.J.; McDermott, C.; Vijge, M.J.; Cashore, B. Trade-offs, co-benefits and safeguards: Current debates on the breadth of REDD. *Curr. Opin. Environ. Sustain.* **2012**, *4*, 646–653.
14. UNFCCC Report. In Proceedings of the Conference of the Parties on Its Nineteenth Session, Warsaw, Poland, 11–23 November 2013.

15. Ochieng, R.M.; Visseren-Hamakers, I.J.; Arts, B.; Brockhaus, M.; Herold, M. Institutional effectiveness of REDD+ MRV: Countries progress in implementing technical guidelines and good governance requirements. *Environ. Sci. Policy* **2016**, *61*, 42–52.
16. Rosenqvist, A. Earth Observation Support to the UN Framework Convention on Climate Change: The Example of REDD. In *Satellite Earth Observations and Their Impact on Society and Policy*; Onoda, M., Young, O.R., Eds.; Springer: Singapore, 2017; Volume 352, pp. 143–153.
17. Rosenqvist, A.; Shimada, M.; Chapman, B.; McDonald, K.; De Grandi, G.; Jonsson, H.; Williams, C.; Rauste, Y.; Nilsson, M.; Sango, D.; et al. An Overview of the JERS-1 SAR Global Boreal Forest Mapping (GBFM) Project. In proceedings of the 2004 IEEE International Geoscience and Remote Sensing Symposium, Anchorage, AK, USA, 20-24 September 2004; Volume 2, pp. 1033–1036.
18. Rosenqvist, A.; Shimada, M.; Ito, N.; Watanabe, M. ALOS PALSAR: A Pathfinder Mission for Global-Scale Monitoring of the Environment. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 3307–3316.
19. Reiche, J.; Lucas, R.; Mitchell, A.L.; Verbesselt, J.; Hoekman, D.H.; Haarpaintner, J.; Kellndorfer, J.M.; Rosenqvist, A.; Lehmann, E.A.; Woodcock, C.E.; et al. Combining satellite data for better tropical forest monitoring. *Nat. Clim. Chang.* **2016**, *6*, 120–122.
20. Ryan, C.M.; Hill, T.; Woollen, E.; Ghee, C.; Mitchard, E.; Cassells, G.; Grace, J.; Woodhouse, I.H.; Williams, M. Quantifying small-scale deforestation and forest degradation in African woodlands using radar imagery. *Glob. Chang. Biol.* **2011**, *18*, 243–257.
21. Reiche, J.; Hamunyela, E.; Verbesselt, J.; Hoekman, D.; Herold, M. Improving near-real time deforestation monitoring in tropical dry forests by combining dense Sentinel-1 time series with Landsat and ALOS-2 PALSAR-2. *Remote Sens. Environ.* **2018**, *204*, 147–161.
22. Cartus, O.; Santoro, M.; Kellndorfer, J. Mapping forest aboveground biomass in the Northeastern United States with ALOS PALSAR dual-polarization L-band. *Remote Sens. Environ.* **2010**, *124*, 1–13.
23. Lucas, R.; Rebelo, L.-M.; Fatoyinbo, L.; Rosenqvist, A.; Itoh, T.; Shimada, M.; Simard, M.; Souza-Filho, P.W.; Thomas, N.; Trettin, C.; et al. Contribution of L-band SAR to systematic global mangrove monitoring. *Mar. Freshw. Res.* **2014**, *65*, 589–616.
24. Shimada, M.; Isoguchi, O. JERS-1 SAR mosaics of Southeast Asia using calibrated path images. *Int. J. Remote Sens.* **2002**, *23*, 1507–1526.
25. Kalamandeen, M.; Gloor, E.; Mitchard, E.; Quincey, D.; Ziv, G.; Spracklen, D.; Spracklen, B.; Adami, M.; Aragão, L.E.; Galbraith, D. Pervasive Rise of Small-scale Deforestation in Amazonia. *Sci. Rep.* **2018**, *8*, 1600.
26. Godar, J.; Gardner, T.A.; Tizado, E.J.; Pacheco, P. Actor-specific contributions to the deforestation slowdown in the Brazilian Amazon. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 15591–15596.
27. Gibbs, H.K.; Munger, J.; L'Roe, J.; Barreto, P.; Pereira, R.; Christie, M.; Amaral, T.; Walker, N.F. Did Ranchers and Slaughterhouses Respond to Zero-Deforestation Agreements in the Brazilian Amazon? *Conserv. Lett.* **2015**, *9*, 32–42.
28. Almeida-Filho, R.; Shimabukuro, Y.E.; Rosenqvist, A.; Sanchez, G.A. Using dual-polarized ALOS PALSAR data for detecting new fronts of deforestation in the Brazilian Amazonia. *Int. J. Remote Sens.* **2009**, *30*, 3735–3743.
29. Etter, A.; Mcalpine, C.; Wilson, K.; Phinn, S.; Possingham, H. Regional patterns of agricultural land use and deforestation in Colombia. *Agric. Ecosyst. Environ.* **2006**, *114*, 369–386.

Pizano, C.; García, H. *El Bosque seco Tropical en Colombia*; UBC Press: Vancouver, Canada, 2014.

30. Armenteras, D.; Gast, F.; Villareal, H. Andean forest fragmentation and the representativeness of protected natural areas in the eastern Andes, Colombia. *Biol. Conserv.* **2003**, *113*, 245–256.
31. Galvis, G.; Iván Mojica, J. The Magdalena River fresh water fishes and fisheries. *Aquat. Ecosyst. Health Manag.* **2007**, *10*, 127–139.
32. JAXA Calibration Result of ALOS-2/PALSAR-2 JAXA Standard Products. Available online: http://www.eorc.jaxa.jp/ALOS-2/en/calval/calval_index.htm (accessed on 15 June 2018).

Werner, C.; Wegmller, U.; Strozzi, T.; Wiesmann, A. Gamma SAR and interferometric processing software. In proceedings of the ERS-ENVISAT Symposium, Gothenburg, Sweden, 16–20 October 2000.

33. Lopes, A.; Touzi, R.; Nezry, E. Adaptive speckle filters and scene heterogeneity. *IEEE Trans. Geosci. Remote Sens.* **1990**, *28*, 992–1000.

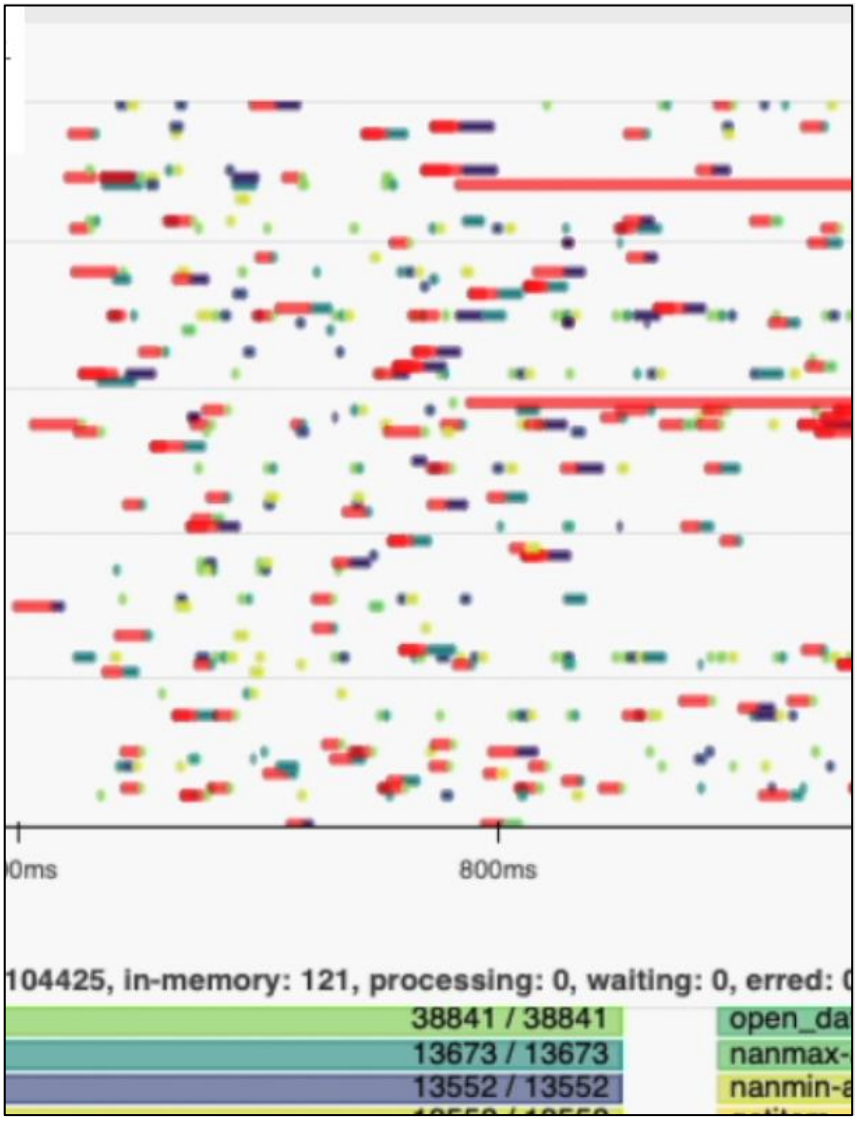
Tou, J.T.; Gonzalez, R.C. *Pattern Recognition Principles*; Addison-Wesley: Reading, Mass, USA, 1974.

34. Richards, J.A. *Remote Sensing Digital Image Analysis*; Springer: Berlin/Heidelberg, Germany, 1999.
35. Zhu, Z.; Woodcock, C.E. Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sens. Environ.* **2012**, *118*, 83–94.
36. Vermote, E.F.; El Saleous, N.; Justice, C.O.; Kaufman, Y.J.; Privette, J.L.; Remer, L.; Roger, J.C.; Tanre, D. Atmospheric correction of visible to middle-infrared EOS-MODIS data over land surfaces: Background, operational algorithm and validation. *J. Geophys. Res. Atmos.* **1997**, *102*, 17131–17141.
37. Schmidt, G.; Jenkerson, C.; Masek, J.G. *Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) Algorithm Description*; U.S. Geological Survey: Reston, VA, USA, 2013.

Masek, J.G.; Vermote, E.F.; Saleous, N.; Wolfe, R.; Hall, F.G.; Huemmrich, K.F.; Gao, F.; Kutler, J.; Lim, T.K. *LEDAPS Calibration, Reflectance, Atmospheric Correction Preprocessing Code, Version 2*; ORNL DAAC: Oak Ridge, TN, USA, 2013.

38. Zhu, Z.; Woodcock, C.E. Automated cloud, cloud shadow, and snow detection in multitemporal Landsat data: An algorithm designed specifically for monitoring land cover change. *Remote Sens. Environ.* **2014**, *152*, 217–234.
39. Ouaidrari, H.; Vermote, E.F. Operational Atmospheric Correction of Landsat TM Data. *Remote Sens. Environ.* **1999**, *70*, 4–15.
40. Motohka, T.; Shimada, M.; Uryu, Y.; Setiabudi, B. Using time series PALSAR gamma nought mosaics for automatic detection of tropical deforestation: A test study in Riau, Indonesia. *Remote Sens. Environ.* **2014**, *155*, 79–88.
41. Reiche, J.; Verbesselt, J.; Hoekman, D.; Herold, M. Fusing Landsat and SAR time series to detect deforestation in the tropics. *Remote Sens. Environ.* **2015**, *156*, 276–293.
42. Whittle, M.; Quegan, S.; Uryu, Y.; Stüewe, M.; Yulianto, K. Detection of tropical deforestation using ALOS-PALSAR: A Sumatran case study. *Remote Sens. Environ.* **2012**, *124*, 83–98.
43. Olofsson, P.; Foody, G.M.; Stehman, S.V.; Woodcock, C.E. Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation. *Remote Sens. Environ.* **2013**, *129*, 122–131.
44. Olofsson, P.; Foody, G.M.; Herold, M.; Stehman, S.V.; Woodcock, C.E.; Wulder, M.A. Good practices for estimating area and assessing accuracy of land change. *Remote Sens. Environ.* **2014**, *148*, 42–57.
45. Gelautz, M.; Frick, H.; Raggam, J.; Burgstaller, J.; Leberl, F. SAR image simulation and analysis of alpine terrain. *ISPRS J. Photogramm. Remote Sens.* **1998**, *53*, 17–38.
46. Herrera, G.; Gutiérrez, F.; García-Davalillo, J.C.; Guerrero, J.; Notti, D.; Galve, J.P.; Fernández-Merodo, J.A.; Cooksley, G. Multi-sensor advanced DInSAR monitoring of very slow landslides: The Tena Valley case study (Central Spanish Pyrenees). *Remote Sens. Environ.* **2013**, *128*, 31–43.
47. Cigna, F.; Bateson, L.B.; Jordan, C.J.; Dashwood, C. Simulating SAR geometric distortions and predicting Persistent Scatterer densities for ERS-1/2 and ENVISAT C-band SAR and InSAR applications: Nationwide feasibility assessment to monitor the landmass of Great Britain with SAR imagery. *Remote Sens. Environ.* **2014**, *152*, 441–466.
48. Rubiano, K.; Clerici, N.; Norden, N.; Etter, A. Secondary Forest and Shrubland Dynamics in a Highly Transformed Landscape in the Northern Andes of Colombia (1985–2015). *Forests* **2017**, *8*, 216–217.
49. Jeon, S.B.; Olofsson, P.; Woodcock, C.E. Land use change in New England: A reversal of the forest transition. *J. Land Use Sci.* **2012**, *9*, 105–130.
50. Olofsson, P.; Torchinava, P.; Woodcock, C.E.; Baccini, A.; Houghton, R.A.; Ozdogan, M.; Zhao, F.; Yang, X. Implications of land use change on the national terrestrial carbon budget of Georgia. *Carbon Balance Manag.* **2010**, *5*, 4.
51. Olofsson, P.; Kuemmerle, T.; Griffiths, P.; Knorn, J.; Baccini, A.; Gancz, V.; Blujdea, V.; Houghton, R.A.; Abrudan, I.V.; Woodcock, C.E. Carbon implications of forest restitution in post-socialist Romania. *Environ. Res. Lett.* **2011**, *6*, 045202.
52. Seto, K.C.; Woodcock, C.E.; Song, C.; Huang, X.; Lu, J.; Kaufmann, R.K. Monitoring land-use change in the Pearl River Delta using Landsat TM. *Int. J. Remote Sens.* **2002**, *23*, 1985–2004.
53. Foody, G.M. Status of land cover classification accuracy assessment. *Remote Sens. Environ.* **2002**, *80*, 185–201.

54. Lehmann, E.A.; Caccetta, P.; Lowell, K.; Mitchell, A.; Zhou, Z.-S.; Held, A.; Milne, T.; Tapley, I. SAR and optical remote sensing: Assessment of complementarity and interoperability in the context of a large-scale operational forest monitoring system. *Remote Sens. Environ.* **2015**, *156*, 335–348.
55. Shimada, M.; Ohtaki, T. Generating Large-Scale High-Quality SAR Mosaic Datasets: Application to PALSAR Data for Global Monitoring. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2010**, *3*, 637–656.



Chapter 3: Consistent and robust deforestation monitoring: A cloud-based computing workflow to access, structure, and analyze Earth Observations big data.

This chapter was designed to integrate the latest cloud-based technology for EO data access, organization, analysis, and delivery of open source and cloud-based solutions. Two cloud-based solutions were integrated: (1) the Nebari platform: open-source routines where developed and tested to process EO to detect deforestation events in one of the Colombian hotspots (Caquetá, department); (2) with the Software for Earth Big Data Processing Prediction, and Organization (SEPPO), radar EO pre-processing and deforestation change detection routines were scaled on a cloud-based environment to efficiently process big data and generate value-added products in record time.

This chapter will be published as a scientific paper during the second semester of 2023. The paper is already reviewed by the multiple authors, and the next step is to define between three potential journals to publish this research:

- **Computers & Geosciences Journal**, ScienceDirect:
    - Special Section on Highlighting Earth and Planetary Sciences research by authors based in the Global South, **Special Issue.**
- **Journal of Cloud Computing**, Springer Open
- **Big Data and Cognitive Computing**, MDPI.

This research is a collaboration between this PhD project and Earth Big Data LLC. This research started the year 2021 where Carlos Pedraza worked during his internship with Earth Big Data LLC, when Josef Kellndorfer Ph.D introduced Carlos to cloud-computing technologies developed by EBD, the SEPPO software, and supported by Amazon Web Services. Followed by the introduction to Jupyter Hub and Nebari platform through the Earth Science Information Partners (ESIP). ESIP, a collaborative platform to establish and continuously improve science-based end-to-end processes that increase the quality and value of Earth science products and services.

Based on the knowledge acquired through this internship and the access to advanced cloud-based technologies; Carlos, designed, configured, and executed technologies and routines to access and analyze to Sentinel-1 time series data to generate deforestation change detections for one of hotspots of deforestation in Colombia.

The present research involved the following collaborations:

Earth Big Data LLC (EBD) and Earth Science Information Partners (ESIP). Through the collaboration with EBD and ESIP Carlos had his first introduction to cloud computing technologies working with multiple researchers: Josef Kellndorfer and Marcelo Villa from EBD, and Richard Signell from United States Geological Survey (USGS). Based on knowledge obtained during these collaborations Carlos configured multiple cloud-based technologies for deforestation change detection.

A cloud-based workflow to develop and implement solutions for forest monitoring using Earth Observation big data.

Carlos Pedraza[1,2,] *, Nicola Clerici[2], Marcelo Villa[1], Josef Kellndrofer[1]

[1] Earth Big Data LLC, Woods Hole, Massachusetts, USA

[2] Department of Biology, Faculty of Natural Sciences, Universidad del Rosario, Kr 26 # 63B-48, Bogotá, D.C., Colombia

*** Correspondence:** carlos@earthbigdata.com

Keywords: deforestation detection, cloud computing, Sentinel-1, AWS, Time series, Earth Observations, Auto-scaling computing.

1. Introduction

In recent years, the world has seen an exponential growth in the generation of data across multiple fields of research. This unprecedented surge in data production has been fueled by advancements in technology and the development of sophisticated data collection methods. For example, in the field of Earth Observation (EO), massive amounts of remotely sensed data are being continuously acquired. In 2017, at least 277 EO satellites were in orbit and NASA's Earth Observing System Data and Information System (EOSDIS) data archive reached 23.8 petabytes with an average growth rate of 15.3 terabytes per day (Wang et al., 2020). In 2021, the archive held more than 59 petabytes of data and is expected to expand to around 250 petabytes by 2025 (Earth Science Data Systems, 2019). Furthermore, with the upcoming launch of the joint NASA-Indian Space Research Organization Synthetic Aperture Radar (NISAR) mission in 2024, the overall daily ingest of data in the archive is estimated to grow up to 114 terabytes (Josh Blumenfeld, 2017).

While the surge of available EO data provides a great opportunity to enhance our knowledge of Earth systems and phenomena, it also entails big challenges for scientists and technology information experts. The sheer volume, velocity and variety of EO data demands innovative approaches to process, analyze and visualize this data (Nativi et al., 2015). Traditional computing infrastructures can be bound by hardware and their provisioning, maintenance and upgrading can often prove difficult to achieve (Fisher, 2018). Furthermore, small organizations or companies with occasional high demands of computational resources can be limited by the financial burdens of acquiring large amounts of hardware for sporadic processing purposes.

Cloud-based systems are a cost-effective alternative that offer flexible and scalable computational and storage resources to a wide range of users. By leveraging on-demand access to infrastructure and services, researchers can develop complex EO applications that rely on the processing of extensive and diverse datasets (Yao et al., 2019). Many space agencies and companies are hosting their imagery on cloud object storage through public cloud providers (e.g., Amazon, Google and Microsoft), enabling better user accessibility and discovery as well as interoperability with cloud-based solutions (Abernathey et al., 2021). Users can now deploy infrastructure in the same or in a geographically close region where the data is stored, resulting in faster data transfer and a high-throughput processing capacity. In addition, as new data paradigms and formats are being developed and established, imagery data access and analysis is becoming more efficient and straightforward (Giuliani et al., 2019; Xu, 2022).

SEPPO (which stands for Software for Earth big data Processing, Prediction modeling, and Organization) is an in-house cloud-based system to address the growing demand of computational resources in EO applications. Built on top of Amazon Web Services (AWS), it aims to simplify the processing of huge amounts of data. Although SEPPO was initially designed with the idea of allowing the processing of several satellite images in parallel, it can be used to run batches of any other user-defined computational job.

2. Software for Earth big data Processing, Prediction modeling, and Organization (SEPPO) - Overview

SEPPO is based on five different AWS products: (i) Elastic Compute Cloud (EC2), (ii) Simple Queue Service (SQS), (iii) Relational Database Service (RDS), (iv) Simple Storage Service (S3), and (v) Elastic File System (EFS). These products interact to provide the ability to stage and process groups of computational jobs in the AWS Cloud Figure 11. The general workflow consists in users creating queues for each group of processing jobs that are then executed in parallel by a predefined number of cloud instances with specific hardware and software configurations. Jobs are language-independent programs that usually (but are not required to) have input and output parameters that can range from cloud storage objects (e.g., S3) to files sitting on HTTP or FTP servers.

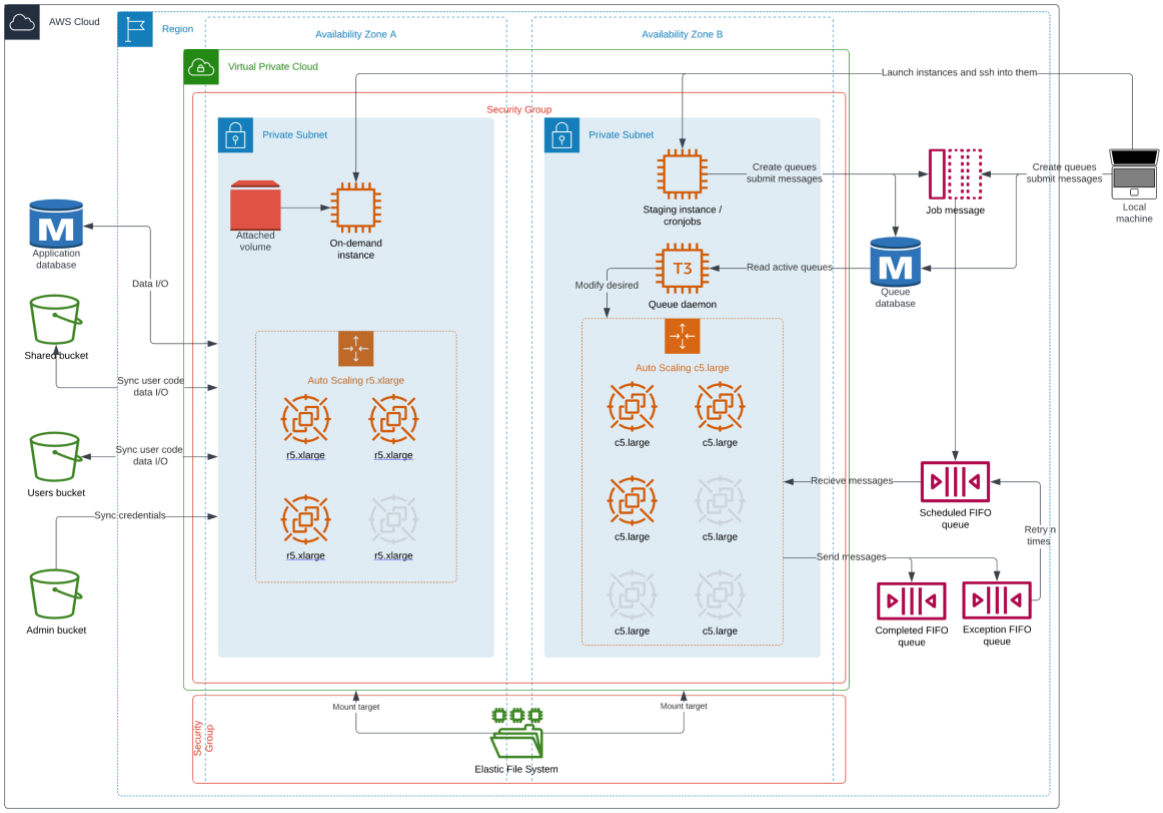


Figure 11. SEPPO general architecture workflow describing the main components of SEPPO software in an AWS cloud infrastructure, to access, storage, and process to be used in multiple EO applications. All queues generated in a local machine are stored as messages in a database located in AWS where jobs are executed as First-in-First-Out (zoter) by autoscaling groups with specific instance types and AMI depending on queue metadata.

The core of SEPPO lies within EC2, as it provides scalable computing capacity in the cloud. It allows users to define and create Auto Scaling groups that automatically launch virtual machines based on the user's demand to process the jobs submitted in the queues. Each of these groups has a specific instance type associated with it, enabling users to have multiple fleets of instances with different hardware specifications that fit diverse computational requirements. Additionally, each group is linked with an Amazon Machine Image (AMI), a snapshot of a configured and provisioned instance, that is used to launch other instances with the same software setup. This ensures that the instances running the jobs have the correct software requirements to do so. A daemon, denominated the queue processor, runs on every machine polling the queues for messages (i.e., jobs) to process sequentially. After it has processed the current message, it polls the queue for the next available

message. Once there are no messages left in the queue, it looks for other relevant queues matching the hardware and software specifications to poll from. If there are no queues or messages left to process, the instance terminates itself.

Jobs are stored as messages in First-In-First-Out queues managed by SQS. Each message contains a single command that is read and executed by an arbitrary instance from the corresponding Auto Scaling group. In addition, queue metadata is stored in a small database - the queue database - managed by RDS. Metadata is useful to define what instance type and AMI (i.e., hardware and software configuration) should be associated with the Auto Scaling group that will launch the instances to run the jobs, queue status (i.e., suspended, active, completed and waiting) and queue dependency. When a user submits a queue via SEPPO, a new entry with all the metadata is added to the queue database and three FIFO queues are created in SQS: a scheduled queue, a completed queue and an exception queue. The scheduled queue contains all the messages that the user submits as jobs and is the queue where the instances read from. Once an instance successfully finishes a job, it stores a message in the complete queue with relevant information about the duration of the processing. On the contrary, when an instance encounters an error while processing a job, it creates a message on the exception queue with the corresponding error log and returns the original message to the scheduled queue. This is done an arbitrary number of times to ensure that failed jobs are reprocessed in case there were external errors such as connection timeouts.

To control the logic behind users creating queues and submitting jobs and the Auto Scaling groups launching instances to process them, a permanent process (the queue daemon) - is running on a small instance (t3.nano with 2 vCPUs and 500 MB of RAM). This process runs as a scheduled task (cron job) every minute to check whether there are entries in the queue database with an active status. In case there are, it reads the required hardware and software specifications and searches for an Auto Scaling group with these characteristics. If it exists, the daemon modifies the desired capacity of the group (i.e. number of instances) to match the number of jobs to process. Auto Scaling groups can be previously configured to have a maximum capacity that depends on user needs but also on AWS service quotas. If the requested number of instances is higher than the maximum capacity of the group, the number will be capped. The requested (or capped) number of instances is then automatically launched to process the jobs. Just before an instance terminates itself, it reduces the desired capacity of the group by one to make sure no new instances are launched unnecessarily. Once all the jobs in one or several queues are processed, the desired capacity of the group becomes zero and there are no instances left. The queue daemon updates the status of the corresponding queue database entries to completed. Queues can have dependencies between each other, meaning that a certain queue might only be activated once another queue it is depending on is completed. The queue daemon handles the activation of those queues automatically by updating their waiting status to ‘active’.

While AMIs are configured to have the required software to run the jobs submitted by users, they do not necessarily have the user's code nor the necessary credentials to run it (e.g., database connections or HTTP server authentications). When SEPPO is deployed, three private buckets are created in S3 to store user data, code and credentials. Every time an instance from a given Auto Scaling group is launched, it copies specific objects from those buckets to the local file system. By the time the instance starts processing jobs, it already has access to the code, configuration, credentials and any other files defined by the user. SEPPO users are set up as local users in the instance and all relevant files are downloaded via S3 access on startup This way a user can replicate his local development or production environment across all the instances without the need of manually

provisioning them. These buckets (and any other bucket the user might have access to) also allow users to provide input data for their jobs and to store their processing outputs.

SEPPO was designed with a multi-user approach in mind to allow teams to work within the same AWS account and have multiple users leverage existing resources in it. Account administrators are allowed to create several users who are automatically attached a set of default opinionated permission policies for the AWS services that SEPPO uses. While these policies can be modified and new ones can be attached, they aim to follow security best practices and set least-privilege permissions for users. This way, users are initially granted access only to the strictly necessary actions, while administrators have the flexibility to gradually expand user privileges as needed. Resources such as the queue daemon, the queue database and the Auto Scaling groups are shared within all users in one account, allowing multiple users to scale their processing without the need of duplicating resources. These resources are deployed and configured using a special super user that only administrators have access to: the captain. When two or more users specify the same Auto Scaling group to process their jobs, instances from that group will process all relevant queues in a first-come-first-served basis unless users assign priority to their queues. Queues are resources that are user-specific instead of being shared. To avoid the accidental deletion of queues or submission of jobs to incorrect queues, these are tagged with the name of the user that created them and the default permissions allow users to only interact with the queues they have created.

There are two resources that must be permanently running for users to be able to scale batch workloads with SEPPO: the queue database and the queue daemon. Furthermore, the three private buckets containing user credentials, code and data must exist. These resources can be deployed in any available AWS region, a specific geographic location where Amazon maintains multiple physical data centers. AWS Cloud currently spreads across 31 different geographic regions serving 245 countries and territories and is expected to include 5 new regions in the future. While these permanent resources are usually deployed in only one region for cost-saving purposes, SEPPO allows users to create queues and Auto Scaling groups and launch instances in multiple regions, enabling them to scale their processing jobs close to where their data sits. The queue database can store metadata for queues across multiple regions, while the queue daemon has the capability to modify the desired capacity of Auto Scaling groups in those regions as well.

3. Case Study: Deforestation monitoring in an Amazonian deforestation hotspot: Caquetá, Colombia

4.1 Study Area

The concept of cloud processing environment and highly performance environments to generate deforestation monitoring was tested in one of the areas with highest deforestation rates in Colombia: the Caquetá department (Correa Ayram et al., 2020), (excluding the Solano municipality, **Figure 12**. Solano was not included due the continuous and unaltered forest cover, where no deforestation events are expected. The study area extends for 90,108 Km2 and includes 15 municipalities of Caquetá, corresponding to 53.04% and 3.04% of the total area of the department and the country, respectively. Caquetá is characterized by three distinct morphological units: the eastern flank of the Eastern mountain range, the foothills, and the Amazon plain; the region experiences a decrease in rainfall from December to February, while May and June are the wettest months (Amazonia, 2005). Caquetá has historically witnessed one of the highest deforestation rates in the country, particularly during the period between 1990 and 2010, with a deforestation rate of 0.77% per year based on a six-

teen year study (Murad and Pearse, 2018). The causes of deforestation in the area are associated with various factors, including complex colonization pressure, the intensification of coca cultivation, expansion of cattle ranching activities, and land appropriation (Murad and Pearse, 2018). The Caquetá deforestation hotspot was selected as study area to evaluate cloud processing environment and high-performance systems due the importance to deploy a cost-efficient, elastic, transparent, and consistent deforestation monitoring, reporting and verification system that can be accessible to multiple local, regional, national, and global stakeholders.

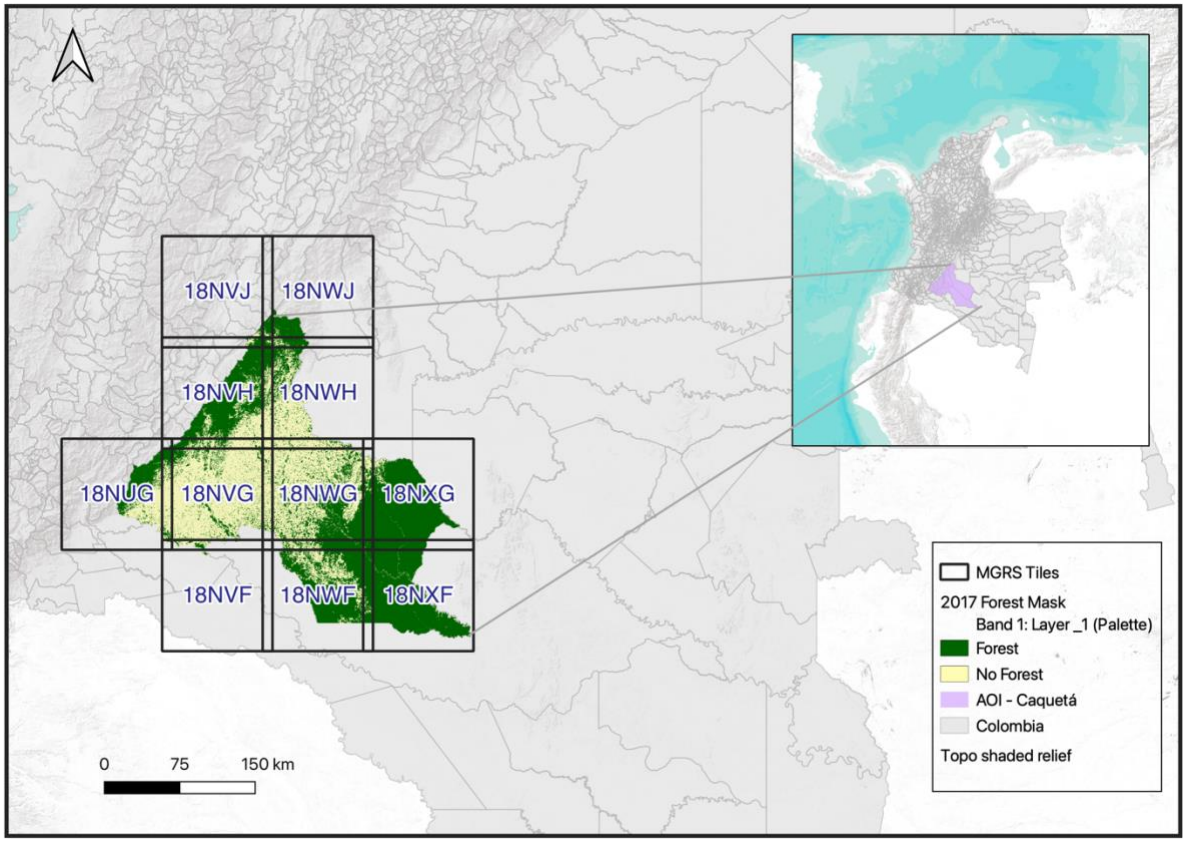


Figure 12. Caquetá State study area (purple), except for Solano municipality, and forest extent by the end of 2017 (green) and MGRS tiles used to process Sentinel-1 products (black squares)

4.1 Deforestation monitoring through the interoperability of open source and cloud-based technologies to process large volumes of Earth Observation Data

The deforestation monitoring was conducted in three distinct phases, through SEPPO and Nebari two cloud-based systems that seamlessly inter-operate. In the first phase, the SEPPO software was employed to pre-process all available Sentinel-1 imagery to generate EO time series stacks. In the second phase, the Nebari platform facilitated the co-development and testing of Python routines. This phase is primarily centered around a single tile, enabling to fine-tune and optimize the routines before scaling them up to the larger study area. Nebari provided a reliable and efficient environment for the iterative refinement of the monitoring techniques. Once the deforestation monitoring workflow had undergone thorough testing and verification, the third phase utilized SEPPO software to scale the monitoring process across the 11 tiles within the study area (**Figure 12**).

1. Earth Observation Data pre-processing using SEPPO

The SEPPO software was parameterized to access, structure, pre-process, and deliver Sentinel-1 EO data for deforestation monitoring purposes. All available Copernicus Sentinel-1 mission's satellite

Single Look Complex (SLC) imagery, from the Interferometric Wide Swath (IW) with 250 km swath width across-track, including Sentinel-1A (2014) and Sentinel-1B (2016), was accessed through the Alaska Satellite Facility (ASF) platform, through the ASF-DAAC data repository (i.e., an AWS S3 bucket in the us-west-2 region). The pre-processing workflow was divided into three different group of jobs that were submitted to independent processing queues: (1) Radiometric Terrain Correction (RTC), (2) tiling, stacking and multi temporal filtering, and (3) conversion to a cloud-native geospatial data format (i.e. Zarr). Each of these groups was assigned a different Auto Scaling group for processing (Table 5). Initially, 250 km tracks were sliced in 110 km * 110 km tiles based on the Military Grid Reference (MGRS) to generate 20-meter resolution multilook geocoded images radiometrically and terrain corrected with a 30 meters resolution Copernicus DEM, in Universal Transverse Mercator (UTM) projection (**Figure 12**). Pre-processing routines were implemented using the Gamma® software (Werner et al., 2000) and scaled by SEPPO software in AWS cloud-computing infrastructure to stack all available imagery for each tile and path on next-generation files formats (NGFF's) as Zarr files (Miles et al., 2020) that correspond to large N-dimensional arrays that provides an efficient way for object storage and parallel computing on the cloud.

Table 5. Auto scaling groups with the minimum and maximum AWS EC2 machines for pre-processing routines applied to Sentinel-1 imagery (i) radiometric and terrain correction (RTC), (ii) tiling, stacking, and filtering, (iii) and convert to Zarr format, available in SEPPO

Auto Scaling Group	Min	Max	Jobs	vCPU	Memory (GiB)
r5.24xlarge_1	0	100	Tile, stack and filter	96	768
r5.4xlarge_1	0	300	Convert to Zarr	16	128
r5.xlarge_1	0	1000	RTC	4	32

2. *Deforestation monitoring: routines refinement using Nebari Hub and scaling with SEPPO*

Nebari is an open-source data science platform that was used to develop, test, and launch Python routines to detect and quantify deforestation events in Caquetá integrating EO Zarr stacks generated using SEPPO in combination with tools and technologies for geospatial analysis. Nebari uses Terraform (Brikman, 2017), which is a infrastructure as code tool to efficiently deploy cloud infrastructure based on configuration templates. This is done through the execution of a plan based on existing infrastructure and the expected configuration to perform all proposed operation in the correct order, respecting dependencies, and providing secure authentication processes for private cloud environments. Complementary to infrastructure as code available through Terraform, Kubernetes was implemented for deployment of clusters. Kubernetes corresponds to a revolutionary way that leads the container orchestration (Shah and Dubaria, 2019; Vayghan et al., 2019). Kubernetes in AWS (AWS EKS) provided mechanisms to control the cluster by orchestration of containers with workers that executed all deforestation analysis workloads. These containers are scheduled and orchestrate to run machines on the cloud in a way to scale applications based on resource utilization metrics were monitoring mechanisms can maintain, deploy, heal, and scale containers to ensure that every workload can be completed without failures.

Built upon this modern infrastructure for cloud cluster management, Dask library (Rocklin, 2015) was used for parallel and distributed computing using the Python programming language. Through the Dask Gateway, Kubernetes was launched in a shared, centrally managed cluster environment providing a secure, multi-tenant, flexible and robust to failure processing. Zarr arrays, through Dask and Xarray libraries, were divided in multiple subsets of the original array, called chunks, in a way that every chunk can be efficiently processed by available workers. Unlike traditional approaches, where all operations are immediately executed, Dask arrays create a task queue where operations are mapped over blocks based on metadata. Computation is deferred until values are explicitly requested to be computed, for example when writing to disk is requested. When computation is triggered, data is loaded into memory, and the processing occurs in a sequential, block-by-block manner. This concept is called lazy computing.

Nebari was deployed on the AWS us-west-2 region using an adaptive scaling approach with a maximum of 20 workers each one with 8 CPU's with 32 GB each. For each 110 km * 110 km tile chunks were optimized as: 20, 512, 512 for time, latitude, and longitude array dimensions, respectively. Python language libraries for geospatial analysis as GDAL, NumPy, Pandas, Xarray, HoloViews, and Cartopy were used to develop deforestation monitoring routines (Table 6. All libraries integrated in the routines can inter-operate with Dask to ensure the optimization of computational operations.

Table 6. Python based libraries integrated in the development of geospatial routines for deforestation monitoring

Library	Description	Source
GDAL	Raster and vector data manipulation, format conversion, and geospatial analysis	https://gdal.org
Dask	Parallelize Python code	https://dask.org/
NumPy	Numerical operations on multi-dimensional arrays	https://numpy.org/
GeoPandas	Data manipulation and analysis, spatial operations on geometric types	https://geopandas.org/
Xarray	Handle and analyze multi-dimensional data, particularly large-scale data sets	https://xarray.dev/
hvplot	Data exploration and visualization	https://hvplot.holoviz.org/
HoloViews	Create interactive visualizations	https://holoviews.org/
Rasterio	Organize and store gridded or raster datasets	https://pypi.org/project/rasterio/

3. Deforestation detection: the Cumulative Sum approach

The algorithm implemented in this work to evaluate the trajectories of vegetation structure is based on time series techniques of backscatter values associated with the Sentinel-1 C-band. The algorithm analyses the backscatter time series to detect deforestation events from 2017-12-31 to 2023-06-09. Due the available Sentinel-1 earth observation data have been collected over a long period of time; analysis techniques commonly applied to other disciplines, such as economics, can be integrated to analyze patterns, trends and determine thresholds in EO data (Regier et al., 2019). However, environmental time series data, are often characterized by non-normal distributions, serial correlations, data gaps, outliers, and seasonal regime shifts (Darken et al., 2002; Regier et al., 2019), specifically in biological/environmental studies where exogenous disturbances are unavoidable (Tsay, 1988). These characteristics often require manipulation of the data prior to analysis to fit assumptions such as independence of variables, which makes the interpretation of the data complex and generates artifacts in the dataset (Regier et al., 2019).

To overcome these limitations of conventional time series approximations of earth observation data, the Cumulative Sums method (Schweder, 1976) was implemented. The cumulative sums approximation represents a useful statistical tool for the interpretation and analysis of complex time series without the need to pre-process the initial data (Regier et al., 2019). An approach that has been implemented from industrial process (Mawonike et al., 2018), and has recently been widely applied in environmental sciences (Adrian et al., 2006; Beaugrand and Reid, 2003; Norström et al., 2011; Regier et al., 2019), and more recently applied to time series of remotely sensed data (Kučera et al., 2007; Ruiz-Ramos et al., 2020) (Kellndorfer, 2019).

The initial detection of disturbances in the time series was the result of a Cumulative Sums analysis performed on the corrected and calibrated Sentinel-1 images with 20 m pixel size. Cumulative Sum analysis of time series is the basis of classical change point detection by means of mean shift detection, which investigates the change of the mean before and after a change in a time series (Schweder, 1976). It is a distribution-free approach, applicable to short and irregular time series to detect gradual and sudden changes. Initially, geometric means and time series residuals are calculated from each observation for each pixel in the time series. Then the residuals of the time series are calculated against the geometric mean of the minimum and maximum values of the time series. With the advantage that the maximum value of the cumulative sum of the residuals is independent of the temporal occurrence of the change. In this way, the thresholds applied to restrict the analysis based on the maximum or minimum values of the cumulative sum curve are more consistent over time. From the time series of the residuals, we calculate the time series of the cumulative sum of the residuals. The detection of deforestation change corresponds to the values of the cumulative curve that are higher than the geometric mean (slope). Additionally, the magnitude of change is estimated based on the difference of the maximum and minimum values of the cumulative sums to determine, high values indicate strong changes in the fall of backscatter (deforestation) compared to low values at the same point in time. To determine if these potential change points correspond to deforestation, bootstrapping was performed to compare with multiple randomly generated cumulative curves, and to determine the confidence level of detection.

The Cumulative Sums algorithm was applied to all forest patches standing at the 2017-12-31, based on the initial implementation proposed by Kellndorfer (Kellndorfer, 2019) using ALOS PALSAR L-BAND SAR sensor, and adapted in this work to Sentinel-1 C-band sensor. A forest mask generated from the integration of 2017 Sentinel-2 imagery using the red, NIR, and SWIR2 spectral

bands and calibration points, 300 and 380 for forest and no-forest classes respectively, in a pixel-based supervised random forest machine learning algorithm (Pedregosa et al., n.d.). Forest mask based on Sentinel-2 imagery, at 10 meter spatial resolution, was re sampled to 20 meter resolution to ensure co-registry with Sentinel-1 imagery. The Cumulative Sums algorithm was implemented using the SEPPO Software for all 11 tiles using a auto scaling group of r5.4xlarge machines with a maximum of 300 machines in AWS EC2.

4. Results

4.1 Earth Observation Data pre-processing using SEPPO

Using the software SEPPO in AWS cloud-based infrastructure, 2790 single imagery from Sentinel-1 for all 11 tiles for 77 different satellite passes, was pre-processed to obtaining radiometric and terrain corrected imagery stacks which served as input for deforestation time series detection. Every single image pre-processing mean time was 17.83 minutes, stacking in Zarr files 29.93 minutes, and multi-temporal speckle filtering 30.44 minutes Table 7. If add up the individual times of each task associated with the pre-processing job group (RTC,Zarr, and speckle), 829.06 hours will be the time it takes for the analysis without paralleling computation through the cloud. Nevertheless, the entire process was optimized to be completed in 1.23 hours utilizing parallel computing on a cloud-based environment managed by SEPPO, up to a maximum of 100 CPUs were fired up for tiling, stacking, and filtering, while 1000 CPUs were employed for the RTC process. The total data size for all final products was 523.6 GiB, and the tasks performed included RTC, tiling, multitemporal speckle filter, and VRT routines.

Table 7. SEPPO processing time (minutes) in each job group associated with pre-processing and analysis of Sentinel-1 time series (2017-01-01 to 2023-06-09) for deforestation detection. *Number of individual EO acquisitions from Sentinel-1. **Number of tiles. ***Number of Sentinel-1 paths for all tiles.

Job Group	No. of Jobs	Mean (Std) Minutes	Total hrs individual jobs	Total hrs parallel computing
RTC and tiling	2790*	17.83 (1.54)	829.06	1.23
VRT to Zarr	11**	29.93 (9.74)	5.49	0.89
Speckle filtering	11**	30.44 (6.81)	5.58	0.67
Change detection	77***	58.55 (32.02)	75.14	2.72

A total of 77 passes of Sentinel-1 associated to 11 tiles (Table 7) were integrated into the Cumulative Sums algorithm to detect deforestation events from 2017-12-31. Change detection for

each analyzed passes were performed in 58.55 minutes each in average and the entire process of analysis was completed in 2.72 hours by SEPPO compared to 75.14 hours if the individual process were performed in a sequence by traditional approaches (Table 7).

4.2 Deforestation change detection using CumulativeSum

Figure 3shows the forest and deforestation trends during the analyzed period from December 31, 2017, to June 9, 2023. The spatial distribution of deforestation events highlights that the most significant deforestation events occurred in the lowlands of Caquetá, specifically in the municipalities of San Vicente del Caguán and Cartagena de Chaira. These areas present extensive deforestation events, primarily concentrated along the Caguan, Yarí, Cuemaní, and Camuya rivers. However, the foothills and mountainous regions of Caqueta also presente some instances of deforestation. Nevertheless, these events were relatively dispersed in number and less extensive compared to the deforestation observed in the lowlands.

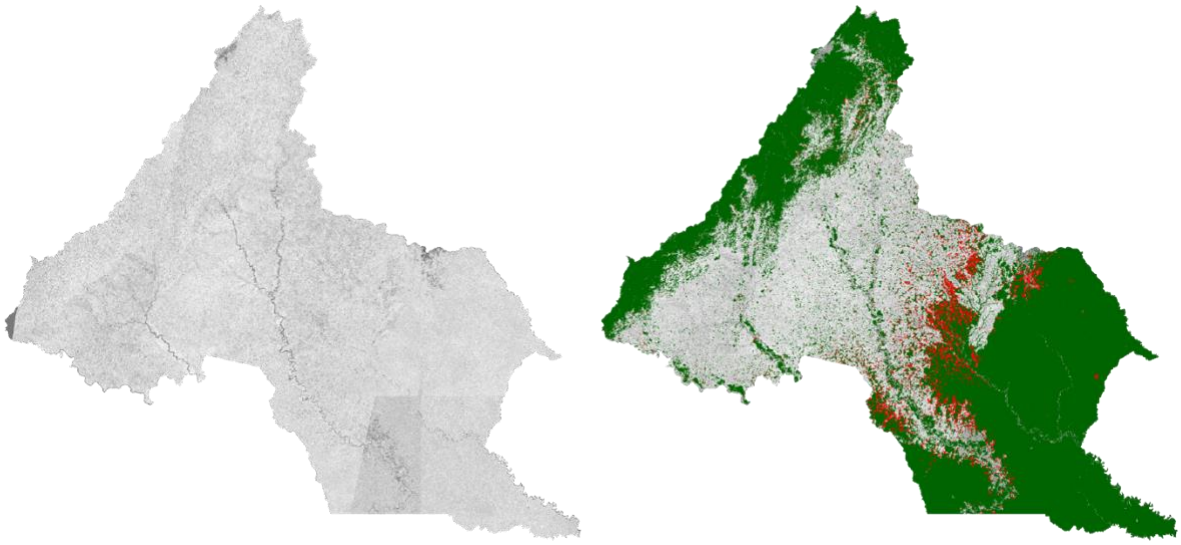


Figure 13. Sentinel-1 mosaic for the latest imagery acquired (left) and deforestation detection (red areas) results from the Cumulative Sums analysis (right). All imagery are descending 069 path acquired on the 2023-06-01 for 18NVJ, 18NWJ, 18NVH, 18NWH, 18NUG, 18NVG,18NWG, 18NVF, and 18NWF tiles; for 18NXG and 18NXF acquisition dates are 2021-12-07 and 2021-12-21, for ascending paths 150 and 171 respectively

Based on land cover change estimation areas (Table 8), in 2017, it was estimated that there were 2,669,237 hectares of forested areas, accounting for approximately 49% of the total Area of Interest (AOI). On the other hand, non-forest areas were estimated to cover 2,106,421 hectares, constituting around 51% of the total AOI. By 2023, the results obtained through the integration of Sentinel-1 time series EO data using the Cumulative Sums approach have detected a total of 83,302 hectares deforested, representing approximately 2% of the total AOI area.

Table 8. Landcover extent dynamics for 2017-12-31 to 2023-06-09 period

Landcover	**2017**	**2023**

	Hectares	Area proportion	Hectares	Area proportion
Forest	2,669,237	0.49	2,585,935	0.54
No Forest	2,106,421	0.51	2,106,421	0.44
Deforestation	NA	NA	83302	0.02

A robust approach to detect deforestation events and estimate confidence levels associated to potential deforestation events. Through Cumulative Sums, valid deforestation pixels (Figure 14, center), along with the identification of invalid deforestation pixels (Figure 14, right) and confidence level pixel (Figure 14, left) were estimated. The magnitude of change in deforestation pixels was determined by calculating the difference between the maximum and minimum values of the cumulative sum of residuals over time. To identify deforestation pixels accurately (Figure 14, center), a bootstrap approach was employed, involving the random reordering of time series (bootstraping = 100) and comparing the resulting Cumulative Sums curves. The confidence level was then estimated based on the frequency with which the Cumulative Sums value exceeded the randomly generated CumSum values. This threshold of 1 was applied to the confidence level, ensuring a conservative estimation of deforestation pixels (Figure 14, center) and discard invalid pixel for deforestation (Figure 14, right).

Figure 14. Confidence level and derived masks after bootstrapping the process of change point detection. Left: Confidence Level. Center: Valid change points (red). Right: Invalid change points (gray).

Results from the Cumulative Sums approach for deforestation area detection are further complemented by the algorithm's ability to identify the date of change using high-density time series Earth observation (EO) data from missions such as Sentinel-1. **Figure 15** showcases imagery from December 2017 and June 2023, highlighting two coordinates of deforestation detection: point A and point B. The deforestation pixels are examined through the time series backscatter variation and cumulative sums of the residuals time series (Cumulative Sums). In the figure **Figure 15**, the top panel displays the backscatter time series, while the bottom panel illustrates the Cumulative Sums time series and random bootstrap Cumulative Sums. The backscatter values for the deforested pixel are plotted against the geometric mean estimate between the maximum and minimum backscatter values (red line). The random bootstrap Cumulative Sums values are also plotted in the bottom panel (gray lines). The detected change point date is indicated by a vertical black line. These visual representations provide a comprehensive understanding of the deforestation dynamics and the reliability of the change detection algorithm.

Figure 15 provides a concise summary of the two predominant deforestation patterns observed through the application of the Cumulative Sums algorithm on Sentinel-1 EO time series data. The first deforestation change pixel, labeled as A, represents a single change event characterized by a noticeable drop in the backscatter signal at the beginning of 2022. This deforestation detection was further validated by the cumulative sum of the residuals, which exhibited an inflection point in the time series, indicating a significant change. The deviation from randomly bootstrapped values further supported the accuracy of the detection. Moving on to deforestation change pixel B, both the backscatter and Cumulative Sums time series patterns indicate multiple change events occurring over time for the same pixel. The initial deforestation event took place in early 2019, evident by a drop in backscatter value. This was followed by a subsequent increase in backscatter after mid-2019. Finally, a second change detection occurred at the end of 2022, related to a second drop on the backscatter values.

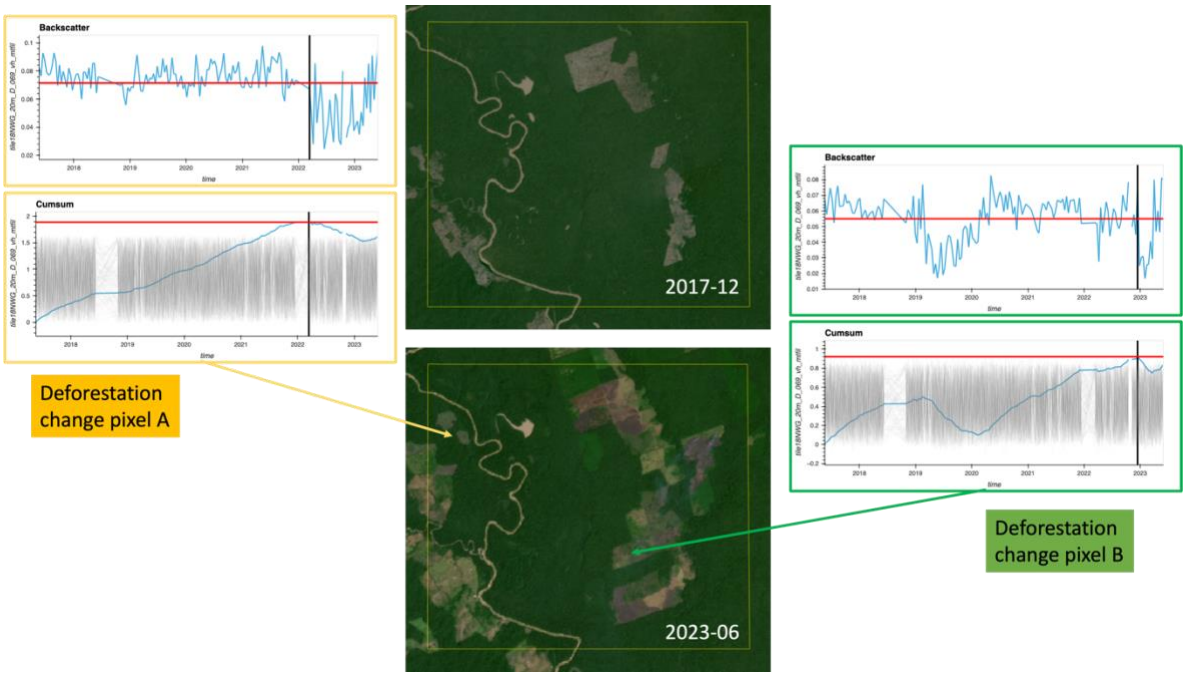


Figure 15. Visualization of change points based on Cumulative Sums results between 2017-12-31 and 2023-06-09. Where backscatter time series (blue line) is plotted with the geometric mean estimate between max(X) and min(X) (red lines) to estimate the Cumulative Sums (S curve) based in residuals. The detected change point location is indication by the vertical black line. Random time series bootstraps (100) are background gray lines.

5. Discussion and Conclusions

This work demonstrates the successful application of open-source and cloud-based software and platforms, i.e. SEPPO and Nebari, in implementing a comprehensive and robust deforestation monitoring workflow. This workflow leveraged high-density Earth observation (EO) time series data, enabling the analysis of significant volumes of information. The utilization of advanced digital infrastructure allowed for efficient storage and processing of large datasets, resulting in improved turnaround times for delivering deforestation monitoring products.

The implementation of SEPPO and Nebari platform facilitated a transparent and traceable monitoring process, ensuring the reliability and reproducibility of the results. The integration of these tools with high-density EO time series data provided a detailed understanding of deforestation

dynamics, enhancing the accuracy and comprehensiveness of deforestation detection. Furthermore, the cloud-based nature of the software enabled the scalability and flexibility necessary to handle substantial volumes of data efficiently. This allowed for seamless processing of large datasets, enabling timely analysis, and reporting of deforestation patterns.

Based on the observed results in Figure 5, two different deforestation patterns were observed. Deforestation in pixel A likely represents a specific deforestation pattern where the cleared area is maintained as pasture, without any significant regrowth or introduction of crops. This pattern is usually explained when the primary purpose of deforestation is to create grazing land for cattle. The prominence of pasture as the predominant land cover aligns with this observation, indicating a potential correlation between deforestation and the establishment of cattle ranching activities. Deforestation in pixel B exhibits a different deforestation pattern where backscatter drop after deforestation can be followed by an increase in backscatter. This can be explained when the deforested area is not completely cleared, as evidenced by the presence of remaining structures or vegetation. Another potential scenario is that forests are cleared, possibly for the introduction of crops or other land uses. The subsequent regeneration observed in the backscatter values after mid-2019 indicates a degree of regrowth or changes in vegetation cover in the cleared areas. This finding suggests a more complex land use transition where deforested areas may be repurposed for agricultural activities or a combination of land uses.

These insights corroborate some limitations of C-band radar data, specifically related to backscatter saturation. The saturation effect can limit the sensitivity of the backscatter signal to accurately detect changes in deforested areas. This limitation implies that the observed patterns and detected change events may not fully capture the extent of deforestation in regions where backscatter values are saturated, especially due to the wavelength of the C-band. Compared to L-band that usually don't show any backscatter increase after deforestation; unless enough time has passed without interventions for the forest to have managed to return to a structural state similar to that of an unmanaged forest. Therefore, additional data sources and analysis techniques may be required to complement the observations made using C-band radar data.

Furthermore, it is important to note that the application of the Cumulative Sums algorithm requires careful parameterization to obtain reliable results. This is particularly crucial when examining historical deforestation patterns. The choice of the time frame for assessment plays a significant role in the accuracy of the detected deforestation events. If the time window is too short, there is a risk of underestimating deforestation events, as smaller-scale changes may go undetected. Conversely, a longer time frame can provide an advantage in identifying older deforestation events. Thus, finding the appropriate balance in selecting the time frame is essential to ensure comprehensive and accurate assessment of deforestation dynamics. The proper parameterization of the Cumulative Sums algorithm enhances its effectiveness as a tool for detecting and monitoring deforestation, contributing to improved understanding and management of forest ecosystems.

6. References

Abernathey, R.P., Blackmon-Luca, C.C., Crone, T.J., Henderson, N., Lepore, C., Banihirwe, A., Hamman, J.J., Gentemann, C.L., McCaie, T.A., Robinson, N.H., Signell, R.P., 2021. Cloud-Native Repositories for Big Scientific Data. Computing in Science & Engineering 23, 26–35. https://doi.org/10.1109/MCSE.2021.3059437

Adrian, R., Wilhelm, S., Gerten, D., 2006. Life-history traits of lake plankton species may govern their phenological response to climate warming. Global Change Biology 12, 652–661. https://doi.org/10.1111/j.1365-2486.2006.01125.x

Amazonia, C., 2005. River Seed Dispersal Modes of the Sandstone Plateau Vegetation of the Middle Caqueta. Methods 37, 64–72. https://doi.org/10.1111/j.1744-7429.2005.03077.x

Beaugrand, G., Reid, P.C., 2003. Long-term changes in phytoplankton, zooplankton and salmon related to climate. Global Change Biology 9, 801–817. https://doi.org/10.1046/j.1365-2486.2003.00632.x

Y Brikman. "Terraform: Up and Running, First Edit". In: O'Reilly Media (2017)

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

Darken, P.F., Zipper, C.E., Holtzman, G.I., Smith, E.P., 2002. Serial correlation in water quality variables: Estimation and implications for trend analysis. Water Resources Research 38, 22-1-22–7. https://doi.org/10.1029/2001wr001065

Earth Science Data Systems, N., 2019. Earthdata Cloud Evolution [WWW Document]. Earthdata. URL https://www.earthdata.nasa.gov/eosdis/cloud-evolution (accessed 5.19.23).

Fisher, C., 2018. Cloud versus On-Premise Computing. AJIBM 08, 1991–2006. https://doi.org/10.4236/ajibm.2018.89133

Giuliani, G., Camara, G., Killough, B., Minchin, S., 2019. Earth Observation Open Science: Enhancing Reproducible Science Using Data Cubes. Data 4, 147. https://doi.org/10.3390/data4040147

Josh Blumenfeld, E.S.W., 2017. EOSDIS Data and Services in the Cloud | Earthdata [WWW Document]. URL https://www.earthdata.nasa.gov/learn/articles/cmr-and-esdc-in-cloud (accessed 5.19.23).

Kellndorfer, J. "Using SAR Data for Mapping Deforestation and Forest Degradation". en. In: Synthetic Aperture Radar (SAR) Handbook:Comprehensive Methodologies for Forest Monitoring and Biomass Esti-mation. Ed. by Africa Flores et al. NASA, 2019. Chap. 3, pp. 65–79.doi: 10.25966/NR2C- S697. url: https://gis1.servirglobal.net/TrainingMaterials/SAR/SARHB_FullRes.pdf (visited on 04/30/2023

Kučera, J., Barbosa, P., Strobl, P., 2007. Cumulative sum charts - A novel technique for processing daily time series of MODIS data for burnt area mapping in Portugal. Proceedings of

MultiTemp 2007 - 2007 International Workshop on the Analysis of Multi-Temporal Remote Sensing Images. https://doi.org/10.1109/MULTITEMP.2007.4293051

Mawonike, R., Chigunyeni, B., Chipumuro, M., 2018. Process improvement of opaque beer (chibuku) based on multivariate cumulative sum control chart. Journal of the Institute of Brewing 124, 16–22. https://doi.org/10.1002/jib.466

Miles, A., jakirkham, Durant, M., Bussonnier, M., Bourbeau, J., Onalan, T., Hamman, J., Patel, Z., Rocklin, M., shikharsg, Abernathey, R., Moore, J., Schut, V., dussin, raphael, Andrade, E.S. de, Noyes, C., Jelenak, A., Banihirwe, A., Barnes, C., Sakkis, G., Funke, J., Kelleher, J., Jevnik, J., Swaney, J., Rahul, P.S., Saalfeld, S., john, Tran, T., bot, pyup io, sbalmer, 2020. zarr-developers/zarr-python: v2.5.0. https://doi.org/10.5281/zenodo.4069231

Murad, C.A., Pearse, J., 2018. Landsat study of deforestation in the Amazon region of Colombia: Departments of Caquetá and Putumayo. Remote Sensing Applications: Society and Environment 11, 161–171. https://doi.org/10.1016/j.rsase.2018.07.003

Nativi, S., Mazzetti, P., Santoro, M., Papeschi, F., Craglia, M., Ochiai, O., 2015. Big Data challenges in building the Global Earth Observation System of Systems. Environmental Modelling & Software 68, 1–26. https://doi.org/10.1016/j.envsoft.2015.01.017

Norström, S.H., Bylund, D., Vestin, J.L.K., Lundström, U.S., 2011. Initial effects of wood ash application on the stream water chemistry in a boreal catchment in central sweden. Water, Air, and Soil Pollution 221, 123–136. https://doi.org/10.1007/s11270-011-0775-z

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., n.d. Scikit-learn: Machine Learning in Python. MACHINE LEARNING IN PYTHON.

Regier, P., Briceño, H., Boyer, J.N., 2019. Analyzing and comparing complex environmental time series using a cumulative sums approach. MethodsX 6, 779–787. https://doi.org/10.1016/j.mex.2019.03.014

Rocklin, M., 2015. Dask: Parallel Computation with Blocked algorithms and Task Scheduling. Presented at the Python in Science Conference, Austin, Texas, pp. 126–132. https://doi.org/10.25080/Majora-7b98e3ed-013

Ruiz-Ramos, J., Marino, A., Boardman, C., Suarez, J., 2020. Continuous forest monitoring using cumulative sums of sentinel-1 timeseries. Remote Sensing 12. https://doi.org/10.3390/RS12183061

Schweder, T., 1976. Some "optimal" methods to detect structural shift or outliers in regression. Journal of the American Statistical Association 71, 491–501. https://doi.org/10.1080/01621459.1976.10480375

Shah, J., Dubaria, D., 2019. Building Modern Clouds: Using Docker, Kubernetes & Google Cloud Platform, in: 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC). Presented at the 2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC), IEEE, Las Vegas, NV, USA, pp. 0184–0189. https://doi.org/10.1109/CCWC.2019.8666479

Tsay, R.S., 1988. Outliers, level shifts, and variance changes in time series. Journal of Forecasting 7, 1–20. https://doi.org/10.1002/for.3980070102

Vayghan, L.A., Saied, M.A., Toeroe, M., Khendek, F., 2019. Kubernetes as an Availability Manager for Microservice Applications. arXiv arXiv:1901.04946.

Wang, L., Yan, J., Ma, Y., 2020. Remote Sensing and Cloud Computing, in: Cloud Computing in Remote Sensing. CRC Press.

Werner, C., Wegmüller, U., Strozzi, T., Wiesmann, A., 2000. GAMMA SAR and interferometric processing software. European Space Agency, (Special Publication) ESA SP 211–219.

Xu, C., 2022. Analyzing large-scale Data Cubes with user-defined algorithms: A cloud-native approach. International Journal of Applied Earth Observation and Geoinformation.

Yao, X., Li, G., Xia, J., Ben, J., Cao, Q., Zhao, L., Ma, Y., Zhang, L., Zhu, D., 2019. Enabling the Big Earth Observation Data via Cloud Computing and DGGS: Opportunities and Challenges. Remote Sensing 12, 62. https://doi.org/10.3390/rs12010062

Chapter 4: Application of a cloud-based infrastructure to deliver specific value-added products for forest monitoring and restoration implementation assessment

This chapter was designed to apply knowledge and technology advances acquired on the two previous chapters, radar application (second) and cloud computing (third), to generate tailor-made forest and restoration products to be delivered to specific users.

This chapter will be published as a scientific paper during the second semester of 2023. The paper is already reviewed by the multiple authors, and next step is to define between two potential Journals to publish this research, here is a description of the potential Journals:

- **Sustainability**, MDPI:
  - **Agricultural and Natural Ecosystems Restoration after Disturbances,** Special Issue.
- **Frontiers in Conservation Science**, Frontiers.
  - **Advances in Privately Protected Areas,** Special Issue

This research is a collaboration between this PhD project and Earth Big Data LLC with Urra Hydropower Sustainable Office to generate specific geospatial products for forest conservation and restoration timeseries change analysis that integrates standard procedures and new approaches. Regarding the forest dynamics analysis this research provide understanding on multiple year forest extent dynamics following the national guidelines of Colombian Forest and Carbon Monitoring System (SMByC), these guidelines were implemented in a cloud-based environment to optimize different stages of the protocol and optimize time product generation. Secondly, new approaches are processed for forest restoration assessment are proposed based on time series analysis using radar EO, all routines are implemented in a cloud-based environment integrating open-source tools.

The present research involved the following collaborations:

Urra Hydropower Environmentally Sustainable. Through the collaboration with the Urra specific user technical requirements were integrated as parameters for analytical routines to fulfill requirements that ensures the integration of research product in formally process aimed to assess the impact of forest conservation and restoration initiatives. Urra provided the ground truth data related to implementation area polygons and dates of restoration implementations for restoration state if implementation assessment.

Forest dynamics and restoration assessment using multi-source Earth Observation data. The case study of the Urra´s Hydropower Restoration Program.

Carlos Pedraza[1,2,*], Nicola Clerici[2], Marcelo Villa[1], Milton Romero[3], Adriana Sarmiento Dueñas[3], Dallan Beltrán Rojas[3], Paola Quintero[4], Mauricio Martínez[4], Josef Kellndrofer[1]

[1] Earth Big Data LLC, Woods Hole, Massachusetts, USA

[2] Department of Biology, Faculty of Natural Sciences, Universidad del Rosario, Kr 26 # 63B-48, Bogotá, D.C., Colombia

[3] 4D elements Consultores, Calle 44A # 53 – 05, Bogotá, Colombia

[4] Urra S.A E.S.P., Environmental Unit, Carrera 2 # 48 – 08, Monteria, Colombia

*** Correspondence:** carlos@earthbigdata.com

Abstract: Implementing the United Nations Framework Convention on Climate Change have become an important goal for environmental and development sectors in Colombia. The objective of this study focuses on the initiatives undertaken by the Urra Hydropower Company to understand forest dynamics and assess forest restoration implementations. This is achieved by integrating remote sensing-based optical and radar Earth Observation (EO) data, which were analyzed in a cloud-based environment. Results demonstrate a substantial decrease in forested areas between 1996-2000: 37,763 hectares, particularly concentrated in the western region of the study area. The accuracy assessment of multiple forest change maps (1996-2021) presented a high precision in detecting deforestation events, while improvements are necessary for accurately representing non-forest areas. Restoration implementation assessment based in time series analysis from Sentinel-1 SAR EO suggests that the majority of the 270 evaluated plots are in intermediate state (82.96%) when compared to reference data. This study underscores the necessity of robust and continuous monitoring systems that integrate ground truth data with EO techniques for enhanced accuracy and effectiveness in forest restoration and conservation endeavors.

Keywords: forest management[1], forest restoration[2], Sentinel-1[3], Sentinel-2[4], Landsat[5], Time series[6], cloud computing[7].

1. Introduction

In the context of the scientific and technical progress in remote sensing, Colombia is taking advantage of the agreements of the United Nations Framework Convention on Climate Change - UNFCCC during the Conferences of the Parties in 2009 and 2010 (COP 15 and 16, respectively), and recently in Warsaw (COP 19; 2013), which requires developing countries to establish national forest monitoring systems that allow quantifying greenhouse gases emissions/absorptions and changes in the extension of forests and forest carbon stocks. For this purpose, forest cover in Colombia was defined as: "Land occupied mainly by trees that may contain shrubs, palms, *guaduas*, herbs and lianas, in which tree cover predominates with a minimum canopy density of 30%, a minimum canopy height (in situ) of 5 m at the time of identification, and a minimum area of 1.0 ha." (Cabrera et al., 2011). Tree cover of commercial forest plantations, palm plantations, and trees planted for agricultural production are excluded. Deforestation is defined as: "the direct and/or induced conversion of forest cover to another type of land cover in a given period of time", and restoration is defined as: "the recovery of forest cover in areas where it was not present in the past" (Cabrera et al., 2011). These definitions are consistent with the criteria defined by the UNFCCC in its decision 11/COP 7 (FAO, Terms and definitions FRA 2020. 2020), with the definition adopted by Colombia under the Kyoto Protocol (Cabrera et al., 2011).

Thus, in 2011, the Colombian Forest and Carbon Monitoring System (SMByC) evaluated different remote sensing data processing techniques for the detection of forest cover and its respective changes over time, generating as a result the "Protocol for digital image processing for the quantification of deforestation in Colombia at the national level - coarse and fine scale" (Cabrera et al., 2011), which today has been the methodological approach for the historical quantification of deforestation at the national level, e.g. for the periods 1990-2000, 2000-2005 and 2005-2010, and in the last decade annually from 2011 to 2021. These guidelines have become the Colombian standard for the evaluation of spatio-temporal forest and deforestation dynamics, which is applied in the case of national regulations for the case that companies must report their effectiveness in protecting forests.

Since the first missions that integrated sensors on satellites to generate Earth Observation data, e.g. Landsat launched in 1972, a large number of orbiting missions have been launched for environmental applications with different characteristics, providing information can be integrated into multiple applications to interpret land surface patterns and monitor changes in land use land cover (Ustin and Middleton, 2021). Satellite-based remote sensing has been proposed especially as a cost-effective way to provide reliable data on forest change dynamics (Onoda and Young, 2017), (Hansen et al., 2010), (Reiche et al., 2016). Multiple methods have been developed based on Earth Observation from satellites, nevertheless there are limitations associated to each approach, including how to balance the integration of multiple sensors and available resources for their implementation (Mitchell et al., 2017). Some of these challenges are related to complex trajectories and rapid changes associated with land cover changes (Lewis et al. 2015), (Murillo-Sandoval et al., 2023), (Joshi et al., 2015), (Joshi et al., 2016), the inconsistency in the series of EO data between different missions (van Oostende et al., 2022), and specifically in tropical areas the continuous presence of clouds generating uncertainties towards quantitative estimation.

The first launch of global pathfinder optical missions with advanced observations strategies (i.e. Landsat) have been providing consistent EO with the limitation of cloud interference; recently Synthetic Aperture Radar sensors (i. e. ALOS-1 PALSAR-1 and Sentinel-1) continues providing persistent data with the advantage of cloud free EO that allow to meet multiple challenges due to use time series analyses to detect and quantify forest trajectories. The use of synthetic aperture radar (SAR) sensors to generate Earth Observations imagery in recent years has increased in tropical regions, where radar sensor provides day/night imagery without interference from atmospheric conditions (Shimada et al., 2014), (Lucas et al., 2014) (Persaud and Cabrera, 2021). This impulse, which has started in the last decade, mainly with the staging of ALOS-1 PALSAR-1, ALOS-2 PALSAR-2 and

Sentinel-1 sensors, which in addition could obtain EO with a high acquisition frequency, every 6 -14 day.

Regarding the integration of all available EO from optical sensors coupled with advanced techniques to reduce spatiotemporal gaps generated by clouds and the increasingly use of EO from SAR sensors in Colombia, most of the research have been focus to wetlands and flooding mapping (Palomino-Ángel et al., 2019), (Estupinan-Suarez et al., 2015), (Ayala et al., 2015), land use land cover mapping (LULC) (Hoekman and Quinones, 1997), (Hoekman and Quinones, 2000), (Anaya et al., 2023) forest mapping and deforestation monitoring (Hoekman and Quiñones, 2002) (Marcela et al., 2016) (Pedraza et al., 2018) , (Anaya et al., n.d.). Thus, there is a lack of standardized protocols and methodologies based on EO to monitor and assess forest restoration initiatives. Implementation of forest restoration initiatives is called one of the most promising climate actions to rapidly remove CO2 from the atmosphere (Koch and Kaplan, 2022), and new analytical approaches based on remote sensing EO to monitor and assess forest restoration initiatives needs to be developed.

This study aims to: (1) detect forest change dynamics in a 25-year timeframe (1996-2000-2005-2010-2015-2021) based on optical sensors in the surrounding of the Urra hydropower area of influence (Colombia), based on national standards for forest and deforestation monitoring, and (2) design a first methodological approach to assess the status of forest restoration implementations from Urra's hydropower offset strategy based on a 5-year time series of Sentinel-1 C-SAR, Landsat, and Sentinel-2 imagery. Thus, this study proposes an operational monitoring and assessment approach for initiatives aimed to preserve forests, reduce deforestation, and restore forests, by applying scalable and cost-effective technologies and methods.

2. Materials and methods

2.1 Study area

The study area, covering almost 89.000 ha, is located in the northwestern sector of Colombia, between 8°8´N 7° 47 ' N and 76° 27' to 76° 05 ' W. It is characterized by an altitude ranging between 0 and 1250 meters above sea level. Annual rainfall varies between 1.500 and 4.000 mm, with a dry season running from December through the end of March. The mean annual temperature fluctuates between 21,1°C to 28°C, with a warm semi-humid regime climate (Rodriguez et al., 2006). Within this region, three major biomes are identified: i) the tropical wet zone biome, which covers 88% of the area and is the largest one, corresponding to areas where humid rainforest predominates at elevations below 800 meters above sea level (m a.s.l.), ii) the pedobiome of the tropical wet zone biome, which corresponds to areas where vegetation and flora types are determined by soil and humidity conditions (Rodriguez et al., 2006). This region, is mainly represented by the Urra reservoir area, covering 9.3% of the area, and iii) the sub-Andean orobiome, which corresponds to the altitude range between 800 to 1,800 meters a.s.l, where a temperate climatic zone predominates along with the humidity provinces: humid, semi-humid, and super-humid (Rodriguez et al., 2006).

The study area includes 90 villages, within the municipalities of Tierralta and Valencia in the department of Córdoba, and Turbo, Apartadó, Carepa and San Pedro de Urabá in the department of Antioquia. In 1996 in the region began the construction of the Urra Hydropower facility, which collects the water of the Sinú river, and that was put into operation in year 2000.

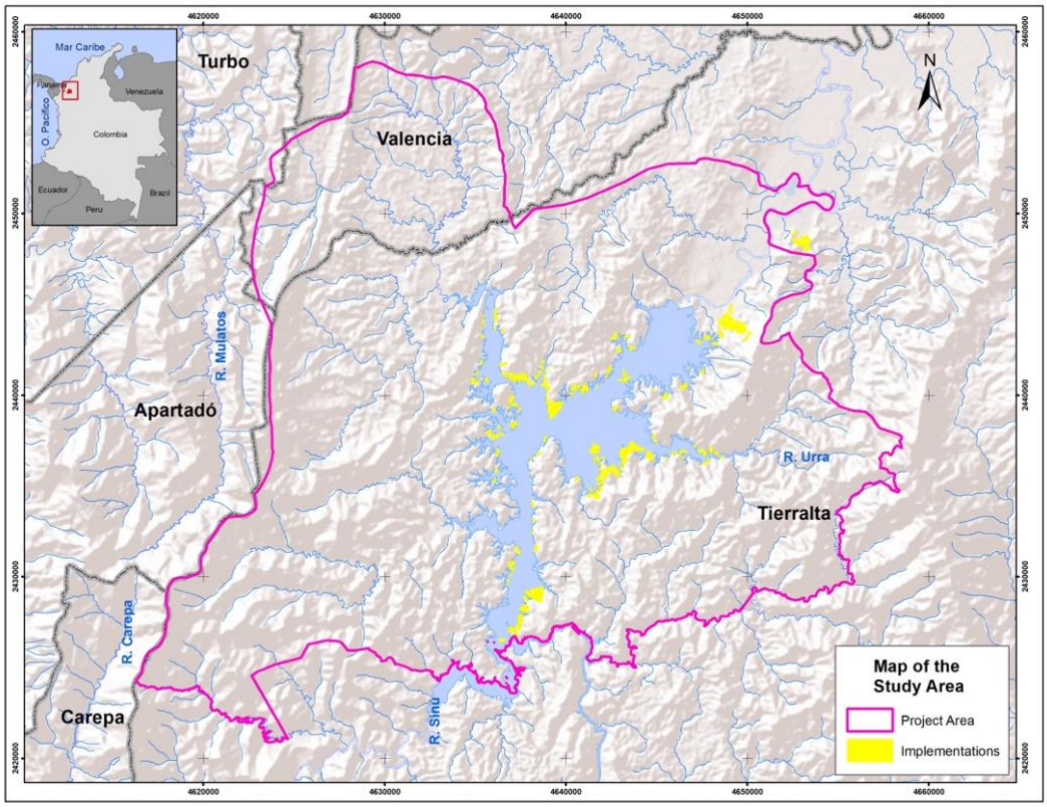


Figure 16. Study area around the influence region of the Urra hydropower project, located in Antioquia Department, Colombia

2.2. Forest dynamics analysis

2.2.1. Sentinel-2 and Landsat pre-processing

A 25-year timeframe (1996-2000-2005-2010-2015-2021) forest change analysis was performed based on the long-term availability of EO from optical sensors (Landsat and Sentinel-2). The Landsat program has been providing continuous observations for about 50 years with 8- or 9- day repeat intervals (Masek et al., 2020). Providing important advantages for long timeframe monitoring applications as the consistency in frequency acquisitions as on multispectral parameters throughout the multiple Landsat programs (Helder et al., 2012). To ensure a complete assessment from 1996 to 2021, multiple Landsat mission products were used, starting with Landsat 5 TM, Landsat 7 ETM+, and Landsat 8 OLI, with a 30 m of pixel spatial resolution accessing to visible bands (red, green, and blue), infrared and panchromatic. Sentinel-2 optical product, Level2a, were also included for the year 2021; imagery from this sensor was included to increase data availability by densifying temporal coverage. Thus, revisit frequency can be improved from 8- or 9-day (based on Landsat) to 2.9 days (integrating Sentinel-2) (Li and Roy, 2017) which can contribute to this study to reduce spatial gaps caused by cloud cover.

Landsat and Sentinel-2 imagery collection 2 Level -2 was accessed through the Microsoft Planetary Computer platform (https://planetarycomputer.microsoft.com). All available imagery of each year with less than 50% cloud cover was transferred to project's cloud-based storage infrastructure, AWS S3 bucket us-east-1. Based on all available imagery for each year analyzed, each band associated to a single acquisition was stacked to generate a multispectral multi-band image (Figure 2A). Next, we performed cloud and shadow masking of each single acquisition scene from

the data series of each year, where every pixel with the presence of clouds and mountain shadows in a single scene was discarded and replaced by data from the closest acquisition with free cloud and shadow pixel from the same year to generate a cloud free mosaic for each year. The masked reflectance surface images were then processed to obtain yearly mosaics with pixels containing only land cover information, i.e. masking and eliminating areas of elements that interfere with the analysis (clouds, banding, shadows, or haze). Orthorectification and/or co-registration procedures were performed for the stacked images to have accurate georeferenced information (Figure 2A). In order to detect changes in the time series it is important that the images are accurately co-registered and/or orthorectified, so that images acquired from different sensors and dates can be directly compared. For the construction of the time series, it is essential to have an accurate pixel-level co-registration between all acquired images for each scene. This adjustment was performed by measuring the difference between pixels and re-projecting the coordinate of the displaced image ends in this same magnitude (considering that the downloaded images are in a projection system with metric units). The UTM projected reference system was maintained during the whole process until the final product, to avoid the loss of co-registration between pixels when applying the geometric adjustments of the cartographic re-projection models.

Precise reflectance and geometric image rectifications are necessary to generate data quality and ensure the delivery of consistent products (Jafarbiglu and Pourreza, 2023), where is necessary to avoid measurement differences to introduce into analysis and avoid false detections, i.e. due to geometric distortions or sensor calibrations. An adjustment to the radiometric signals of the images was thus performed to ensure consistency with each other (Figure 2A), i.e. radiometric calibration, where digital numbers are transformed to surface reflectance that in the case of tropical forest is driven by species composition and vegetation structure characteristics (Asner, 2009). Relative normalization was applied under the assumption that the relationship between the radiances recorded by the sensor on two different dates is spatially homogeneous (Figure 2A). In order to detect deforestation change events for multitemporal forest dynamics analysis, a multivariate Alteration Detection algorithm proposed by (Nielsen et al., 1998) was implemented, which is based on the analysis of canonical correlations, where the 2021 image will be taken as a base and the rest of the images (1996, 2000, 2005, 2010, and 2015) will be normalized with respect to this one. A visual quality control was carried out to ensure that the scenes were correctly co-registered and that they would allow effective multitemporal analysis; this process guarantees the use of different sensors and spatial resolutions so that they can be used in a homogeneous analysis.

.2.2 *Forest multitemporal analysis*

All surface reflectance imagery processed based on Sentinel-2 and Landsat missions were used to generate annual median and last pixel metrics for the red (~665nm), NIR, (~833nm), and SWIR2 (~2190nm) spectral bands, which are the most relevant bands for vegetation change analysis (Galindo et al., 2014a) (**Error! Reference source not found.**C). Last pixel corresponds to the latest pixel acquired w ithout cloud interference near end of the year (December 31st) for the corresponding year. A forest/non forest map was generated for year 2021 as baseline using a pixel-based supervised random forest machine learning algorithm (MLA) included on the sklearn.ensemble Python module (Baranwal et al., 2019)(**Error! Reference source not found.**C). The Random Forest (RF) algorithm has b een widely used in classifying remote sensing data (Tatsumi et al., 2015), (Wang et al., 2015), (Pal, 1996). (Teluguntla et al., 2018), and it is generally adequately handling noise and overfitting of the data, providing quantitative measurement of variables contribution, and handle high data dimensionality (Teluguntla et al., 2018). The RF classifier was initially build using training samples (250 forest and 275 non-forest point samples) generated based on the visual inspection of 2021 high resolution Planet imagery (NICFI , 2021) and 2021 Red, NIR, and SWIR surface reflectance metrics (median and last pixel) (**Error! Reference source not found.**C). Once the first version of the 2021 f orest/non forest map was generated by the RF classifiers, a visual verification of the classified map

was carried out using Planet imagery as quality control, validation, and adjustment procedure (**Error! Reference source not found.**C). If a global accuracy of 95% was not obtained, new samples were generated in regions where misclassifications were observed, the model was recalculated, and this process was iterated until a 95% accuracy was obtained for both classes (**Error! Reference source not found.**C). Sample data and quality control procedures followed the forest definition from the Colombian Forest and Carbon Monitoring System (SMByC).

To generate multitemporal forest change products and estimate deforestation areas between periods 1996-2000, 2000-2005, 2005-2010, and 2015-2021 the RF approach was performed under similar parameters in forest/no forest classification (**Error! Reference source not found.**C). Training points were generated based on visual interpretation of deforested areas from Sentinel-2, Landsat and Planet imagery. Red, NIR, and WNIR bands from two different years, 2021 and 2015, were combined into principal component bands in a single image (**Error! Reference source not found.**C). Principal component analysis has been demonstrated to capture maximum variances in a finite number of orthogonal components based on eigen vectors analysis from the correlation matrix, providing a robust and simple approach to generate input variables for change detection analysis (Deng et al., 2008). The principal component bands generated based on two different years, are highly correlated between unchanged areas, while a low correlation between the significantly changed areas is expected. Once the first version of the 2015-2021 deforestation map was generated by the RF classifiers, a visual verification was performed iterating validation, generation of training points and generation of a new model until accuracies higher than 95% for the deforestation class were obtained (**Error! Reference source not found.**C). Once the expected accuracy was achieved the forest/no forest map of the following year (2015) and the change map (2015-2021) was updated.

.2.3 Sample-based estimation of area and accuracy assessment

Due the importance on the use of area estimations of forest, non-forest, and deforestation based on the thematic maps generated and the need of reproducible protocols for forest and deforestation mapping, an accuracy quantification is needed for reporting and verification purposes (DeFries et al., 2007). A quantitative accuracy assessment for each class of thematic products was carried out to estimate and understand uncertainties on area estimations based bias attributable to omission and commission classification errors (Olofsson et al., 2013) approach has become a well-accepted standard for land change accuracy assessment, including the national standard in Colombia implemented by the SMByC (Galindo et al., 2014b). The initial validation sample size (n) was determined setting the expected standard error Si = 0.01, mapped area proportions for each class, and theorical validation user accuracy defined by the SMByC standards for each class; 99 validation points for forest, 6 for deforestation and 129 for no forest categories were estimated. Nevertheless, the final validation size per category was balanced to a total of 100 points for each class, due to the low numbers obtained for deforestation category explained by the proportion area of this category, the increase of deforestation number of points was implemented to ensure high accuracy on deforestation class to avoid false positives and false negatives. The spatial distribution of validation points was determined by an automated random seed to ensure no bias due to the spatial distribution of validation points. Target user accuracy (U_i) was defined as: 0.9 for the stable forest and non-stable forest cover, since this class is usually considered to have a high accuracy, while for the deforestation 0.8 was assigned.

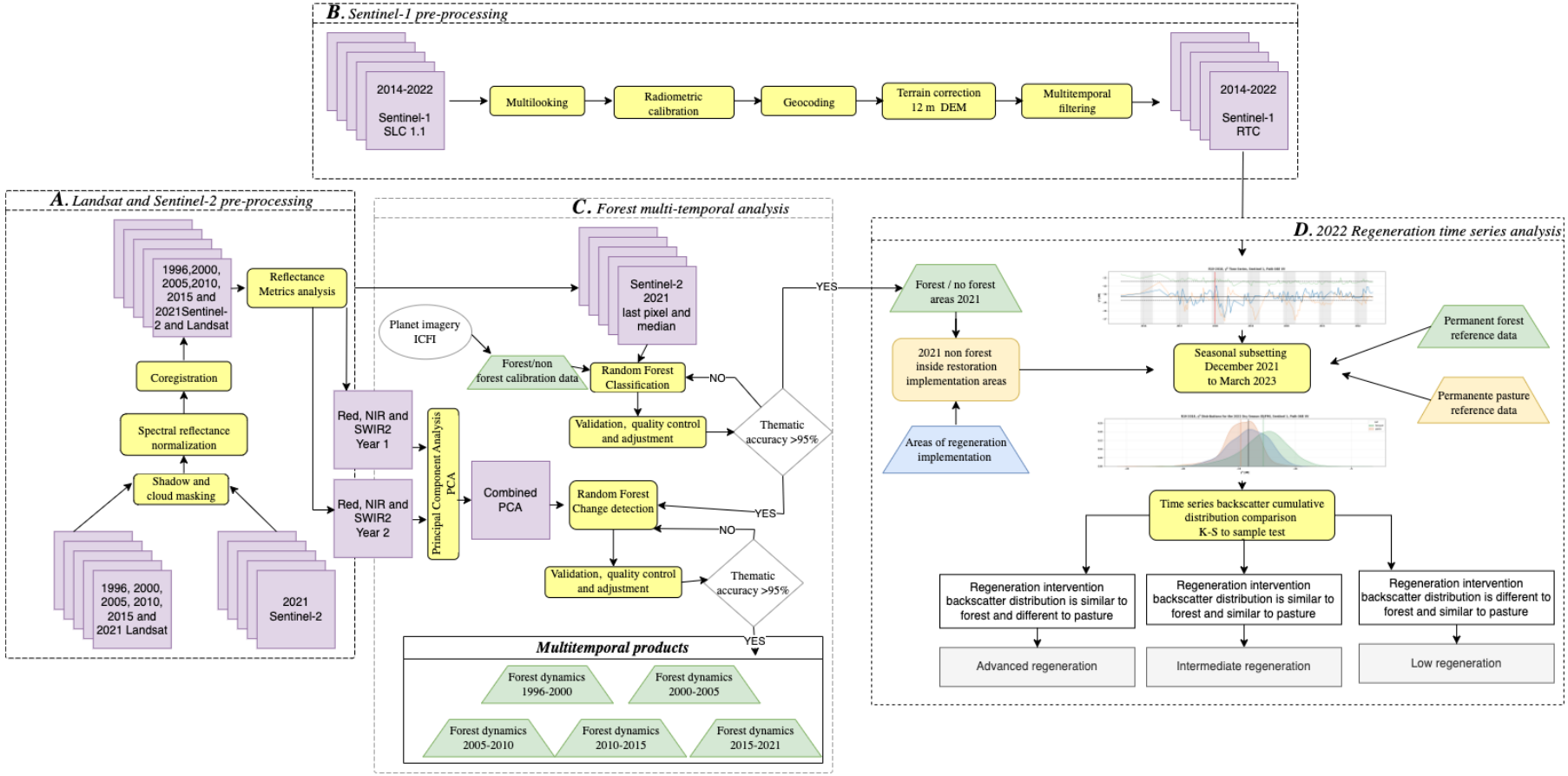


Figure 17. Processing workflow: block A (optical pre-processing) and B (SAR pre-processing) describes all imagery pre-processing generated as inputs to be integrated in block C, to generate multitemporal analysis of forest dynamics, and block D to assess restoration implementations around the URRA hydropower facility.

2.3. Time series analysis for restoration implementation assessment

2.3.1. SAR imagery

Copernicus Sentinel-1 mission launched by the European Space Agency (ESA) is based on a constellation of the two satellites Sentinel-1A (launched in 2014) and Sentinel-1B (launched in 2016), with a revisit time of 12-days each, with a SAR sensor band-c on board. Sentinel-1 SAR imagery was available from the Alaska Satellite Facility (ASF) platform, a mirror of the ESA's Scientific Sentnel-1 hub. All available data from May 2015 to June 2022 was accessed through the ASF-DACC data repository which resides at the AWS S3 bucket region US-west-2. Figure 2B summarizes the five methodological routines for preprocessing Sentinel-1 images, which allows converting the Single Look Complex (SLC) images into 12 m resolution multi-look geocoded, radiometrically and terrain corrected to generate gamma-naught backscatter intensity ($\gamma 0$) images, obtained after pre-processing performed with Gamma® software (Werner et al., 2000). To reduce speckle effects associated to SAR imagery, an enhanced Lee multitemporal filter was applied to the VH polarization, by means of least squares of the signal intensity in a kernel area of 3*3 pixels (Lopes et al., 1990).

2.3.2. Forest restoration status assessment

A time series analysis using Sentinel-1 imagery was used to study vegetation seasonal variations and to assess the status of areas where restoration initiatives were implemented from the year 2014. The polygons of the areas of implementation of restoration measures by the URRA company were integrated (**Error! Reference source not found.**D) and overlapped to the f orest/non-forest areas of the thematic map 2021 (**Error! Reference source not found.**C). The masking process filters the non-forest areas within the intervention plots to perform the restoration status evaluation analysis only in non-forest areas (**Error! Reference source not f ound.**D). For many biomes, seasonal stratification of the time series will improve detection of change events, for example, when dry/wet season conditions introduce significant changes in backscatter due the presence of surface water or phenology variations (Ruiz-Ramos et al., 2020). In this sense, to minimize the influence of seasonal variation in the assessment of structural trajectories of restoration areas, a seasonal trend analysis was introduced in the assessment. Therefore, the first step was to subset the time series data by identifying the dry season, to minimize the precipitation effects on backscatter levels. Dry season months were identified based on time series backscatter from Sentinel-1 from years before 2021 (2017-2020) based on low backscatter values that were validated with precipitation data; imagery from December to March was selected to be included in the analysis as dry season period (Figure 18).

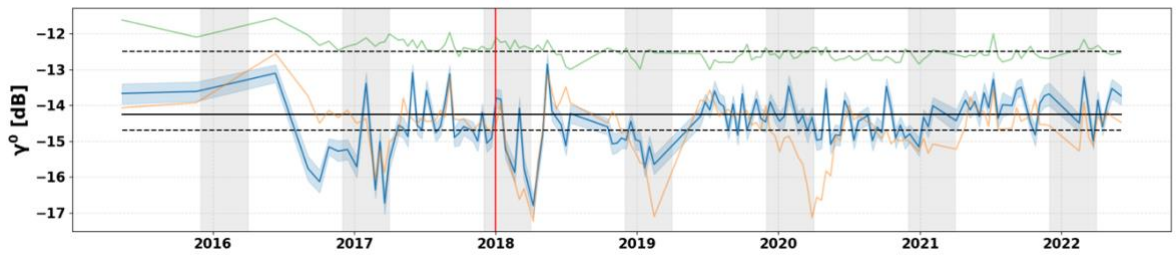


Figure 18. Sentinel-1 vertical-horizontal (vh) backscatter time series profile for an assessed plot (blue line) with implementation year 2018 (vertical red line) contrasted with forest reference time series (green line) and pasture reference time series (yellow line). The dry season was defined from December to March (grey areas).

To assess the current state of the restoration processes a comparative analysis of the dynamics of the vegetation structure from lots undergoing restoration was performed based on Sentinel-1 time series data available (**Error! Reference source not found.**B) to determine if current s tate of the vegetation structure is more similar to reference coverages of well-preserved natural forests or, on the contrary, has a greater affinity with the response of pastures. The assessment was based on a Kolmogorov-Smirnov test, which allows the comparison of the distributions of the backscatter of the implementations with forest and pasture reference data.

$$\mathbf{H_0} = \textbf{The sample of the radar signal backscatter of the regeneration plot and the sample from the reference data belong to the same distribution}$$

$$\mathbf{H_1} = \textbf{The sample of the radar signal backscatter of the regeneration plot and the sample from the reference data DO NOT belong to the same distribution}$$

A total of 270 polygons of intervened lots, which have different years of implementation of reforested and/or revegetated measures (2004-2010, 2012-2018), were analyzed through a comparative analysis of the dynamics in vegetation structure in lots undergoing restoration with respect to two types of reference areas: pasture and forest (Figure 19). Reference data was generated based on the visual inspection of multiyear optical imagery products from Landsat, Sentinel-2, and Planet (1996, 2000, 2005, 2010, 2015, and 2021). Forest reference data areas were defined as consistent forest areas identified in all years and pasture reference data were defined as consistent pasture areas from 2005 to 2021.

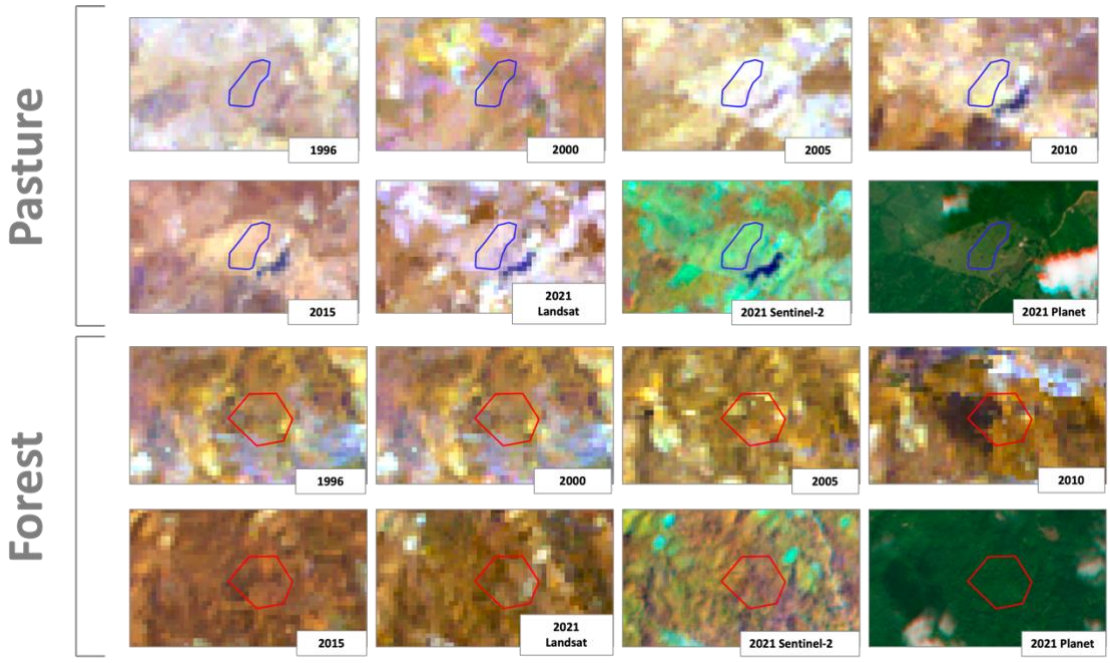


Figure 19. Optical imagery for multiple years was used to generate reference data for forest (magenta) and pastures (cyan) polygons. Landsat and Sentinel-2 imagery correspond to the annual median values combined on NIR (red), SWIR (green), and Red (blue); and Planet Scope December 2021 from the Visual Biannual Archive (ICFI, 2021).

The cross-polarization VH backscatter values from the Sentinel -1 SAR sensor for the most recent dry period (2022) were used as a proxy for the measurement of vegetation aboveground structure (i.e., tree height, canopy size). Multiple studies indicates that SAR cross-polarization (VH in the case of Sentinel-1 or HV in the case of ALOS-PALSAR) shows a higher correlation with biomass compared with co-polarization (VV in the case of Sentinel-1 or HH in the case of

ALOS-PALSAR) to detect vegetation structure changes (M. Susan Moran, 1999), (Rüetschi et al., 2019), (Eriksson et al., 2012), (Kasischke et al., 1997). Backscatter from co-polarization is usually sensible to surface scattering components (Flores-Anderson et al., 2019), frequently applied to detect surface water (Kasischke et al., 1997)

A 95% confidence interval was chosen, which in case the null hypothesis is rejected in favor of the alternative hypothesis if the p-value (significance) is less than 0.05%. This determines the status of the forest restoration implementation based on the following rules: i) if it is in an incipient state, it indicates that it has a greater similarity to pasture; ii) if it is in an intermediate state, it has no similarity to either pasture or forest; and iii) if it is in an advanced state, it has a greater similarity to forest (**Error! Reference source not found.**D).

.4 Computing infrastructure

The Nebari open-source data science platform was used to develop and implement Python routines for restoration assessment, integrating the Sentinel-1 time series backscatter generated using the Software for Earth Big Data Processing, Prediction, Modelling, and Organization (SEPPO) used to access through the cloud to Sentinel-1 imagery and perform all for pre-processing procedures necessary to generate useful time series EO for future analysis. The Nebari platform was configured by Earth Science Information Partners (ESIP-https://www.esipfed.org/), through the Earth Big Data LLC partnership, to run XLarge Instances (8CPU / 32 GB) implementing a parallel computing autoscaling approach with a maximum of 20 workers (CPU).

3. Results

3.1 Optical and radar imagery processing

Sentinel-2 imagery was integrated into the analysis to increase and improve the spatial resolution of thematic products through the generation of the 2021 forest/non forest baseline, achieving a minimum mapping area of between 0.25 ha and 0.5 ha, which is required to scale down to 1:50,000, desired in this study. A total of 9 Sentinel-2 images from the period 2020-2021 were used for this study, corresponding to imagery acquired on the second semester of year 2020. For Landsat, based Surface Reflectance (SR) images annually mosaics were generated as they improve the comparison between multiple images in the same region by considering atmospheric effects, such as aerosol scattering and thin clouds, contributing to the detection and characterization of changes on the Earth's surface. Images correspond to those available in the second half of the year corresponding to the analysis date. For example, for the analysis of the year 2010, images from July 1st to December 31st, 2010, were selected to obtain a composite for the year in question from the historical series of images, metrics, and statistics. A total of 63 images were obtained for the period 1996-2021, to generate yearly median and last pixel partially cloud free mosaics for NIR, SWIR, and Red bands (Figure 20); some pixel for specific years consistently presented the presence on clouds.

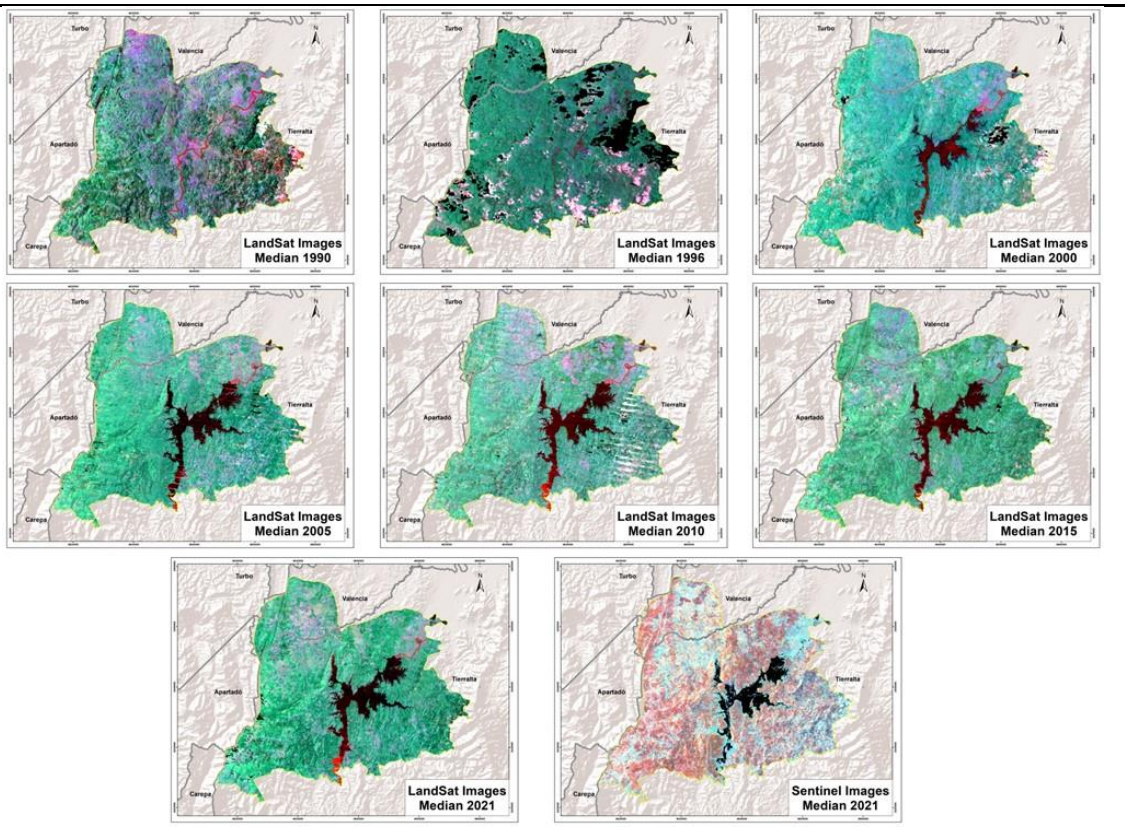


Figure 20. Optical pre-processing results for Sentinel-2 imagery (2021) and Landsat (1996. 2000, 2005, 2010, 2015, and 2021). The mosaics generated correspond to annual median values combined on NIR (red), SWIR (green), and Red (blue).

A total of 285 Sentinel-1 radiometric and terrain corrected images were used for the period 2014-2022, of which 136 correspond to Sentinel-1A satellite and 149 Sentinel-1B, providing Earth's surface observations every six days (Figure 21). All images obtained with a final spatial resolution of 12 meters of pixel size were stacked in Zarr files, a Python library for chunked, compressed, and parallelized N-dimensional arrays (Figure 21), providing a flexible and scalable approach to storing large datasets in a compressed format, allowing for efficient storage and retrieval of data in a cloud-based computing environment.

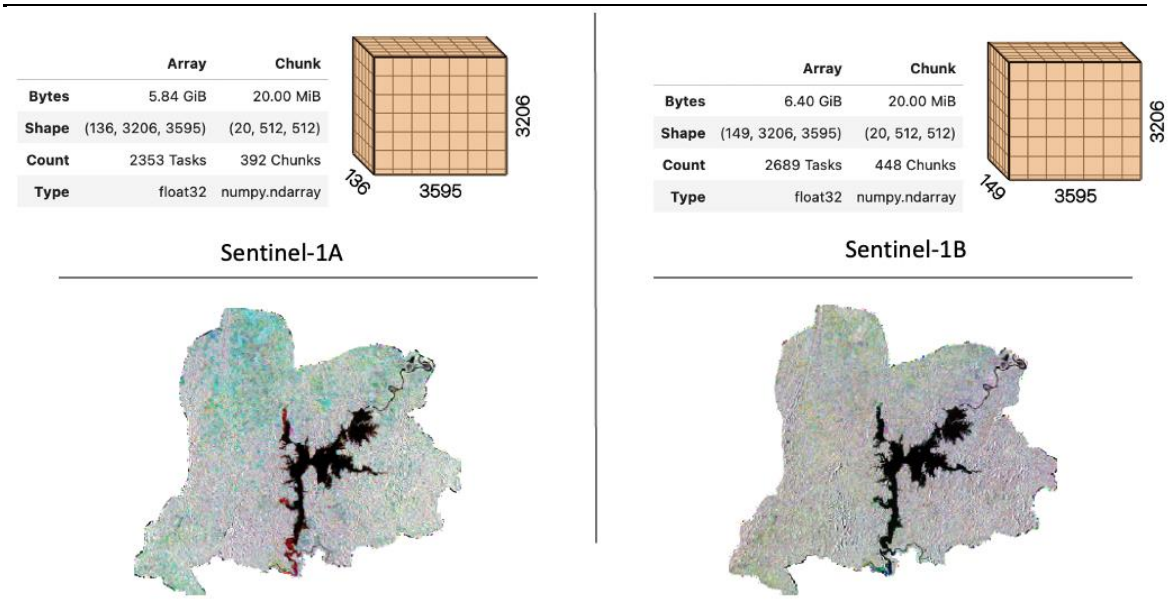


Figure 21. Sentinel-1 radiometric and terrain products obtained from pre-processed imagery product from both satellites (S1A and S1B). Single images were stacked in multidimensional time series array of earth observations.

3.2 Forest and deforestation spatiotemporal dynamics using optical sensors.

Based on the analysis of the forest non-forest map for the year 1996, it was found that the initial baseline for forest areas covered a significant portion of the study area, corresponding to 45,186.9 hectares, which represents 50.7% of the total area. The remaining 49.3% of the area was represented by non-forest areas. However, when examining the dynamics of forest coverage change from 1996 to 2021 (Figure 7), a noticeable decreasing trend in forested areas was observed.

Over the course of the analyzed period, the reported forest area experienced a decline, decreasing from 43,350 hectares in 1996-2000 to 37,763 hectares in 2015-2021. Concurrently, the extent of deforestation increased from 1,837 hectares in 1996-2000 to 2,178 hectares in 2015-2021. In contrast, non-forest areas expanded, growing from 44,004 hectares in 1996-2000 to 49,223 hectares in 2015-2021 (Figure 23).

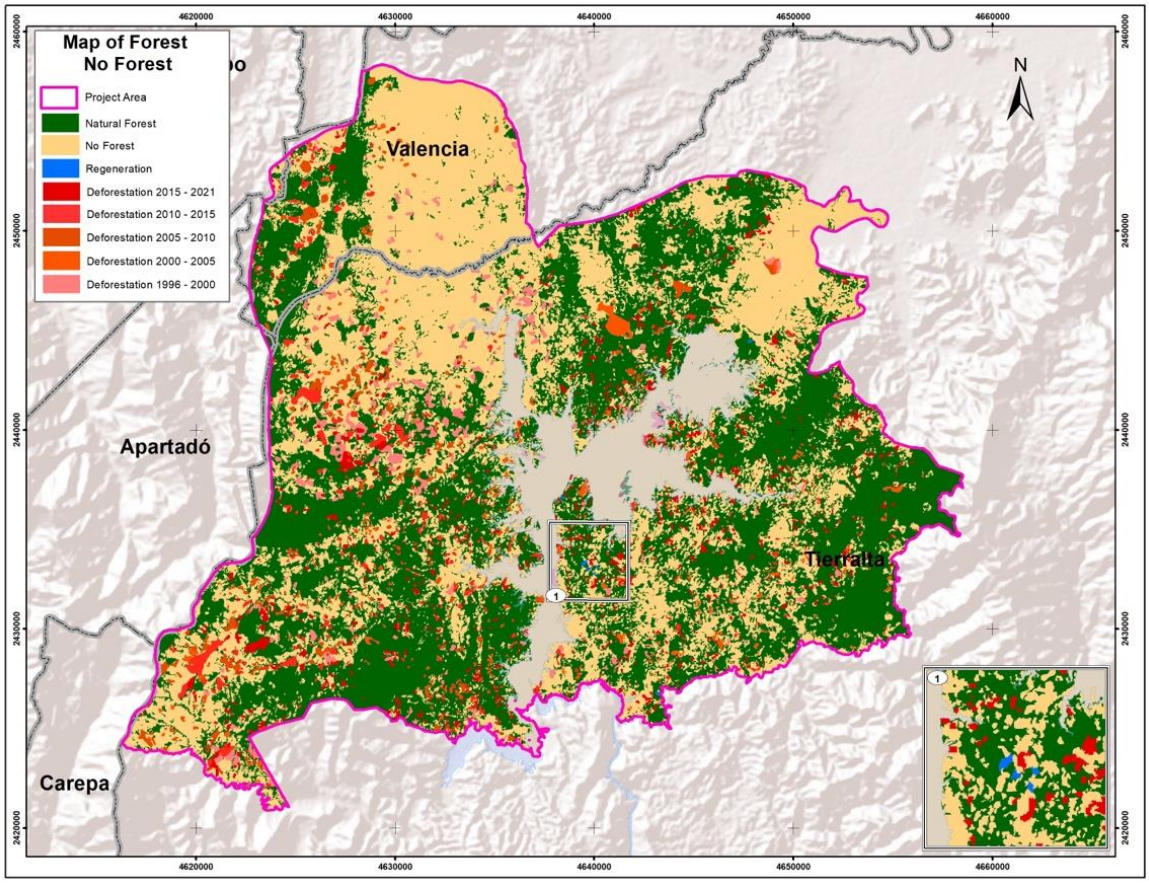


Figure 22. Map of forest, no forest, deforestation, and forest restoration between 1996 and 2021 for the study area. Different red areas represent deforestation events during different periods.

Notably, the analysis also identified reforested and/or revegetated areas, which were first reported in the last two analysis periods. A total of 583 hectares were recorded by the year 2015, and an additional 27 hectares were identified by the year 2021, representing restored forested areas. By the year 2021, forests constituted 42.4% of the study area, covering a total of 37,789 hectares. The remaining 57.63% of the area consisted of non-forest coverages, accounting for 51,401 hectares (Figure 23).

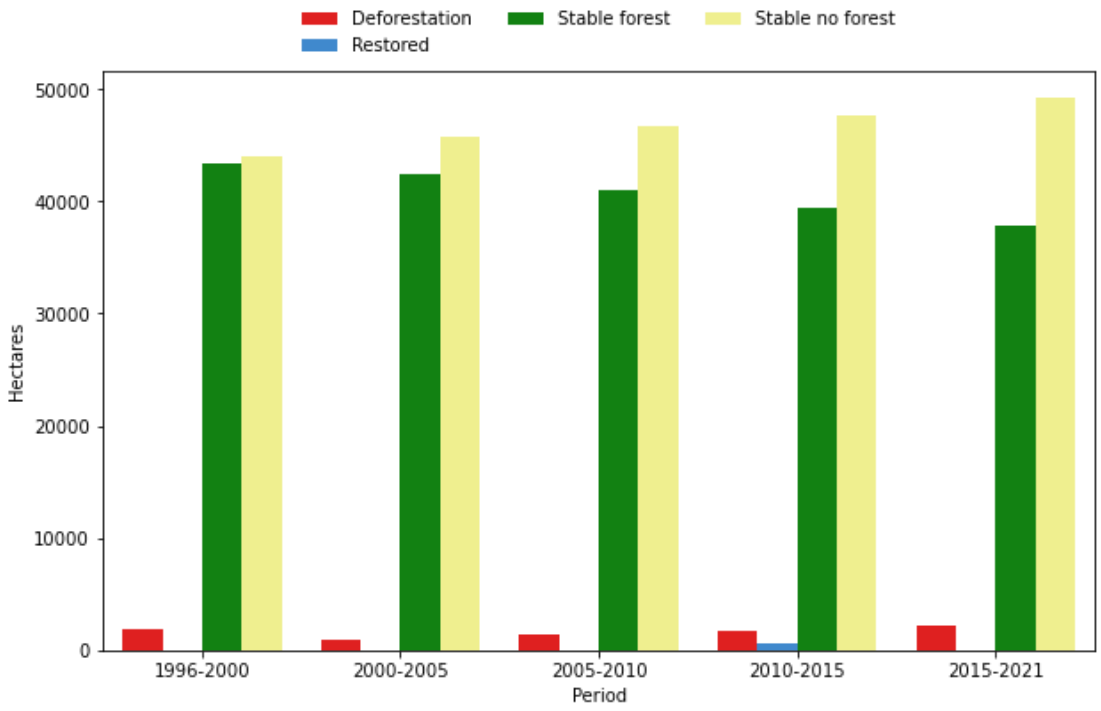


Figure 23. Forest change dynamics estimated extent (hectares) for each of the periods analyzed (1996-2021).

3.3 Validation of forest change dynamics (1991-2021)

Error! Reference source not found. shows the selection sites of the 300 points that served as s amples for the validation of the forest surface change maps for the analyzed study periods. These points were spatially randomly generated for each of the classes: forest, non-forest, and forest loss. No validation points were integrated in the accuracy assessment approach for the reforestation category, as these represented a minimal percentage of the study area, and the areas were concentrated in specific sites that were completely visually confirmed.

The classification accuracy assessment revealed variations in the performance of the map across different time periods. During the period of 1996-2000, the overall accuracy of the forest change map was 89%, user's accuracy for the forest class was 78% of the forest pixels were correctly classified; and producer's accuracy for the forest class was 96%. Conversely, the non-forest class achieved user's accuracy of 100%, and producer's accuracy of 80%. The deforestation class exhibited a user's accuracy of 90% and a producer's accuracy of 95%. In the subsequent time period of 2000-2005, the overall accuracy of the map improved to 96%, compared with 1996-2000 change map. The user's accuracy for the forest class was 93%, indicating a high percentage of correctly classified forest pixels. Additionally, the producer's accuracy for the forest class was 98%, suggesting that the map accurately represented the distribution of forested areas. The deforestation class presented a user's accuracy of 95% and a producer's accuracy of 99%, indicating the map's effectiveness in detecting and representing deforestation events. The non-forest class user's accuracy of 98% and a producer's accuracy of 90%, indicating a slight overestimation of non-forest areas in the map. These results demonstrate the improved accuracy and reliability of the map in capturing land cover changes during this specific period compared to the 1996-2000 period.

Table 9. Accuracy assessment sample for the 1996-2000, 2005-2010, 2010-2015, and 2015-2021 forest/deforestation/non-forest maps for the area of influence of Urra's hydropower. Map categories are rows while the reference categories are columns. Accuracy measures are presented with a 95% confidence interval.

Class	Forest	Def	No Forest	Total	Wi (%)	User's	Producer's	Overall
1996-2000								
Forest	78	4	18	100	0.38	0.78 ± 0.4	0.96 ± 0.3	0.89
Deforestation	3	90	7	100	0.2	0.90 ± 0.02	0.95 ± 0.04	
No Forest	0	0	100	100	0.42	1 ± 0.4	0.8 ± 0.5	
Total	81	94	125	300	1			
2000-2005								
Forest	93	1	6	100	0.48	0.93 ± 0.4	0.98 ± 0.4	0.96
Deforestation	0	95	5	100	0.01	0.95 ± 0.01	0.99 ± 0.01	
No Forest	2	0	98	100	0.51	0.98 ± 0.5	0.9 ± 0.5	
Total	95	96	109	300	1			
2005-2010								
Forest	92	0	8	100	0.46	0.92 ± 0.4	0.99 ± 0.4	0.95
Deforestation	0	99	1	100	0.02	0.99 ± 0.01	0.98 ± 0.02	
No Forest	1	0	97	100	0.52	0.97 ±0.5	0.92 ± 0.5	
Total	93	101	106	300				
2010-2015								
Forest	92	1	7	100	0.45	0.92 ± 0.4	0.97 ± 0.4	0.95
Deforestation	0	95	5	100	0.2	0.95 ± 0.1	0.99 ± 0.02	
No Forest	3	0	97	100	0.53	0.97 ± 0.5	0.89 ± 0.5	
Total	95	96	109	300				
2015-2021								
Forest	92	1	7	100	0.43	0.94 ± 0.4	0.97 ± 0.4	0.96

Deforestation	0	95	5	100	0.2	0.94 ± 0.02	1 ± 0.02
No Forest	3	0	97	100	0.55	0.97 ± 0.5	0.89 ± 0.5
Total	95	96	109	300			

For the period 2005-2010, the overall accuracy of the map remained consistently high at 95%. The user's accuracy for the forest class was 92%, indicating a high level of agreement between the map classification and the actual forested areas. The producer's accuracy for the forest class was 99%, suggesting that the map accurately represented the distribution of forest pixels. The deforestation class user's accuracy was 99% and a producer's accuracy of 98%, indicating the map's effectiveness in detecting and representing deforestation events with a high level of precision. The non-forest class showed a user's accuracy of 97% and a producer's accuracy of 92%, suggesting a slight overestimation of non-forest areas in the map. These results highlight the accuracy and reliability of the map in capturing land cover changes during this specific time period.

From 2010 to 2015, the overall accuracy of the map remained consistently high at 95%. The user's accuracy for the forest class was 92%, indicating a high percentage of correctly classified forested pixels. The producer's accuracy for the forest class was 97%, indicating that the map accurately represented the distribution of forested areas. The deforestation class achieved a user's accuracy of 95% and a producer's accuracy of 99%, demonstrating the map's effectiveness in identifying and representing deforestation events. The non-forest class demonstrated a user's accuracy of 97% and a producer's accuracy of 89%, suggesting a slight underestimation of non-forest areas in the map.

In the most recent time period analyzed, 2015-2021, the overall accuracy of the map remained consistently high at 94.7%. The user's accuracy for the forest class was 92%, indicating a high percentage of correctly classified forested pixels. The producer's accuracy for the forest class was 97%, suggesting that the map accurately represented the distribution of forested areas. The deforestation class achieved a user's accuracy of 95% and a producer's accuracy of 100%, demonstrating the map's effectiveness in detecting and representing deforestation events. The non-forest class showed a user's accuracy of 97% and a producer's accuracy of 89%, indicating a slight underestimation of non-forest areas in the map.

3.4 Restoration implementation assessment using Sentinel-1.

Previous analysis of the forest/non-forest dynamics analysis, showed that three of the implementation plots were found reach a forest succession completed state, indicating that, as of the analysis cutoff date in March 2022, these specific plots exhibited forest coverage in accordance with the definition provided by the SMByC.

Based on time series analysis performed utilizing VH polarization from Sentinel-1, cumulative distributions derived for each of the 270 implementation areas were contrasted to forest and pasture reference data (Figure 24). The two-sample Kolgomorov-Smirnov test revealed that 16 plots (5.93%) demonstrated an incipient restoration state, as evidenced by the vegetation structure estimates that displayed a greater resemblance to reference grassland areas compared to forests (Figure 25, providing an incipient example). Among the intervened plots subjected to evaluation (totaling 224), a significant majority of 82.96% (Table 10) were found to be in an intermediate state of the restoration process. These plots did not exhibit significant differences

when compared to the two reference coverages: grasslands and forests (Figure 10, exemplifying the intermediate state). Furthermore, 27 of the evaluated plots, accounting for 10% of the total (as outlined in Table 10), displayed significant differences when compared to the grassland reference data, but did not exhibit significant differences in comparison to forests. Consequently, these plots were classified as being in an advanced state of reforestation (Figure 25).

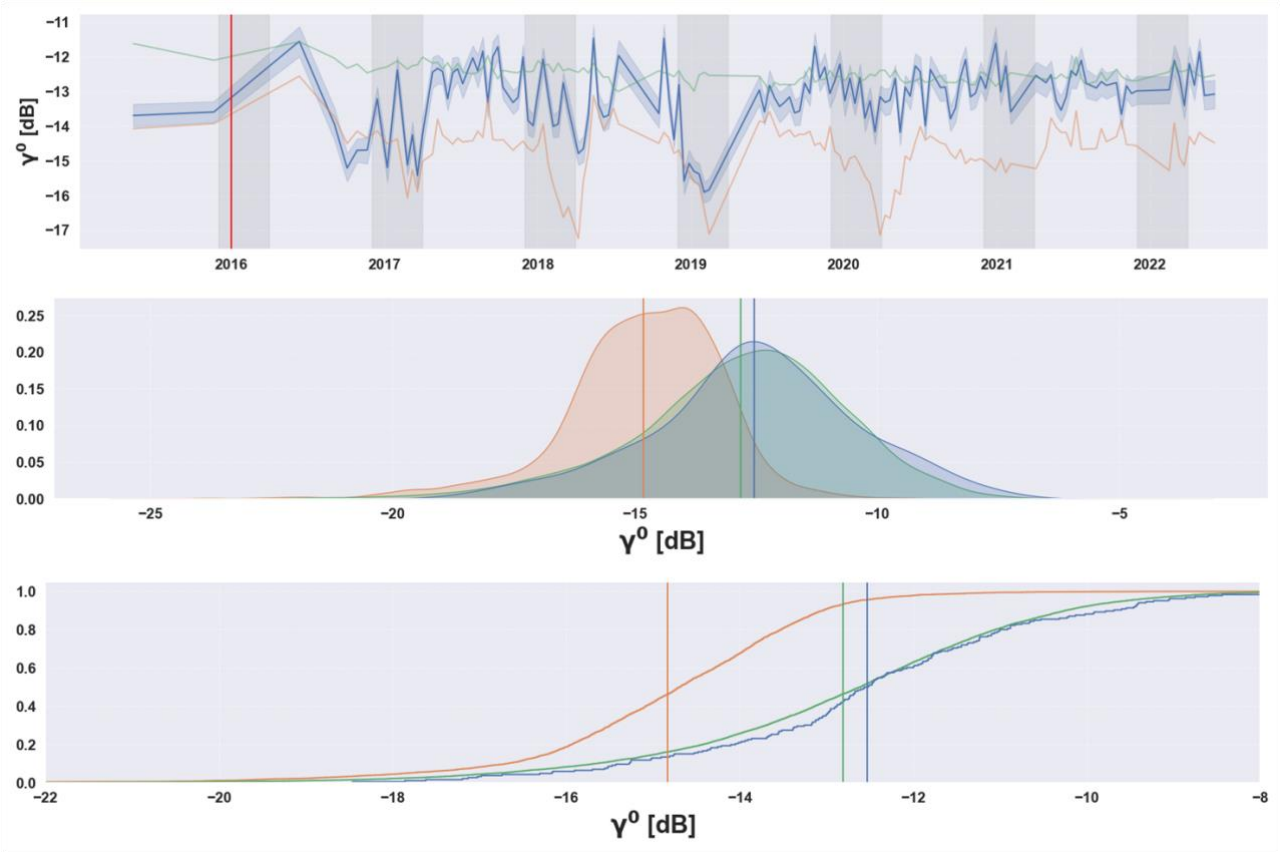


Figure 24. Synthetic Aperture Radar (SAR) time series mean backscatter values (top) and cumulative distributions (middle and bottom) associated to reference areas, forest (green) and pasture (orange); and assessed implementation plot (blue) classified as advanced. Implementation of restoration date (red) and dry season period (December, January, February, and March) (grey) are included.

When the number and state of restoration were inspected related to the implementation years, some years showed a higher number of implementation plots in advanced state of restoration. The years 2018, 2016, 2013, and 2014 demonstrated a substantial presence of 7, 5, 5, and 4 plots in the advanced state, respectively. Additionally, the year 2008 showed 2 plots in the advanced implementation state, while the years 2009, 2012, 2015, and 2017 each had one plot in this state (supplementary material Figure 1 and Table 1).

Table 10. Results of restoration state assessment for 2022 dry season for the 270 implementation plots evaluated based on Sentinel-1 VH backscatter.

Regeneration state	Number of implementations
Incipient	16

Intermediate	224
Advanced	27
Completed	3

No implementation plots from the years 2004, 2005, 2006, 2007, and 2010 were observed to achieve an advanced state in the restoration process. Instead, most plots from these years exhibited intermediate states within the reforestation process, indicating ongoing progress towards full restoration. To explore the relationship between the restoration state and the implementation year, a chi-square test was conducted. The results of the test indicated that there is no statistically significant association between the restoration state and the implementation year (Chi-square statistic: 44.2823; degrees of freedom: 39; and p-value: 2.5860e-01), suggesting that the restoration state is independent of the specific year of implementation.

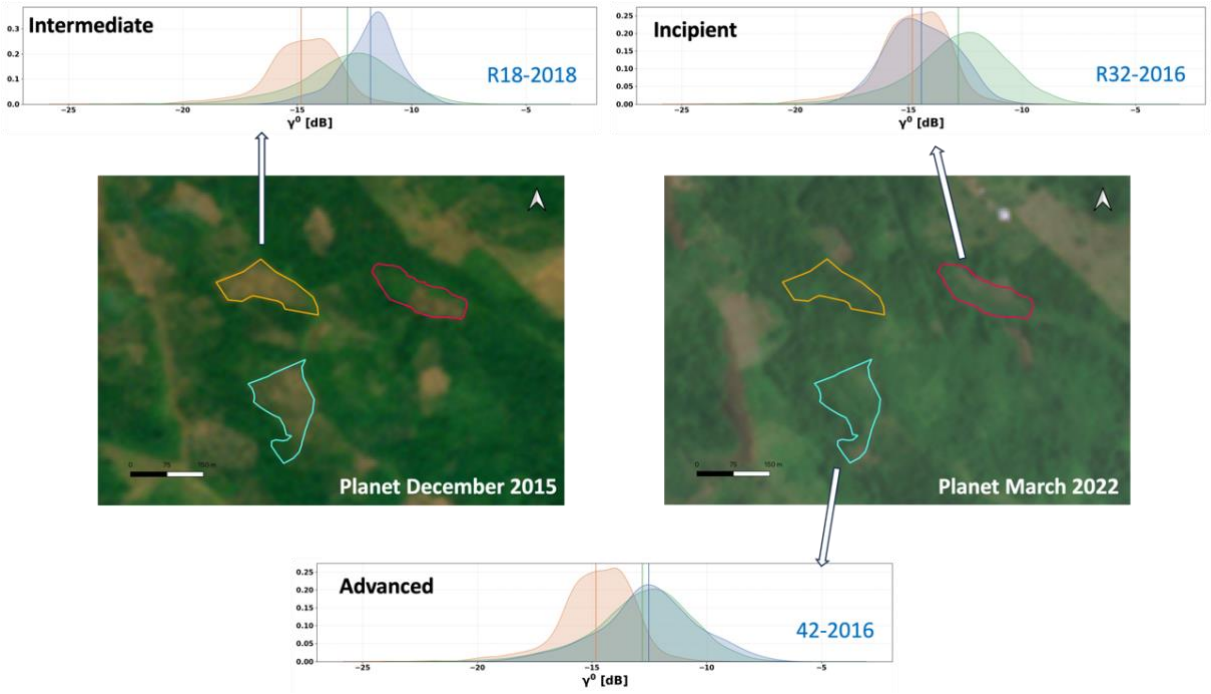


Figure 25. Assessment of three implementation plots based on backscatter cumulative distributions and reference data for the 2022 dry season. The 42-2016 plot (cyan polygon and blue distribution) exhibited similarity to forest reference (green distribution) rather than pasture reference (orange distribution). The R32-2016 plot (red polygon) underwent an incipient implementation assessment, considering its similarity to pasture reference (orange distribution) and dissimilarity to forest distribution (green distribution). The R32-2016 plot (orange polygon) indicated an intermediate restoration state due to differences in distribution (blue) compared to forest (green distribution) and pasture (orange distribution).

4. Discussion

4.1 Forest and deforestation spatiotemporal dynamics using optical sensors.

The results demonstrate that the proposed method allows for optimal identification of forest loss in the study area. The method developed enables the detection of changes occurring within forest coverage, providing a quantitative measure and spatial location of the observed change events. This is achieved through the creation of composite images that eliminate areas without information and extract pixels for generating information-rich images. The forest and

deforestation spatiotemporal dynamics analysis from 1996 to 2021 reveals a deforestation trend in forest change. The initial baseline in 1996 showed that forests covered a significant portion of the study area, accounting for 50.7% of the total area. However, over the period analyzed, there was a noticeable decrease in forested areas, with a corresponding increase in non-forest areas. Deforestation played a significant role in this trend, as the forest area declined from 43,350 hectares in 1996-2000 to 37,763 hectares in 2015-2021, while non-forest areas expanded from 44,004 hectares to 49,223 hectares during the same period. By 2021, forests constituted 42.4% of the study area, resulting in a loss of approximately 7,397 hectares, which corresponds to 8.3% of the total area. Deforestation has primarily concentrated in the western part of the study area. The effects of deforestation are particularly evident near the rural areas located far from the water reservoir. This indicates that there are actors and actions (especially related to land use) that influence beyond the scope of the hydropower reservoir. Consequently, the ten closest most affected rural areas in terms of forest loss area Nain, Murmullo Alto, Maria de Jesus, Casajales, Kilometro 14, Zumbona, Chispas, La Osa, Alto Quimary, and Angostura (Figure 22). Evidence of forest recovery is observed in areas that are not associated with or adjacent to the intervention plots carried out by the company. This finding suggests the presence of ecological processes, e.g., natural succession, that are promoting restoration. It is important to note that although the largest areas of forest recovery were detected in the year 2015, this does not necessarily imply that the recovery occurred specifically between 2010 and 2015. The visual inspection, which identified pastures in 1996, indicates that the recovery could have started before the previous period.

Accuracy assessment of the forest change map revealed variations in its performance across different time periods. The overall accuracy of the map consistently remained high for most of the periods, except the 1996-2000 with lowest overall, user's and producers' accuracies. This is possibly since the Landsat 7 mission have low frequency of revisits compared to later missions, making it difficult to generate cloud-free mosaics, as was the case of the 1996 mosaic, for which there were areas where no cloud-free pixels were obtained in 1996 and 2000 mosaics. Furthermore, since the first Landsat missions in 1972, several enhancements have been made to improve spatial, radiometric, and spectral resolutions. These improvements can have varying effects on the final products, such as mosaics, which in turn can influence the accuracy assessment. The user's accuracy for the forest class also remained consistently high, ranging from 78% to 92%, indicating a high percentage of correctly classified forested pixels. The producer's accuracy for the forest class ranged from 96% to 100%, suggesting that the map accurately represented the distribution of forested areas. Notably, the map showed a high level of effectiveness in detecting and representing deforestation events, with user's accuracies ranging from 90% to 99% and producer's accuracies ranging from 95% to 100% for the deforestation class. However, there were slight discrepancies in the representation of non-forest areas, with user's accuracies ranging from 97% to 100% and producer's accuracies ranging from 89% to 92%. These results demonstrate the overall high accuracy and reliability of the map in capturing land cover changes, particularly in detecting deforestation events. Nevertheless, there is room for improvement in accurately representing non-forest areas. Overall, the findings highlight the importance of regularly updating and refining the classification methods to enhance the accuracy and precision of the forest change map.

The analysis of forest/non-forest dynamics and time series analysis of the implementation plots revealed that three implementation plots reached a completed state, indicating successful forest restoration according to the provided definition. Furthermore, a small percentage (5.93%) of plots demonstrated an incipient restoration state, resembling grassland areas more than forests. Most evaluated plots (82.96%) were in an intermediate state, showing no significant differences compared to both grassland and forest reference data. Additionally, 10% of the plots displayed significant differences compared to grasslands but not forests, indicating an advanced state of reforestation.

4.2 Restoration implementation assessment using Sentinel-1.

Examining the implementation years, certain years showed a higher number of plots in the advanced growth state, such as 2018, 2016, 2014, and 2013. However, there was no statistically significant association between the restoration state and the implementation year, suggesting that the restoration status is not associated to the implementation year. Integration of additional variables associated to the type implementations, active or passive, species composition, among others, that in conjunction with continuous vegetation structure time series provided by SAR EO could generate more insights for the evaluation of restoration process.

To effectively evaluate the success and impact of restoration strategies, it is crucial to combine on-the-ground data with remote sensing observations. Integrating these two sources of information allows for a comprehensive assessment of the restoration process, including the identification of factors that contribute to successful restoration and the identification of areas that may require further intervention. Implementing monitoring systems, which integrates ground truth data and EO data enables a more robust and accurate evaluation of restoration efforts. By continuously assessing the progress and outcomes of restoration strategies, decision-makers and stakeholders can make informed decisions regarding resource allocation, adaptive management, and the refinement of restoration approaches.

5. Conclusions

The proposed method for detecting and quantifying deforestation events using cloud-based and open-source science platforms proved to be effective in monitoring forest dynamics in the study area. The method allowed for the creation of composite images that facilitated the extraction of information-rich images, enabling the detection of spatially and quantitatively significant changes in forest coverage. The impacts of deforestation were particularly evident near rural areas, emphasizing the influence of actors and actions beyond the scope of the hydropower reservoir. However, the study also highlighted areas where forest recovery was observed, suggesting the presence of ecological processes promoting restoration.

The accuracy assessment of the forest change map demonstrated its high performance in detecting and representing deforestation events, with consistently high user's and producer's accuracies. While the accuracy for the forest class remained high, there were slight discrepancies in representing non-forest areas, indicating room for improvement. Regular updates and refinement of classification methods are crucial to enhance the accuracy and precision of the forest change map.

The analysis of implementation plots revealed that currently most of the implementation plots are in an intermediate state. Certain years showed a higher number of plots in an advanced state of reforestation, but no statistically significant association was found between the restoration state and the implementation year. Further integration of additional variables, such as implementation type and species composition, along with continuous time series data from SAR, could provide deeper insights into the evaluation of the restoration process.

To effectively evaluate the success and impact of restoration strategies, combining on-the-ground data with remote sensing observations is crucial. This integrated approach allows for a comprehensive assessment of the restoration process, aiding in the identification of factors contributing to successful restoration and areas requiring further intervention. By continuously monitoring and evaluating restoration efforts, decision-makers and stakeholders can make informed decisions regarding resource allocation, adaptive management, and the refinement of restoration approaches.

6. References

56. Anaya, A., Guti, H., Pacheco-pascagaza, A.M., n.d. Drivers of Forest Loss in a Megadiverse Hotspot on the Pacific Coast of Colombia 1–16.
57. Anaya, J.A., Rodríguez-Buriticá, S., Londoño, M.C., 2023. Clasificación de cobertura vegetal con resolución espacial de 10 metros en bosques del Caribe colombiano basado en misiones Sentinel 1 y 2. Revista de Teledetección 29–41. https://doi.org/10.4995/raet.2023.17655
58. Asner, G.P., 2009. Automated mapping of tropical deforestation and forest degradation: CLASlite. Journal of Applied Remote Sensing 3, 033543. https://doi.org/10.1117/1.3223675
59. Ayala, C.F., Suárez, L.M.E., Rojas, S., Aponte, C., 2015. de Colombia Identification and mapping of Colombian inland wetlands 1–22.
60. Baranwal, A., Bagwe, B.R., M, V., 2019. Machine Learning in Python 12, 128–154. https://doi.org/10.4018/978-1-5225-9902-9.ch008
61. Cabrera, E., Vargas, D.M., Galindo, G., García, M.C., Ordoñez, M.F., 2011. Memoria técnica de la cuantificación de la deforestación histórica nacional – escalas gruesa y fina.
62. DeFries, R., Achard, F., Brown, S., Herold, M., Murdiyarso, D., Schlamadinger, B., de Souza, C., 2007. Earth observations for estimating greenhouse gas emissions from deforestation in developing countries. Environmental Science and Policy 10, 385–394. https://doi.org/10.1016/j.envsci.2007.01.010
63. Deng, J.S., Wang, K., Deng, Y.H., Qi, G.J., 2008. PCA-based land-use change detection and analysis using multitemporal and multisensor satellite data. International Journal of Remote Sensing 29, 4823–4838. https://doi.org/10.1080/01431160801950162
64. Eriksson, L.E.B., Fransson, J.E.S., Soja, M.J., Santoro, M., 2012. Backscatter signatures of wind-thrown forest in satellite SAR images, in: 2012 IEEE International Geoscience and Remote Sensing Symposium. IEEE, Munich, Germany, pp. 6435–6438. https://doi.org/10.1109/IGARSS.2012.6352732
65. Estupinan-Suarez, L.M., Florez-Ayala, C., Quinones, M.J., Pacheco, A.M., Santos, A.C., 2015. Detection and characterizacion of Colombian wetlands using Alos Palsar and MODIS imagery. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci. XL-7/W3, 375–382. https://doi.org/10.5194/isprsarchives-XL-7-W3-375-2015
66. Flores-Anderson, A.I., Herndon, K.E., Thapa, R.B., Cherrington, E., 2019. SAR Handbook: Comprehensive Methodologies for Forest Monitoring and Biomass Estimation. THE SAR HANDBOOK Comprehensive Methodologies for Forest Monitoring and Biomass Estimation 1–307. https://doi.org/10.25966/nr2c-s697
67. Galindo, G., Espejo, O.J., Rubiano, J.C., Vergara, L.K., Cabrera, E., 2014a. Protocolo de procesamiento digital de imágenes para la cuantificación de la deforestacion en Colombia V.2. IDEAM, Instituto de Hidrología, Metereología y Estudios Ambientales 1–225. https://doi.org/10.4324/9781315780245
68. Galindo, G., Espejo, O.J., Rubiano, J.C., Vergara, L.K., Cabrera, E., 2014b. Protocolo de procesamiento digital de imágenes para la cuantificación de la deforestacion en Colombia V.2. IDEAM, Instituto de Hidrología, Metereología y Estudios Ambientales 1–225. https://doi.org/10.4324/9781315780245
69. Hansen, M.C., Stehman, S.V., Potapov, P.V., 2010. Quantification of global gross forest cover loss. Proceedings of the National Academy of Sciences of the United States of America 107, 8650–8655. https://doi.org/10.1073/pnas.0912668107
70. Helder, D.L., Karki, S., Bhatt, R., Micijevic, E., Aaron, D., Jasinski, B., 2012. Radiometric calibration of the jandsat MSS sensor series. IEEE Transactions on Geoscience and Remote Sensing 50, 2380–2399. https://doi.org/10.1109/TGRS.2011.2171351
71. Hoekman, D.H., Quiñones, M.J., 2002. Biophysical forest type characterization in the Colombian Amazon by airborne polarimetric SAR. IEEE Transactions on Geoscience and Remote Sensing 40, 1288–1300.
72. Hoekman, D.H., Quinones, M.J., 2000. Land cover type and biomass classification using AirSAR data for evaluation of monitoring scenarios in the Colombian Amazon. IEEE Trans. Geosci. Remote Sensing 38, 685–696. https://doi.org/10.1109/36.841998
73. Hoekman, D.H., Quinones, M.J., 1997. Land cover type and forest biomass assessment in the Colombian Amazon. IGARSS'97. 1997 IEEE International Geoscience and Remote Sensing Symposium Proceedings. Remote Sensing - A Scientific Vision for Sustainable Development 4.

74. Jafarbiglu, H., Pourreza, A., 2023. Impact of sun-view geometry on canopy spectral reflectance variability. ISPRS Journal of Photogrammetry and Remote Sensing 196, 270–286. https://doi.org/10.1016/j.isprsjprs.2022.12.002
75. Joshi, N., Baumann, M., Ehammer, A., Fensholt, R., Grogan, K., Hostert, P., Jepsen, M.R., Kuemmerle, T., Meyfroidt, P., Mitchard, E.T.A., Reiche, J., Ryan, C.M., Waske, B., 2016. A review of the application of optical and radar remote sensing data fusion to land use mapping and monitoring. Remote Sensing 8, 1–23. https://doi.org/10.3390/rs8010070
76. Joshi, N., Mitchard, E.T.A., Woo, N., Torres, J., Moll-Rocek, J., Ehammer, A., Collins, M., Jepsen, M.R., Fensholt, R., 2015. Mapping dynamics of deforestation and forest degradation in tropical forests using radar satellite data. Environmental Research Letters 10, 34014. https://doi.org/10.1088/1748-9326/10/3/034014
77. Kasischke, E.S., Melack, J.M., Dobson, M.C., 1997. The use of imaging radars for ecological applications - A review. Remote Sensing of Environment 59, 141–156. https://doi.org/10.1016/S0034-4257(96)00148-4
78. Koch, A., Kaplan, J.O., 2022. Tropical forest restoration under future climate change. Nature Climate Change 12, 279–283. https://doi.org/10.1038/s41558-022-01289-6
79. Li, J., Roy, D.P., 2017. A global analysis of Sentinel-2a, Sentinel-2b and Landsat-8 data revisit intervals and implications for terrestrial monitoring. Remote Sensing 9. https://doi.org/10.3390/rs9090902
80. Lopes, A., Touzi, R., Nezry, E., 1990. Adaptive speckle filters and scene heterogeneity. IEEE Trans. Geosci. Remote Sensing 28, 992–1000. https://doi.org/10.1109/36.62623
81. Lucas, R., Rebelo, L.M., Fatoyinbo, L., Rosenqvist, A., Itoh, T., Shimada, M., Simard, M., Souza-Filho, P.W., Thomas, N., Trettin, C., Accad, A., Carreiras, J., Hilarides, L., 2014. Contribution of L-band SAR to systematic global mangrove monitoring. Marine and Freshwater Research 65, 589–603. https://doi.org/10.1071/MF13177
82. M. Susan Moran, 1999. Principles and Applications of Imaging Radar, Manual of Remote Sensing, 3rd Edition, Volume 2. Eos, Transactions American Geophysical Union 80, 67-67–67.
83. Marcela, Q., Martín, V., Ana María, P.-P., Carlos, F., Lina M., E.-S., César, A., Úrsula, J., Claudia, H., Dirk, H., 2016. Un enfoque ecosistémico para el análisis de una serie densa de tiempo de imágenes de radar Alos PALSAR, para el mapeo de zonas inundadas en el territorio continental colombiano. Biota Colombiana 16, 63–84. https://doi.org/10.21068/c2016s01a04
84. Masek, J.G., Wulder, M.A., Markham, B., McCorkel, J., Crawford, C.J., Storey, J., Jenstrom, D.T., 2020. Landsat 9: Empowering open science and applications through continuity. Remote Sensing of Environment 248. https://doi.org/10.1016/j.rse.2020.111968
85. Mitchell, A.L., Rosenqvist, A., Mora, B., 2017. Current remote sensing approaches to monitoring forest degradation in support of countries measurement, reporting and verification (MRV) systems for REDD+. Carbon Balance and Management 12. https://doi.org/10.1186/s13021-017-0078-9
86. Murillo-Sandoval, P.J., Kilbride, J., Tellman, E., Wrathall, D., Van Den Hoek, J., Kennedy, R.E., 2023. The post-conflict expansion of coca farming and illicit cattle ranching in Colombia. Nature Publishing Group UK. https://doi.org/10.1038/s41598-023-28918-0
87. NICFI, 2021. Norway's international climate and forest initiative (NICFI). https://www.
88. nicfi.no/
89. Nielsen, A. a., Conradsen, Knut., Simpson, J., James, 1998. Multivariate Alteration Detection (MAD) in Multispectral, Bi-temporal Image Data: A new approach to change detection studies. Technical Report 1, 1–28.
90. Olofsson, P., Foody, G.M., Stehman, S.V., Woodcock, C.E., 2013. Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation. Remote Sensing of Environment 129, 122–131. https://doi.org/10.1016/j.rse.2012.10.031
91. Onoda, M., Young, O.R., 2017. Satellite earth observations and their impact on society and policy. https://doi.org/10.1007/978-981-10-3713-9
92. Pal, M., 1996. Random forests for land cover classification Mahesh Pal Department of civil engineering National Institute of technology , Kurukshetra. Symposium A Quarterly Journal In Modern Foreign Literatures 3510–3512.

93. Palomino-Ángel, S., Anaya-Acevedo, J.A., Simard, M., Liao, T.-H., Jaramillo, F., 2019. Analysis of Floodplain Dynamics in the Atrato River Colombia Using SAR Interferometry. Water 11, 875. https://doi.org/10.3390/w11050875
94. Pedraza, C., Clerici, N., Forero, C.F., Melo, A., Navarrete, D., Lizcano, D., Zuluaga, A.F., Delgado, J., Galindo, G., 2018. Zero deforestation agreement assessment at farm level in Colombia using ALOS PALSAR. Remote Sensing 10, 1–18. https://doi.org/10.3390/rs10091464
95. Persaud, H., Cabrera, I., 2021. Eficiencia de las imágenes de radar para el monitoreo a tiempo casi real de bosques tropicales en Guyana. Scielo.Org.Pe 28, 577–592.
96. Reiche, J., Lucas, R., Mitchell, A.L., Verbesselt, J., Hoekman, D.H., Haarpaintner, J., Kellndorfer, J.M., Rosenqvist, A., Lehmann, E.A., Woodcock, C.E., Seifert, F.M., Herold, M., 2016. Combining satellite data for better tropical forest monitoring. Nature Climate Change 6, 120–122. https://doi.org/10.1038/nclimate2919
97. Rodriguez, N., Armenteras, D., Morales, M., Romero, M., 2006. Ecosistemas de los Andes Colombianos. Instituto de Investigación de Recursos Biológicos Alexander von Humboldt, Bogotá, Colombia.
98. Rüetschi, M., Small, D., Waser, L., 2019. Rapid Detection of Windthrows Using Sentinel-1 C-Band SAR Data. Remote Sensing 11, 115. https://doi.org/10.3390/rs11020115
99. Ruiz-Ramos, J., Marino, A., Boardman, C., Suarez, J., 2020. Continuous forest monitoring using cumulative sums of sentinel-1 timeseries. Remote Sensing 12. https://doi.org/10.3390/RS12183061
100. Tatsumi, K., Yamashiki, Y., Canales Torres, M.A., Taipe, C.L.R., 2015. Crop classification of upland fields using Random forest of time-series Landsat 7 ETM+ data. Computers and Electronics in Agriculture 115, 171–179. https://doi.org/10.1016/j.compag.2015.05.001
101. Teluguntla, P., Thenkabail, P., Oliphant, A., Xiong, J., Gumma, M.K., Congalton, R.G., Yadav, K., Huete, A., 2018. A 30-m landsat-derived cropland extent product of Australia and China using random forest machine learning algorithm on Google Earth Engine cloud computing platform. ISPRS Journal of Photogrammetry and Remote Sensing 144, 325–340. https://doi.org/10.1016/j.isprsjprs.2018.07.017
102. Ustin, S.L., Middleton, E.M., 2021. Current and near-term advances in Earth observation for ecological applications. Ecological Processes 10, 1–57. https://doi.org/10.1186/s13717-020-00255-4
103. van Oostende, M., Hieronymi, M., Krasemann, H., Baschek, B., Röttgers, R., 2022. Correction of inter-mission inconsistencies in merged ocean colour satellite data. Frontiers in Remote Sensing 3, 1–17. https://doi.org/10.3389/frsen.2022.882418
104. Wang, J., Zhao, Y., Li, C., Yu, L., Liu, D., Gong, P., 2015. Mapping global land cover in 2001 and 2010 with spatial-temporal consistency at 250m resolution. ISPRS Journal of Photogrammetry and Remote Sensing 103, 38–47. https://doi.org/10.1016/j.isprsjprs.2014.03.007
105. Werner, C., Wegmüller, U., Strozzi, T., Wiesmann, A., 2000. GAMMA SAR and interferometric processing software. European Space Agency, (Special Publication) ESA SP 211–219.

Chapter 5: Conclusions and recommendations

Based on the conclusions from the respective chapters, it is recommended to utilize the following methods and approaches for robust, cost-efficient, transparent, and high-quality monitoring of conservation and restoration initiatives:

SAR systems for forest monitoring

Regarding, the application of SAR-based systems has shown promise in quantifying forest extent and change in any atmospheric conditions. Further research should focus on improving geometric correction procedures to reduce errors related to topography features. Radar dense time series can enhance change analysis when combined with ground truth data, allowing for calibration of radar sensors, and reducing classification uncertainties. Additional research is needed to generate robust and accurate forest products at local scales. Exploring the multiple operation modes and processing approaches of ALOS PALSAR and other SAR sensors will provide a better understanding of their potential contribution to local-scale forest monitoring.

Integration of cloud-based solutions and EO data

Our results highlight the effectiveness of open-source and cloud-based software platforms like SEPPO and Nebari for deforestation monitoring. Leveraging high-density Earth observation (EO) time series data enables better analysis and detection of deforestation patterns. The integration of tools with high-density EO time series data improves the accuracy and comprehensiveness of deforestation detection. The cloud-based nature of the software allows for scalability and flexibility in handling large datasets, enabling timely analysis and reporting of deforestation patterns. It is important to consider additional data sources and analysis techniques alongside C-band radar data to complement observations and overcome limitations related to backscatter saturation.

The integration of cloud-based and open-source science routines proves to be an effective approach monitoring forest dynamic, that can be scalable to other applications like in this case with restoration implementations assessment. Composite images created through this method facilitate the extraction of information-rich images, enabling the detection of significant changes in forest coverage. Regular updates and refinement of classification methods are crucial to improve the accuracy and precision of forest change maps, particularly for non-forest areas. Integration of additional variables, such as implementation type and species composition, along with continuous time series data from SAR, provides deeper insights into the evaluation of the restoration process. The combination of on-the-ground data with remote sensing observations allows for a comprehensive assessment of restoration efforts, supporting decision-making, resource allocation, adaptive management, and the refinement of restoration approaches.

By adopting these methods and approaches, stakeholders and decision-makers can obtain reliable and detailed information on conservation and restoration initiatives. This knowledge facilitates informed decision-making and effective management of forest ecosystems, ultimately contributing to the achievement of sustainable environmental goals.

Modern tools for current and future challenges

Modern tools have the potential to significantly improve actions for current and future environmental challenges by offering adaptability and the ability to evolve and interoperate. Firstly, advanced remote sensing technologies and data analysis techniques provide valuable insights into the state of the environment at various scales. High-resolution satellite imagery, LiDAR data, and hyperspectral sensors can capture detailed information about land cover, vegetation health, and ecosystem dynamics. These tools enable more accurate monitoring of deforestation, habitat loss, and biodiversity changes, facilitating targeted conservation efforts and evidence-based decision-making.

Secondly, the integration of geospatial data with other sources, such as climate models, socio-economic data, and citizen science information, allows for a holistic understanding of environmental challenges. Geographic Information Systems (GIS) and data platforms provide the infrastructure to collect, store, analyze, and visualize diverse datasets, enabling interdisciplinary collaboration and the identification of complex environmental interactions. By integrating these tools, policymakers and researchers can gain a comprehensive understanding of the interconnected factors influencing environmental issues, leading to more effective strategies for mitigation and adaptation.

Thirdly, modern tools promote data interoperability and collaboration among stakeholders. Open data standards, open-source software, and cloud-based platforms facilitate data sharing, integration, and the development of innovative solutions. This interoperability allows for the seamless exchange of information and knowledge between different organizations, disciplines, and sectors. By leveraging collective intelligence and pooling resources, stakeholders can address environmental challenges more efficiently and make informed decisions based on a wider range of perspectives and expertise.