CAPSTONE PROJECT PROPOSAL

Michelle MARCHESINI VANEGAS
Earlham College
mmarch22@earlham.edu

Abstract – The Amazon Jungle or Amazonia, is a tropical moist forest located in the Amazon River basin. The Amazon rain forest covers 6.7 million square kilometers spanning Brazil, Bolivia, Peru, Ecuador, Colombia, Venezuela, Guyana, Suriname, and French Guiana. It is the largest area covered in trees in the world, and home to 10% of known species on Earth, making it one of the most important places for wildlife. However, 400 billion trees are under severe threat due to cattling, and logging, and illegal crops. In Colombia, the armed conflict between government forces, the FARC guerrilla, paramilitary groups, and other actors drove illicit activities such as coca cultivation and illegal logging—deep into remote Amazon areas, exacerbating deforestation. This project aims to assess the effectiveness of one of the components of the Colombian Peace Agreement (2016) to reduce the rates deforestation in the Colombian Amazonia by analyzing satellite imagery.

1 Introduction

Biodiversity refers to the different kinds of life coexisting or cohabiting in one area, including animals, plants, bacteria, and fungi. Each one of these actors contribute to their ecosystem in different ways and work together to support our essentials such as clean water, food, shelter and even medicine. The Amazon rain forest is one of the most important places in the world for biodiversity; despite covering only around 1% of the planet's surface, the Amazon rain forest is home to 10% of all the wildlife species we know [1].

The Colombian Armed Conflict, has exacerbated the illegal logging of millions of hectares in Colombia. Beginning in the 1980s, booming U.S. demand sparked successive "cocaine bonanzas," as remote-area farmers cleared hundreds of hectares of native forest to grow coca and reap unprecedented profits. The resulting deforestation has created vast, fragmented patches of degraded jungle throughout the Amazon basin. Coca (*Erythroxylum coca*) is a shrub native to South America. In their natural form, coca leaves contain mild alkaloids that provide gentle stimulation, aid digestion, and alleviate altitude sickness without producing any psychoactive effect. However, organized crime groups process these same leaves into cocaine, a potent and addictive stimulant. Because coca plants thrive in warm, humid environments, illicit plantations have proliferated across Colombia. In response, a range of government agencies and NGOs have launched alternative-livelihood programs—promoting sustainable agriculture, agroforestry, and ecotourism—to redirect local incomes away from illegal crops and toward environmentally sound practices that benefit both families and forests.

The first coca crops dedicated to the production of cocaine in Colombia were detected in 1986 [2]. Over the ensuing decades, competition among guerrilla factions (notably FARC), paramilitary groups, and government forces to control coca cultivation zones drove cultivation ever deeper into remote Amazonian forests, where illicit clearings and informal roads accelerated deforestation. With the 2016 Peace Agreement came the promise of dismantling these networks and reclaiming land for sustainable uses, but questions remain about how effectively the agreement has curtailed drug-related forest loss. By comparing satellite-derived deforestation rates before and after 2016, this project will assess whether reductions in coca-driven violence align with measurable improvements in forest conservation. The environmental degradation caused by the armed conflict in Colombia has been a tragic and often forgotten consequence of this protracted confrontation [3]. The presence of illegal armed groups in the lush jungles, fragile coastal areas, and high mountain paramos has affected Colombia's diverse ecosystems and natural resources due to the expansion of illicit crops, illegal mining, and deforestation that have left deep scars on the Colombian landscape [4].

This project uses satellite imagery to monitor and analyze deforestation in the Amazon rainforest. By leveraging high-resolution data from platforms such as PlanetScope, Sentinel-2, and Landsat, I will study the vegetation loss through spectral indices like NDVI and NBR, combined with change detection techniques. The project focuses on identifying patterns and rates of deforestation over time in a defined area of interest within the Colombian Amazon Basin. Special attention is given to overcoming challenges such as cloud cover and data validation. Ultimately, the goal is to support conservation efforts by providing accurate, timely insights into forest degradation and land-use change. By leveraging high-resolution data from platforms such as PlanetScope, Sentinel-2, and Landsat, I will be looking at the vegetation loss through spectral indices like NDVI and NBR, combined with change detection techniques. The project focuses on identifying patterns and rates of deforestation over time in a defined area of interest within the Amazon Basin. Special attention is given to overcoming challenges such as cloud cover and data validation. Ultimately, the goal is to support conservation efforts by providing accurate, timely insights into forest degradation and land-use change.

Growing up near the Amazon instilled in me a profound sense of responsibility for its conservation. My motivation stems from the desire to transition from theoretical knowledge to practical application in a field I hope to dive into. I am eager to utilize the skills I have gained in class to contribute meaningfully to real-world conservation challenges. My goal is for this project to become an open-source tool for deforestation monitoring. I envision a system that not only aids in environmental protection but also serves as an educational resource, empowering and informing local communities, demonstrating how technology can be a powerful ally in the fight against environmental degradation.

2 Survey

2.1 Satellites

Satellite imagery sources vary in spatial resolution, temporal frequency, and sensor type. These characteristics need to be considered in order to understand each of the advantages and limitations each offers.

2.1.1 Landsat Satellite

Landsat, first placed in orbit in 1972, established the U.S. as the world leader in land remote sensing. The Landsat system has contributed significantly to the understanding of the Earth's environment, spawned revolutionary uses of space-based data by the commercial value-added industry, and encouraged a new generation of commercial satellites that provide regional, high-resolution spatial images [5]. Since 1972, NASA has launched nine Landsat satellites with different spectral range bands. Landsat 1–3 were equipped with a Multispectral Scanner (MSS), which recorded data in four spectral bands: two visible and two near-infrared. The next group, Landsat 4–7, carried either the Thematic Mapper (TM) or Enhanced Thematic Mapper (ETM+) sensors, which featured finer spatial resolution. (i.e., pixel size), and increased radiometric resolution (i.e., bit depth) compared to the MSS. This group also had expanded spectral coverage, adding bands in the middle-infrared and thermal-infrared wavelengths. It is worth noting that the middle-infrared is often referred to as the shortwave infrared (SWIR) [6]. Landsat 8 and 9 introduced advanced sensors and spectral capabilities compared to earlier satellites. Landsat 8, has the Operational Land Imager (OLI) and the Thermal Infrared Sensor (TIRS) [7]. Landsat 9's OLI-2 sensor captures 14-bit data (vs. Landsat 8's 12-bit), allowing finer detection of subtle variations in brightness, particularly in dark (e.g., dense forests) and bright (e.g., snow) surfaces. Both satellites share identical 11 spectral bands, spatial resolutions (15–100 m), and swath width (185 km), ensuring seamless integration of data for long-term environmental monitoring[8].

2.1.2 Sentinel Satellite

Optical satellites (like Sentinel-2) capture images using sunlight reflected from Earth. They provide detailed color imagery, but do not do well when there is cloudy weather or darkness. Radar satellites (like Sentinel-1) use microwaves to collect data through clouds, smoke, and even at night. However, their images look less intuitive to the human eye. Combining these two data types, we could obtain the rich spectral details from the optical satellites, the all-weather visibility, and structural information from the radar satellites.

A 2020 study fused Sentinel-1 and Sentinel-2 data to map land cover in Italy. Radar provided cloud-free structural data, while optical data added spectral details—this improved accuracy in identifying crops, forests, and urban areas [9].

In Para, Brazil, PRODES and DETER data sets have been used to monitor the Amazon. They used Sentinel-1 and Sentinel-2 to obtain valuable data despite cloudy conditions, and Landsat-8 for historical comparisons. They trained a CNN on labeled PRODES data, where green, dense textures represent forests, while brown, fragmented patches indicate deforestation. This allowed the model to learn to recognize deforestation patterns, achieved more than 90% precision in segmenting deforestation areas, and detected small-scale illegal logging that manual methods often miss. [10]

2.1.3 Planet Scope

Planet Scope is a satellite imaging system that provides high-resolution continuous views of the Earth using hundreds of Dove satellites in orbit. Planet Scope provide daily, global imagery of the Earth using multi-spectral imagery (Blue, Green, Red and Near-Infrared), and 3-5 meters resolution which allows the user to see trees, roads, and fields[11]. Due to the vast number of satellites in orbit, Planet Scope can image the entire surface of the Earth once per day [11]. One of the benefits of using this data is that PlanetScope can detect small-scale changes within days, such as forest clearing, new burn areas, changes in the water levels and changes in vegetation in fragile ecosystem like wetlands. Its cloud-based access allows easy integration with machine learning, GIS software or APIs.

2.2 Methods

2.2.1 Google Earth Engine (GEE) Workflows

The GEE is a cloud platform allowing users to access satellite imagery from Landsat or Sentinel archives designed for large-scale geospatial analysis. This platform is helpful because it provides algorithms like the NDVI or time-series analysis to be applied to the images to detect changes in the forest cover. Additionally, machine models can be used directly on the GEE to automate deforestation alerts[12].

GEE was used in a study focused on tracking and predicting forests in Brazil. Using satellite imagery from Landsat and Sentinel missions, the study mapped deforestation patterns at five-year intervals and projected forest dynamics up to 2028. A Random Forest classification model was implemented, enabling efficient processing of large datasets in the cloud.

This engine is a great option for reducing processing time with its high computational power, providing access to a vast satellite imagery catalog, including cloud-free images over long time spans, and allowing near real-time monitoring and streamlined integration of machine learning algorithms.

2.2.2 NDVI Differencing: Change detection algorithm

The Normalized Difference Vegetation Index or greenness index measures vegetation's greenness, density, and health using near-infrared and red spectral bands. By comparing the NVDI values of two different dates, the areas of forest loss or regrowth can be analyzed. This index is suitable for estimating vigor throughout the crop cycle based on how plants reflect specific electromagnetic spectrum ranges. This index is helpful to determine how healthy or unhealthy a plant is, based on how it reflects energy and light. Healthy plants are green because their chlorophyll pigments reflect green waves and absorb red waves. Therefore, a healthy plant actively absorbs red light and reflects near-infrared when photosynthesis occurs.

2.3 Challenges and Limitations

Satellite remote sensing has allowed scientist to support natural resource management like never before. However, a significant portion of data are not freely available. This limits the access to information and slows the advancement of monitoring and analysis efforts. Additionally, SRS-based data analysis is expensive due to hardware, software, qualified and trained staff costs. Another challenge is the complexity of integrating SRS-based data to *in situ* data. There is a lack of cooperation between local ecologists and satellite experts which leads to SRS often being underused or undervalued [13].

3 Engineering Design

3.1 Data Collection

I will use Google Earth Engine (GEE) to access public archives of Landsat and Sentinel imagery from 2014 through 2025 over the Colombian Amazon.

3.2 Preprocessing

Within GEE, I will apply built-in routines to mask clouds and shadows automatically and perform basic atmospheric corrections. Once the images are cleaned, I will compute the Normalized Difference Vegetation Index (NDVI) for each date. To reduce noise from stray clouds or seasonal variation, I will aggregate those daily or weekly values into monthly averages providing a single, clear snapshot per month.

3.3 Change Detection

I will detect forest loss using two complementary methods. First, I will perform a simple "threshold differencing" on the monthly NDVI composites: whenever a pixel's greenness drops by more than 0.2 from one month to the next, I will flag it as a potential clearing. Second, I will train a U-Net convolutional neural network in Python, using a set of labeled examples to recognize clear-cut patterns. If the U-Net struggles with very small or subtle clearings, I will fall back on a Random Forest classifier applied to the NDVI-difference stacks.

3.4 Aggregation and Analysis

Once I have produced each month's map of clearings, I will convert the flagged pixels into vector patches and discard any speckles smaller than 0.1 ha. I will then sum the area of all cleared patches in each municipality for every month, producing two time series: one covering January 2014–December 2016 (pre-Agreement) and one covering January 2017–December 2023 (post-Agreement).

3.5 Reporting and Policy Alignment

I will load those time series into a Jupyter notebook or simple web dashboard that overlays key Peace Agreement milestones—such as the November 2016 ratification, so I can directly compare pre and post-agreement deforestation rates. For more robust attribution, I will include nearby control areas unaffected by the Agreement and run a difference-in-differences analysis.

3.6 Risks and Contingencies

- **Cloud cover:** If monthly composites remain noisy, I will expand to quarterly averages or integrate Sentinel-1 radar data as a vegetation proxy.
- **Model accuracy:** If the U-Net overfits, I will simplify its architecture (fewer layers) and rely more on the Random Forest fallback.
- **Policy attribution:** To avoid conflating other land-use changes with peace-related effects, I will compare against matched control regions and time periods.

By building on GEE's data pipelines, Python's machine-learning libraries, and standard vector routines, I will ensure this design is complete, flexible, and accessible.

4 Usage

My project will be a Python-based application designed to assist in the analysis of deforestation by providing a user-friendly interface for satellite imagery. To begin, users must upload a series of GeoTIFF images of the region they will be analyzing from different periods of time. The program then presents a navigable timeline, allowing users to scroll through imagery from different dates to visually track changes over time. The core functionality lies in its ability to apply various filters, such as the Normalized Difference Vegetation Index (NDVI), which transforms the raw satellite data into a clear, color-coded map that allows user to visualize empty patches. This visual output highlights vegetation health, with a shift from green to brown or red patches serving as a primary indicator of deforestation. As development of this project is ongoing, additional features are planned to be implemented as more is learned about the field of conservation and deforestation analysis, ensuring the tool continues to grow in capability. This streamlined process makes the analysis of environmental change more accessible for researchers, conservationists, and students.

5 Evaluation Plan

5.1 Technical Accuracy and Effectiveness

- **Change Detection Validity:** The accuracy of NDVI/NBR-based deforestation detection will be evaluated using time-series visual inspection, and/or cross-validation with official deforestation datasets (IDEAM, Global Forest Watch).
- **Cloud Masking and Preprocessing Quality:** The success of this project will be measured by the project's ability to mitigate cloud cover using filters or multi-temporal composites to preserve consistent spatial coverage.
- **Spatial and Temporal Consistency:** The project will be assessed on how well it identifies deforestation patterns across time and geography, especially post-2016 conflict transitions.

5.2 Project Deliverables

- **Technical Report Quality:** Evaluated based on clarity, depth of analysis, completeness of sections (introduction, literature review, methods, results, references), and incorporation of feedback across drafts.
- **Data Architecture and Graphical Abstracts:** Assessed for logical structure, clarity of pipeline visualization, and consistency with actual implementation.
- **Demonstration Video and Poster:** Judged on communication effectiveness, visual clarity, technical explanation, and audience engagement during the Academic Fair.

5.3 Tool Proficiency and Code Quality

- **Geospatial Tools Usage:** Proper and efficient application of tools such as Google Earth Engine, QGIS, Python (e.g., rasterio, geopandas), and any machine learning libraries.
- **Reproducibility:** Code and documentation on GitLab will be reviewed for clarity, functionality, and the ability to reproduce key outputs with minimal intervention.
- **Data Management:** Evaluation of how well the data is organized, preprocessed, and annotated throughout the project.

5.4 Research Contribution and Reflection

- **Impactful Insights:** The relevance and usefulness of insights produced for conservation or policy-making audiences, especially in relation to the peace agreement's environmental effects.
- **Critical Reflection:** Evaluated through the final report and portfolio, focusing on challenges faced, lessons learned, and potential for future work.

6 Contributions

By creating a replicable pipeline for satellite-based forest monitoring with NDVI and NBR indices, this project will provide a scalable tool for detecting land-use changes in post-conflict areas. The employment of multi-source satellite imagery (PlanetScope, Sentinel-2, and Landsat) with cloud-resilient preprocessing approaches is intended to improve the reliability of remote sensing in tropical regions with persistent cloud cover. In addition to its technical importance, this project promotes environmental justice and conservation by developing visualizations and analyses that may be used to enhance policy conversations and community advocacy activities. The finished tools, code, and visual outputs will be released openly to promote additional research and application in similar contexts.

7 Risks

Satellite remote sensing has allowed scientist to support natural resource management like never before. However, a significant portion of data are not freely available. This limits the access to information and slows the advancement of monitoring and analysis efforts. Additionally, SRS-based data analysis is expensive due to hardware, software, qualified and trained staff costs. Another challenge is the complexity of integrating srs-based data to *in situ* data. There is a lack of cooperation between local ecologists and satellite experts which leads to SRS often being underused or undervalued. [13] Some of the considerations I should take into account for the development of my capstone projects are:

1. **Choosing the right satellite data:** Different satellite or data sources provide different trade-offs such as spatial resolution, revisit time, cost and accessibility.
2. **Cloud Coverage:** This is one of the biggest challenges in tropical regions. Clouds often affect the clarity of

the images, specially during the winter (rainy season in South America). For this, I could try using cloud masking algorithms, dry season data (that could limit the extend to which I can analysis deforestation rates) or use SAR images (i.e Sentinel-1) that can penetrate clouds.

3. **Time Range:** Choosing a specific time range would help me be consistent with dates and seasons, and determine wether I will be monitoring recent events or long-term trends.

4. **Skills:** Some of the skills/tools who were mentioned consistently in the case studies researched, are GIS tools, Remote Sensing Platforms (GEE, SNAP, etc), Python, JavaScript and rasterio, geopandas, earthpy libraries.

7.1 Solutions

8 Further Development

References

[1] WWF. The amazon rainforest. https://www.wwf.org.uk/where-we-work/amazon: :text=Despite Accessed: 30 Aug. 2025.

[2] Sergio Uribe. Evolución de los cultivos de coca en colombia: 1986-2017, Nov 2020.

[3] Deisy Milena Sorzano Rodríguez, Juan Diego Villamizar Escobar, Esthela Galván Vela, and Rafael Ravina Ripoll. Environment and armed conflict: Perspectives of former combatants of the revolutionary armed forces of colombia (farc). *Jurídicas CUC*, 20(1):269–285, 2024.

[4] Francisco Gutiérrez-Sanín. Tough tradeoffs: Coca crops and agrarian alternatives in colombia. *International Journal of Drug Policy*, 89:103156, 2021.

[5] Darrel L Williams, Samuel Goward, and Terry Arvidson. Landsat. *Photogrammetric Engineering & Remote Sensing*, 72(10):1171–1178, 2006.

[6] Nicholas E Young, Ryan S Anderson, Stephen M Chignell, Anthony G Vorster, Rick Lawrence, and Paul H Evangelista. A survival guide to landsat preprocessing. *Ecology*, 98(4):920–932, 2017.

[7] EOS Data Analytics. Landsat 8 satellite: Overview, sensors, and data access, 2024. Accessed 7 Apr. 2025.

[8] NASA. Landsat 9 overview, 2024. Accessed 7 Apr. 2025.

[9] Auravant. Vegetation indices and their interpretation: Ndvi, gndvi, msavi2, ndre, and ndwi, 2021. Accessed: 2025-04-04.

[10] Aldem Bourscheit. Prodes and deter: get to know these strategic systems in the fight against deforestation in the amazon, 2022. Accessed: 2025-04-04.

[11] Planet Labs PBC. Satellite monitoring, 2024. Accessed 7 Apr. 2025.

[12] Maria Antonia Brovelli, Yaru Sun, and Vasil Yordanov. Monitoring forest change in the amazon using multi-temporal remote sensing data and machine learning classification on google earth engine. *ISPRS International Journal of Geo-Information*, 9(10):580, 2020.

[13] Nathalie Pettorelli, William F Laurance, Timothy G O'Brien, Martin Wegmann, Harini Nagendra, and Woody Turner. Satellite remote sensing for applied ecologists: opportunities and challenges. *Journal of Applied Ecology*, 51(4):839–848, 2014.