University of California
Santa Barbara

Informing Forest Conservation Regulations in Paraguay

A Capstone Project submitted in partial satisfaction of the requirements for the degree
of
Master of Environmental Data Science
for the
Bren School of Environmental Science & Management

by
Atahualpa Ayala
Dalila Lara
Alexandria Reed
Guillermo Romero

Committee in charge:
Robert Heilmayr
Naomi Tague
Ruth Oliver

June 2023

PYFOREST
Chaco Conservation

Signature Page

As developers of this Capstone Project documentation, we archive this documentation on the Bren School's website such that the results of our research are available for all to read. Our signatures on the document signify our joint responsibility to fulfill the archiving standards set by the Bren School of Environmental Science & Management.

Atahualpa Ayala Date

Dalila Lara Date

Alexandria Reed Date

Guillermo Romero Date

The Bren School of Environmental Science & Management produces professionals with unrivaled training in environmental science and management who will devote their unique skills to the diagnosis, assessment, mitigation, prevention, and remedy of the environmental problems of today and the future. A guiding principle of the School is that the analysis of environmental problems requires quantitative training in more than one discipline and an awareness of the physical, biological, social, political, and economic consequences that arise from scientific or technological decisions.

The Capstone Project is required of all students in the Master of Environmental Data Science (MEDS) Program. The project is a six-month-long activity in which small groups of students contribute to data science practices, products or analyses that address a challenge or need related to a specific environmental issue. This MEDS Capstone Project Technical Documentation is authored by MEDS students and has been reviewed and approved by:

Robert Heilmayr Date

Ruth Oliver Date

Acknowledgments

We would like to express our gratitude to the individuals and organizations that contributed to the success of this project. Specifically, Christina Goralewski and Antonella Mascheroni with the National Forestry Institute (INFONA). We are also grateful to our Client and faculty advisor, Robert Heilmayer, for their continuous support and insightful feedback. We would also like to extend our thanks to the faculty members and PhD candidates at the University of California, Santa Barbara (UCSB) and staff at the National Center for Ecological Analysis and Synthesis (NCEAS) who provided us with valuable guidance and feedback. Their contributions have played a significant role in enhancing the overall quality and impact of this research.

Abstract

The Paraguayan Chaco is facing significant environmental degradation due to deforestation driven primarily by cattle ranching. Paraguay has implemented laws enforced by the National Forestry Institute (INFONA), requiring landowners to submit land use plans (LUPs). Approved LUPs identify areas authorized for deforestation, define locations of forest reserves, and include additional protective measures. With tensions rising between economic development and forest conservation, policymakers are considering modifying these laws. This project addresses the challenges faced by policymakers through a four-step approach: (1) Assess property owners' compliance with deforestation regulations through the use of geospatial overlays, (2) Estimate protected forest area under different potential laws in the undeveloped Paraguayan Chaco region by creating mock properties with simulated LUPs, (3) Integrate the simulated LUP configurations into a machine learning model to examine how potential changes to land use laws might influence future deforestation patterns, and (4) To inform forest conservation regulations by providing INFONA and other stakeholders with an interactive platform for exploring the results. The study determined that ~83% of properties with LUPs between 2019 and 2020 exhibited land use compliance. Under four potential variations of land use laws, we estimate a simulated range of protected forest between 2.11 and 5.5 million hectares (ha). Our machine learning model uses pixel-wise probabilities to generate a ten-year projection of deforestation trends and an estimated range of 2.10 and 5.2 million ha. The interactive tool provides INFONA with a way to visualize the results and will aid policymakers in making informed decisions regarding future land use policy requirements, implementation, and enforcement.

Table of Contents

Executive Summary

The Gran Chaco is an 86,966,945-hectare (ha) ecoregion extending across Argentina, Brazil, Bolivia, and Paraguay (de et al., 2021). The Gran Chaco contains South America's second-largest forest and is characterized by its diverse landscapes, including arid thorn forests, cacti clusters, and palm savannas that experience seasonal flooding. This ecological diversity is home to a wide array of wildlife, such as 3,400 types of plants, 500 bird species, 150 mammal species, and 220 species of reptiles and amphibians. However, the forest is threatened by agricultural and economic development, primarily for beef and soy production, leading to extensive deforestation and environmental degradation (World Wildlife Fund, 2016). Despite the Gran Chaco extending over several countries, about 23% of it is located in the western region of Paraguay. This portion remarkably accounts for approximately 60% of the entire national territory of Paraguay (Milán & González, 2022).

Paraguay has a variety of laws in place to conserve forest cover in its eastern region. To develop land in the Paraguayan Chaco, cattle ranchers must submit land use plans (LUPs) that comply with forestry laws enforced by Paraguay's National Forestry Institute (INFONA). These plans must include a forest reserve covering 25% of the designated land, a 100-meter hedgerow buffer, a 100-meter riparian forest, and paddocks of less than 100 ha each where deforestation is authorized (Paraguay, 1986; Paraguay, 1995). With tensions rising between the need for economic development and forest conservation, policymakers are considering modifying these laws in response to various stakeholders' needs. Several bills have been submitted to the Senate and the Chamber of Deputies in the past years with that purpose(National Congress, 2009; National Congress, 2017). However, it remains unclear what potential modifications to these laws could mean for forests.

Policymakers are currently facing challenges in understanding how much forest is conserved under the existing laws and the potential impact of modifying these laws. Madam President of INFONA, Cristina Goralewski, has stated "What is not measured cannot be improved, so we will start working on a restoration plan. We want to begin a process of reducing land use change in our country. First, we need to know those data and why they occur" (ABC Color, 2022).

This capstone project focuses on the 13,233,220 ha of the Paraguayan Chaco in the western region of the country, where 88.1% (667,731 ha) of the deforestation between 2017 and 2020 occurred (Instituto Forestal Nacional, 2022). To address this issue, our project aimed to evaluate how changes to the laws governing private LUPs would affect the extent of forest cover and agricultural land in the Paraguayan Chaco. We outlined the following objectives to achieve this goal:

1. Assess the compliance of property owners' execution of their approved LUPs and quantify deforestation rates and forest cover.
2. Estimate the areas designated as protected forests based on current and alternate laws for undeveloped regions.
3. Predict deforestation patterns in the undeveloped region of the Paraguayan Chaco, taking into account potential changes in the laws governing LUPs.

4. Inform forest conservation regulations by providing INFONA and other stakeholders with an interactive platform for exploring the results of our geospatial analysis, land use simulations, and deforestation predictions in the Paraguayan Chaco.

Our approach to achieving our first objective was to utilize geospatial overlays to compare deforestation data against the permitted use of land; this provided us with the information to quantify deforestation rates and compliance with LUPs. To accomplish our second objective, we developed a geospatial simulation that divides the undeveloped Chaco region into mock properties. We applied a set of custom functions for each mock property to generate counterfactual LUP configurations based on various legal scenarios. These counterfactual scenarios represent potential modifications to the current laws governing forest reserves. The simulated LUP configurations were then incorporated into our machine learning model for predicting deforestation patterns (Objective 3). This model integrates historical deforestation data, LUP information, and relevant environmental features to forecast the potential outcomes of different policy scenarios. The integration of the counterfactual LUPs into the ML model allows us to examine how changes to land use laws might influence future deforestation patterns in the undeveloped region of the Paraguayan Chaco. Finally, to make our results accessible to policymakers and the public (Objective 4), we have developed an R Shiny dashboard that provides an interactive platform for exploring the data and visualizations generated from Objectives 1-3. The dashboard is designed to be user-friendly and intuitive, ensuring that stakeholders with varying technical expertise can easily understand the impacts of forest policies on the Paraguayan Chaco.

Our partnership with INFONA holds the potential to significantly impact the future of land use policy requirements, implementation, and enforcement through the use of our results and interactive tool. This tool offers comprehensive data visualization and serves as an inclusive platform that democratizes access to critical information, making it easily accessible to all individuals.

Problem Statement

The Paraguayan Chaco is a 13,233,220-ha ecoregion and the second-largest forest in South America. The Paraguayan Chaco lost approximately 667,731 ha of forest between 2017 and 2020 due to deforestation driven by cattle ranching and cropping. Paraguay has a variety of laws in place to conserve forest cover in this region. To develop land in the Paraguayan Chaco, cattle ranchers must submit LUPs that comply with forestry laws enforced by Paraguay's National Forestry Institute INFONA. These plans must include a forest reserve covering 25% of the designated land, a 100-meter hedgerow buffer, a 100-meter riparian forest, and paddocks of less than 100 ha where deforestation is authorized. With tensions rising between the need for economic development and forest conservation, policymakers are considering modifying these laws in response to various stakeholders' needs. Potential modifications to these laws could lead to an increase or decrease in the required forest reserves.

However, policymakers are currently facing challenges in understanding how much forest will be conserved under the existing laws and the potential impact of modifying these laws. To address this issue, our project aims to estimate how changes to the rules governing private LUPs would affect the extent of forest cover and agricultural land in the Paraguayan Chaco.

Project Objectives

To achieve this goal, we have outlined the following objectives:

1. Assess the compliance of property owners' execution of their approved LUPs and quantify deforestation rates and forest cover.

2. Estimate the areas designated as protected forests based on current and alternate laws for undeveloped regions.

3. Predict deforestation patterns in the undeveloped region of the Paraguayan Chaco, including variations in the laws governing LUPs; and

4. Inform forest conservation regulations by providing INFONA and other stakeholders with an interactive platform for exploring the results of our geospatial analysis, land use simulations, and deforestation predictions in the Paraguayan Chaco.

This project aims to equip the National Forestry Institute (INFONA) with a dynamic and interactive tool to assess LUP compliance statistics and explore current and alternate laws' impacts on the Paraguayan Chaco and future deforestation patterns.

Summary of Solution Design

The solution design for this project addresses the four main objectives outlined in the project scope, focusing on understanding the relationship between LUPs, deforestation, and compliance with forestry laws in the Paraguayan Chaco. The approach leverages various datasets and enables the application of geospatial analysis, machine learning, and data visualization techniques to achieve the desired outcomes

Datasets

This section describes the study boundary and the critical datasets used in the analysis, all provided by INFONA. Table 1 includes the description of all datasets used in this analysis in Appendix A.

Study Boundary:

The boundary for this study was established using the **Political Boundaries** dataset, focusing on the departments of Alto Paraguay, Presidente Hayes, and Boquerón. These departments encompass the western region of Paraguay, an area widely known as El Chaco, which is the dominant biome in this region (Figure 1).

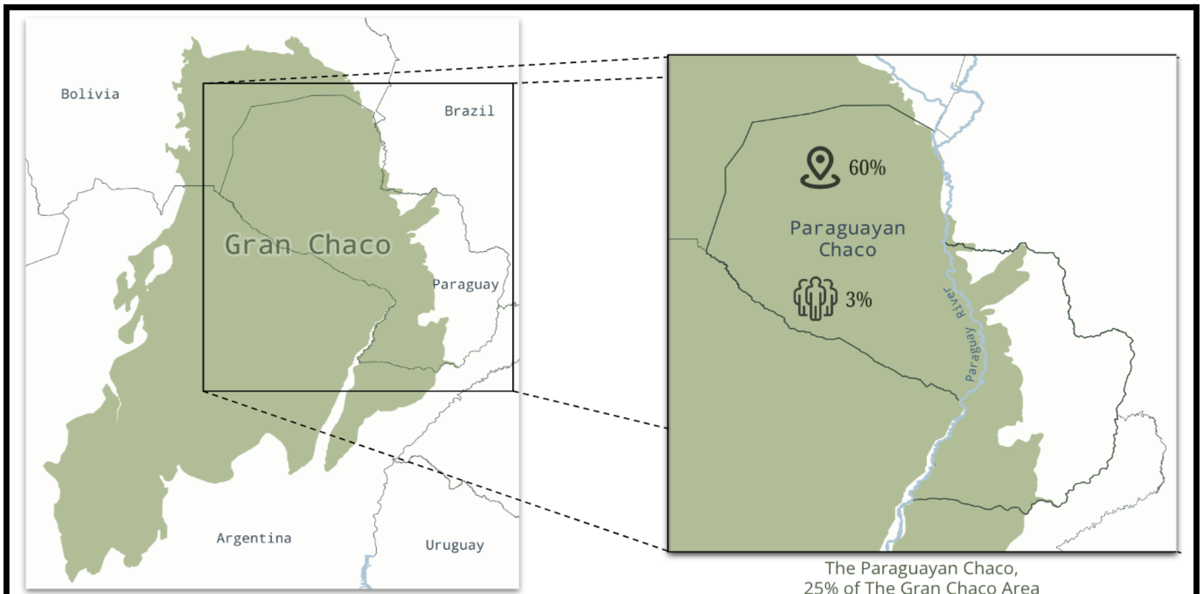


Figure 1: The study boundary of the western region of Paraguay.

Property Boundary:

The **Property Boundary** dataset contains vector geometry outlining the perimeters of 3,612 properties registered with INFONA from 1994 to 2022. Key columns in this dataset include the property's unique identifier, the year of registration, and the date of initial approval.

Land Use Plan:

The **Land Use Plan** dataset corresponds to the property_boundary dataset, with each unique property ID in **Property Boundary** having multiple corresponding rows in **Land Use Plan**. Each row in **Land Use Plan** provides vector geometry that illustrates the approved segmentation of the property by INFONA.

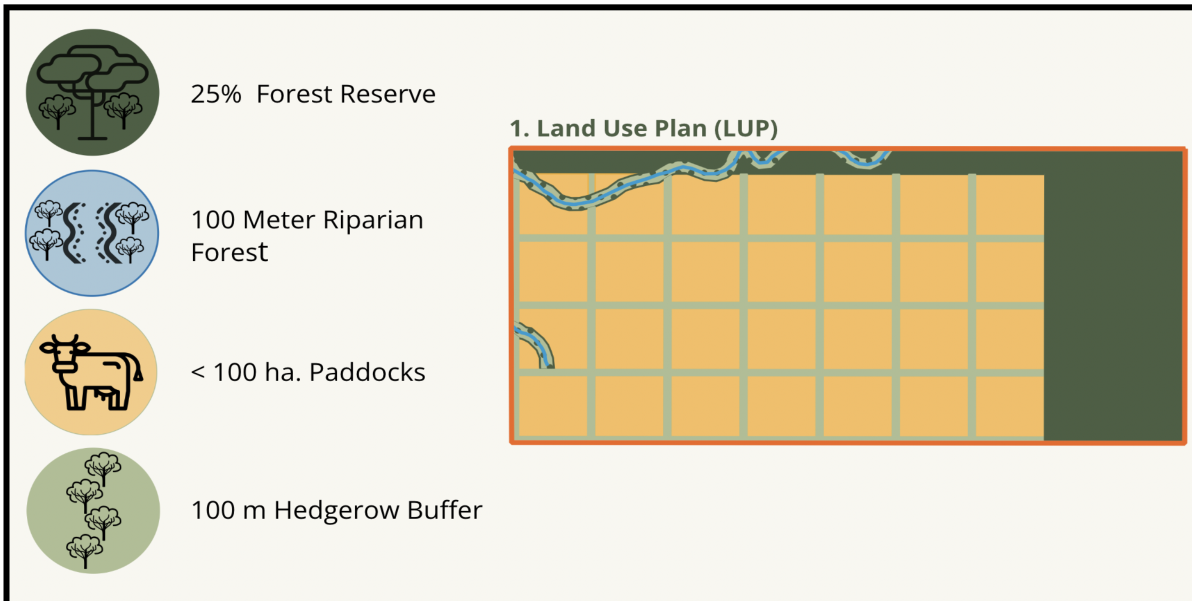


Figure 2: An example land use plan with all required components

Forest Loss & Forest Cover:

The raster dataset from INFONA's 2020 report titled "Nuestros Bosques" details yearly **Forest Loss** (i.e., deforestation) and **Forest Cover** for the periods of 2000-2020. Pixels of **Forest Loss** indicates a change from forest cover in the previous year, with the area of each pixel determined by the dataset's resolution at 900 meters squared.

Objective 1: Assessing LUPs, Quantifying Deforestation & Forest Cover

To understand and mitigate the impacts of deforestation, it is crucial to evaluate property owners' compliance with their approved LUPs and accurately quantify deforestation rates and forest cover. This region, rich in biodiversity, has been experiencing significant deforestation, making it a crucial study area.

The first task is to assess whether property owners are executing their approved LUPs as intended. This involves a detailed analysis of various datasets, including property boundaries, LUPs, and forest loss data, all provided by INFONA, the National Forestry Institute of Paraguay.

By leveraging geospatial overlays, we compare forest loss data against permitted land use. This approach allows us to identify areas where deforestation has occurred without authorization, which is a key indicator of non-compliance with approved LUPs. This approach allows us to identify areas where deforestation has occurred without authorization, providing valuable insights into the effectiveness of current forest regulations and land management practices.

Land use plan assessment:

- The assessment process involves determining the active private properties for each year of analysis, using the **Property Boundary** dataset. The *active-inactive.qmd* and active-lup.qmd files preprocess the data to identify the unique identifiers of active properties and active LUPs for each year from 2011 to 2020. These identifiers are then used to subset the **Land Use Plan** dataset.

- Each row of a LUP subset is a vector polygon of the approved land use type. The *lup_asssessment.qmd* analysis uses the yearly subsets of LUPs and overlays them with the corresponding **Forest Loss** dataset to determine the cell count per land use type. Each cell of the **Forest Loss** dataset represents a deforested area.

- Yearly subsets of the **Land Use Plan** dataset contain a categorical column of *'GRUPOS,'* identifying the approved land use type. The analysis done in *lup_asssessment.qmd* uses the *'GRUPOS'* column to filter by the land use types of *'authorized area'* and *'forest reserve'* (*'AREA_AUTORIZADA,' 'BOSQUES'*).

- Pixel counts were converted to an area for each property and land use type. Pixel counts greater than zero in the area designated as a forest reserve is considered illegal deforestation, placing the property out of compliance with its approved LUP.

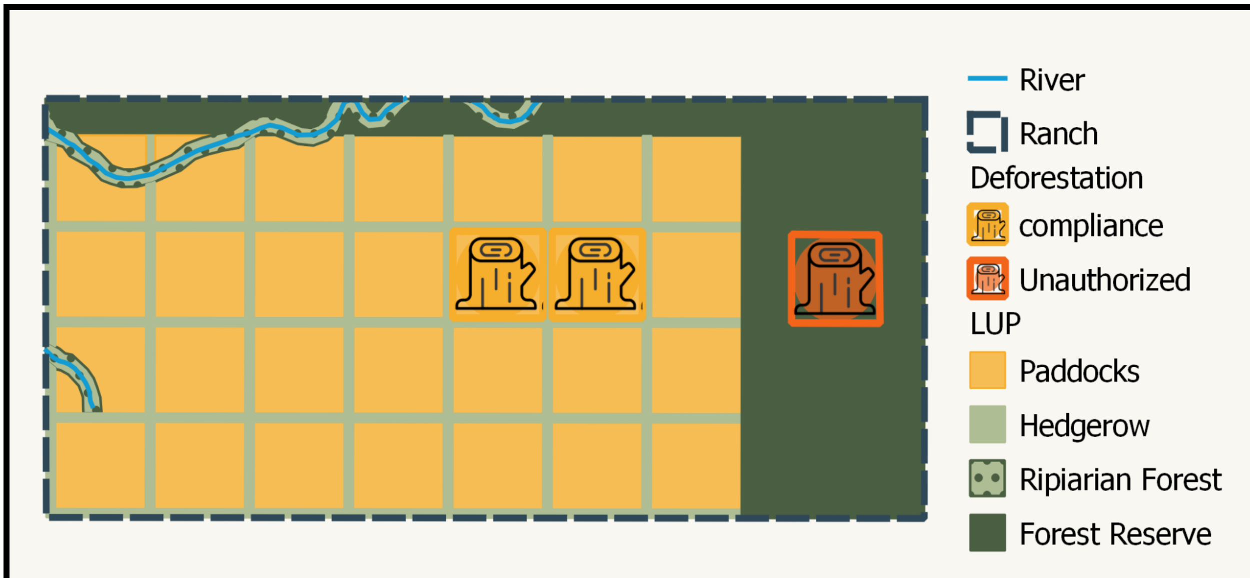


Figure 3: Illustrating authorized/unauthorized deforestation within areas approved according to LUP

Deforestation rates:

- A comprehensive analysis was carried out on the assessment results, aimed at determining deforestation rates and accurately quantifying total areas across various levels, including national, department, district, and individual property scales.

Forest cover:

- A similar approach was taken to quantify forest cover across time using the yearly **Forest Cover** datasets in conjunction with the **Study Boundary** dataset. An overlay of the **Study Boundary** and **Forest Cover** was performed to extract cell counts. Each cell of the **Forest Cover** dataset is considered tree cover. This dataset was then subset into districts and departments.

Objective 2: Estimating Forest Reserve in Undeveloped Regions

To quantify the area of forest that would be designated for protection under current and alternative laws, we have developed a simulation model that divides the undeveloped Paraguayan Chaco into mock properties. We apply custom functions for each mock property that generate LUP configurations based on the specified legal requirements. By simulating various scenarios with different conservation requirements, we can estimate the potential impact of policy changes on forest conservation in the Paraguayan Chaco.

Current Forest Law

- This scenario follows the current policy and legal requirements for LUPs enforced by INFONA. It includes a 25% forest reserve, a 100-meter hedgerow buffer, a 100-meter riparian forest, and paddocks of less than 100 ha for authorized deforestation. The purpose of this scenario is to simulate the continuation of existing practices and policies without any significant changes (Paraguay, 1986; Paraguay, 1995).

Alternative Forest Laws:

Promotes Forest Conservation

- This scenario aims to enhance forest conservation efforts. It proposes increasing the forest reserve requirement to 50%, along with maintaining a 100-meter hedgerow buffer, a 100-meter riparian forest, and paddocks of less than 100 ha. The objective is to simulate the potential outcomes of a policy that was proposed in Paraguay's National Congress in 2017 that would prioritize the preservation and protection of forests (National Congress, 2017).

Prioritize Cattle Production

- This scenario aims to find a balance between cattle production and forest conservation. It proposes a 25% total forest cover, which includes the combined area of the 100-meter riparian forest and 100-meter hedgerow buffer. Any additional forest area required to reach the 25% target would be designated as forest reserve. This policy includes paddocks of less than 100 ha. The intention is to simulate potential effects of a policy goal that prioritizes land use for economic purposes, while maintaining a 25% forest cover goal.

Law Ambiguity

- This scenario addresses a potential ambiguity in the law's interpretation. It suggests that if a property has been deforested beyond the approved amount, an immediate reforestation of 5% of the property is required in the areas of forest reserve. This is in addition to maintaining the 100-meter hedgerow buffer, the 100-meter riparian forest, and paddocks of less than 100 ha. However, some property owners have interpreted this policy as allowing them to deforest their entire property and only needing to replant 5%. This misinterpretation could lead to substantial deforestation, undermining the current forest law's original intent (National Congress, 2009).

These scenarios were selected to evaluate and compare the potential outcomes of different policy approaches concerning land use planning and forest conservation. Simulating a range of scenarios makes it possible to assess the trade-offs, benefits, and impacts associated with each policy option. The chosen scenarios represent distinct policy directions that Paraguay has considered adopting.

Defining the undeveloped region of the Paraguayan Chaco:

Within the study boundary, two distinct regions are defined:

1. **The developed region** takes into account urban areas (including roadways), indigenous lands, national parks, and all private cattle ranches registered with INFONA. To develop a ranch, the landowners require deforesting part of the property to plant grass for cattle. Deforestation occurring within the property requires submission of a LUP for approval by INFONA.

2. **The undeveloped region,** which accounts for approximately 40% of the Paraguayan Chaco, is defined as the study area excluding the developed region.

The undeveloped region is composed of properties that are likely to become cattle ranches in the future. This area lacks property boundaries and associated LUPs, motivating our creation of mock properties and simulated LUPs, under current and alternative forest laws, in the undeveloped region.

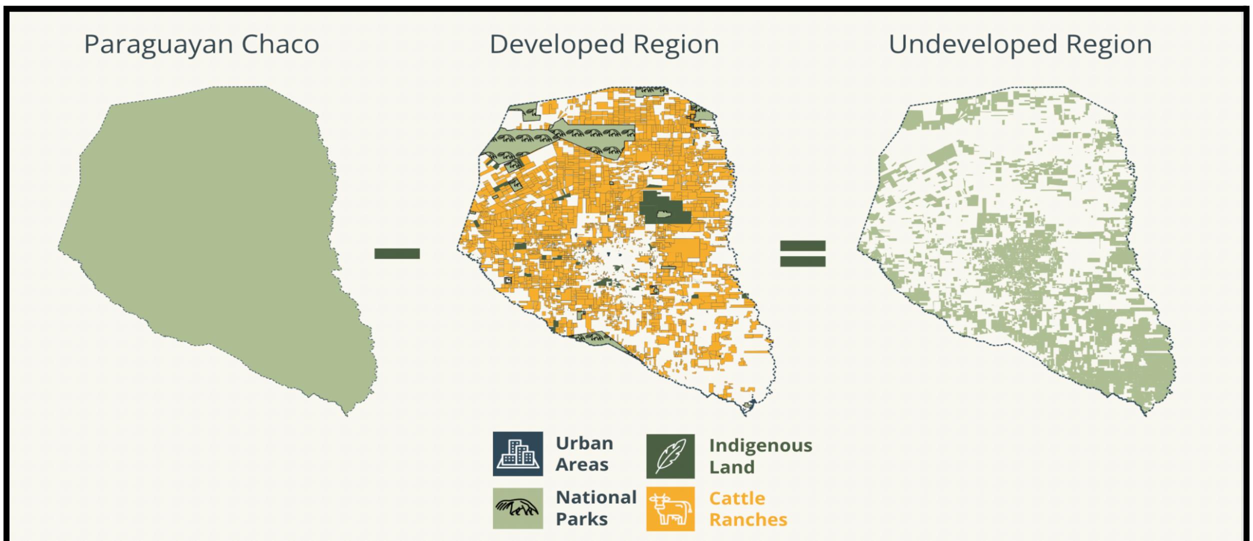


Figure 4: Undeveloped Region of the Paraguayan Chaco

Creating the mock properties in the undeveloped region:

- The custom function *propety_dimensions* was used to determine the dimensions to pass to the *cellsize* parameter of the R library *sf* function *st_make_grid* to make properties of 4,000 ha. The value for the property size was based on the average size of the land plots that the national government sold through the National Institute of Rural Development and Land (INDERT) for livestock farming establishments in the western region (Rojas Villagra & Areco, 2017).

- Only properties above 20 ha were selected, as this is the minimum required to register LUPs (Instituto Forestal Nacional, 2001).

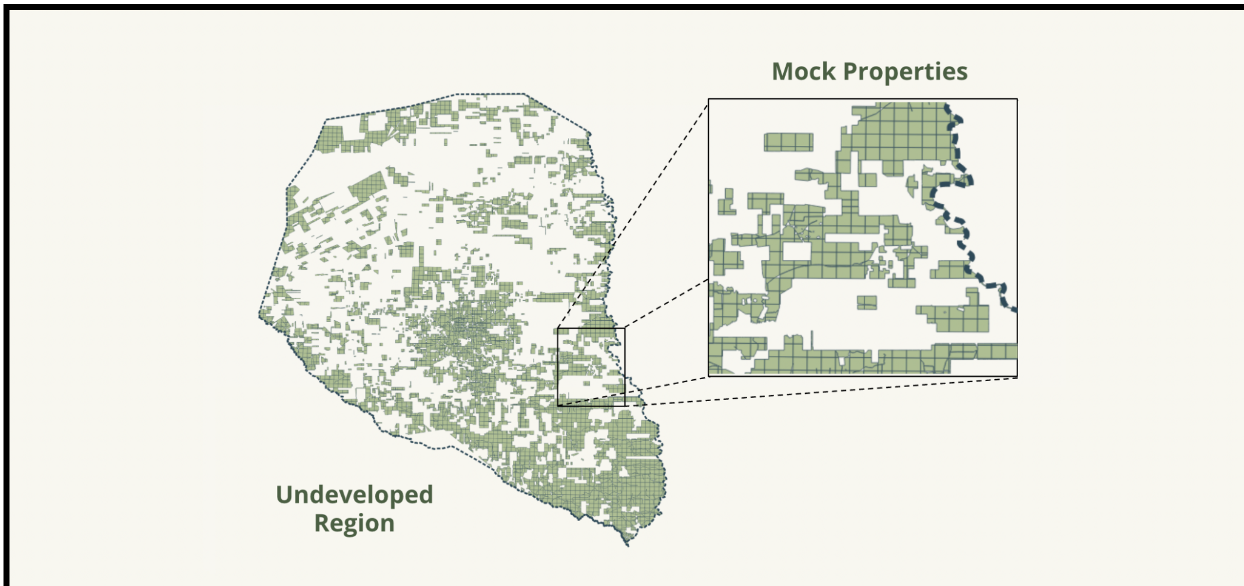


Figure 5: Mock properties in the undeveloped region of the Paraguayan Chaco

Creating simulated land use plans for each mock property:

- Each property iterates through a series of custom functions to create a simulation of LUPs. Each LUP has three main categories: forest reserve, paddocks, and hedgerows. If a river crosses the property, an additional category of riparian corridor is added. The functions are flexible enough to create different-size properties and paddocks with customizable aspect ratios. In addition, flexibility is extended to the width of the hedgerows and riparian corridors.

    - A new *create_optimal_dimensions.qmd* file was created for each simulation where, after the first pass of each property through the functions, the dataset was filtered repeatedly and reran, lowering the paddock size each time. For each simulation a minimum requirement is set for the category of forest reserve (i.e. 25%, 50%, 5%), the iterative process of lowering the paddock size ensures that each property has less than a three percent margin of error (e.g. <28%, <53%, <8%) but above the minimum forest reserve requirement. The repeated process could have been incorporated into the parallel process, but the process was done manually to remain consistent and maintain a quality check. This process returns the optimal dimensions to reach the minimum forest reserve and provides the area statistics for each category type within a property.

- Once the optimal dimension was determined for each property, the create_maps_with_optimal_dimensions.qmd was created for each simulation. This follows the same process as the previous pass and returns a dataset of the polygons of the category types created with the optimal dimensions for visualizations.

- The prioritize cattle production scenario, allowing the hedgerows between paddocks to be counted towards the 25% minimum, required reorganizing the *lup_simulator.qmd* file and the order of functions called. Changes are reflected in the following: *lup_simulator_hedges.qmd, create_optimal_dimensions_hedges.qmd, and create_maps_with_optimal_dimensions_hedges.qmd.*

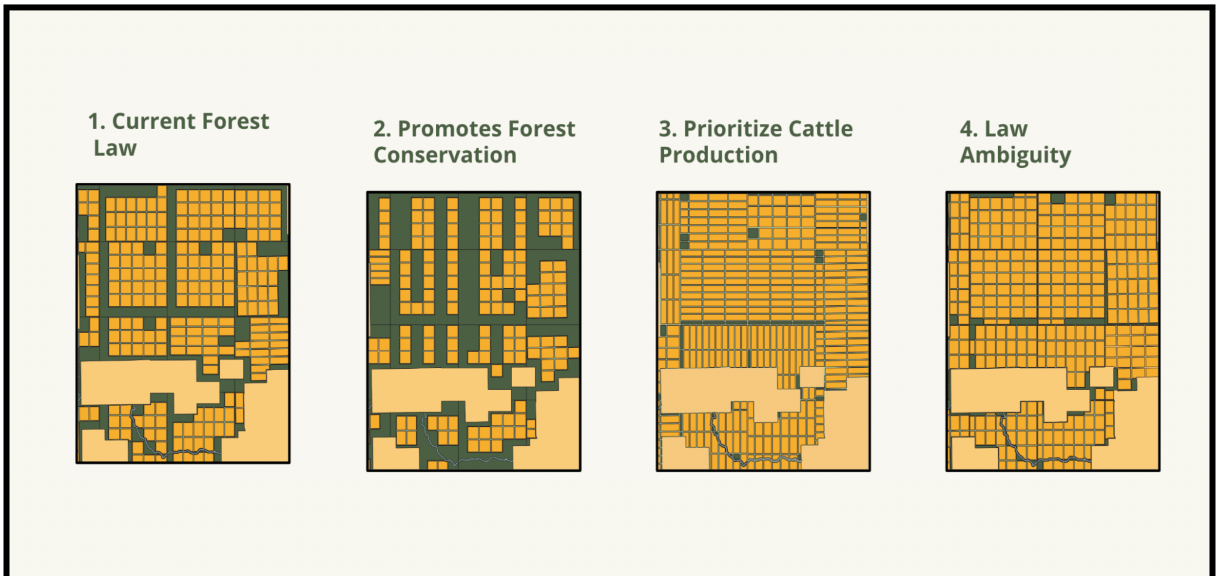


Figure 6: Examples of the simulated land use plans for each scenario.

Objective 3: Predicting Deforestation Patterns Under Alternative Laws

To determine how policy changes will affect deforestation across the Paraguayan Chaco in the next ten years, we first established the pattern of the actual occurrence of deforestation and its relationship to LUPs in the developed region of the Paraguayan Chaco. The mock properties and simulated LUPs of Objective 2 provide the underlying framework for predicting deforestation under different alternative forest laws over the next decade. The simulated LUPs provide a means to determine potential deforestation patterns in areas without registered LUPs and examine how these patterns may be influenced by changes in forest laws.

Machine Learning Approach:

The application of a Random Forest machine learning algorithm was decided on to predict deforestation patterns in the Paraguayan Chaco over the next decade. The choice of this algorithm is justified by several of its characteristics that are particularly relevant to the task at hand. Random Forest is capable of managing large datasets with a multitude of variables. This is a critical feature given the complex nature of deforestation, which is influenced by a wide array of factors. The algorithm is also robust to overfitting, a quality that enhances the reliability of its predictions, even in the presence of noise and outliers in the data. This robustness is achieved by averaging the results of many individual decision trees, each trained on a different subset of the data. Another advantage of Random Forest is its interpretability. It provides measures of variable importance, which can be instrumental in identifying the factors that exert the most influence on deforestation. Furthermore, Random Forest is adept at modeling complex, non-linear relationships between variables. Given the intricate interplay of factors contributing to deforestation, this ability to capture non-linear relationships is likely to be beneficial. Finally, Random Forest is known for its strong performance in predictive tasks. It has been demonstrated to provide accurate and reliable predictions across a range of contexts.

Datasets:

The datasets utilized in this study are integral to the analysis and the subsequent predictions of deforestation patterns. The first dataset comprises LUPs that were active up until 2010. This dataset serves as a cornerstone of our analysis, providing a historical context for land use and its potential impact on deforestation. The second dataset of significance is the Hansen dataset of forest cover change (Hansen, 2013). We utilized the 'treecover2000' variable, where the pixels with greater than an amount 10% of tree cover were selected and then masked with the deforestation that occurred from 2000 to 2010, this manipulation of the dataset establishes what we refer to as 'treecover2010'; This provides a binary snapshot of the state of tree cover at the end of the first decade of the 21st century. The binary ‘treecover2010’ was used as a mask to all other datasets so that the variables reflect where there is the possibility of deforestation. To capture the geographical and environmental characteristics of the region, the features of distance to roads and rivers (Appendix A, Table 1), travel time to ports and cities (Nelson, 2019), soil types (ISRIC - World Soil Information, 2023), and average annual precipitation from 2000 to 2010 (Funk et al., 2015) were included. These datasets provide important context for understanding the physical and environmental factors influencing deforestation patterns. The target variable of our model is the deforestation data from 2011 to 2020, sourced from the Hansen dataset. This dataset provides a record of deforestation events over this decade, which we aim to predict based on our selected features.

Training and Validation:

For all model runs, we utilized the Balanced Random Forest Classifier from the imblearn library. This model is an ensemble method that combines the predictions of several base estimators built with a given learning algorithm to improve generalizability and robustness over a single estimator. We employed RandomizedSearchCV for hyperparameter tuning, which performs a random search on hyperparameters, offering more efficiency than an exhaustive search like GridSearchCV.

The training and validation of our machine learning model is a two-step process designed to establish the predictive power of our features and optimize the model's performance. In the first step, we train the model using only the LUPs active up to and including 2010 as the feature to predict the target variable, the deforestation data from 2011 to 2020. This initial model allows us to assess the predictive power of the LUPs and establish a baseline for further analysis.

Following this, we proceed to the second step, incorporating all features into the model. These features include the LUPs of the developed region, geographical and environmental variables such as distance to roads and rivers, travel time to ports and cities, soil types, and average annual precipitation from 2000 to 2010. By training the model with this comprehensive set of features, we aim to enhance its predictive accuracy and reliability.

After training and validating the model with all the features, we identify the best-performing model based on its predictive performance. This model is then saved for future use, ensuring that the insights gained from the training and validation process can be effectively applied to predict future deforestation patterns in the Paraguayan Chaco under different scenarios of forest law changes.

Predictions:

The predictive model, trained and validated on data from the developed region, is now applied to the undeveloped region of the Paraguayan Chaco. The first step in this process involves using the simulated LUPs for the undeveloped region. Although deforestation data is not used as a target in this case, the LUPs serve as an essential feature, providing a basis for the initial deforestation predictions.

Subsequently, we generate the same set of features for the undeveloped region as we did for the developed region. These features include the simulations, reflecting current and alternative laws, and geographical and environmental variables such as distance to roads and rivers, travel time to ports and cities, soil types, and average annual precipitation from 2000 to 2010.

With these features and the simulated LUPs, we employ the trained model to predict deforestation in the undeveloped region. We use both the 'predict' and 'predict_proba' methods of the model. The 'predict' method provides the most likely outcome (deforestation or no deforestation), while the 'predict_proba' method gives the probabilities for each possible outcome. This dual approach allows us to obtain a nuanced understanding of the potential deforestation patterns in the undeveloped region under different scenarios of forest law changes.

Objective 4: Creating an Interactive Dashboard

To make our results accessible to policymakers and the general public, we have developed an RShiny dashboard that provides an interactive platform for exploring the data and visualizations generated from Objectives 1-3. The dashboard is designed to be user-friendly and intuitive, ensuring that stakeholders with varying technical expertise can easily understand the impacts of forest policies on the Paraguayan Chaco.

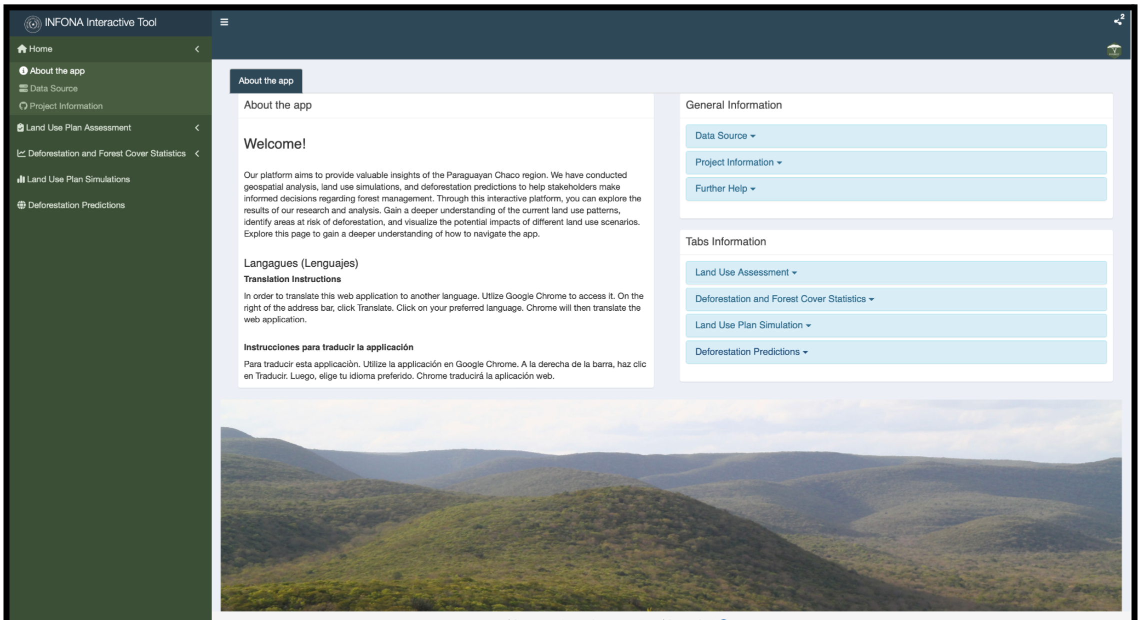


Figure 7: Home page of the interactive RShiny dashboard

The **Home** page of the RShiny dashboard is designed to provide the user with all the necessary information about the dashboard's purpose, functionality, and features. The home page also contains sections dedicated to general and tab information. These sections cover essential aspects such as the data source, project information, and further assistance or guidance to help navigate and make the most out of the various tabs within the dashboard.

The **Land Use Plan Assessment** page includes two subsections. The first sub-section enables users to access information based on political boundaries, while the second sub-section allows information retrieval by property ID. Within the political boundary section, users can explore maps, plots, and data specific to each department and district. They have the option to choose between unauthorized or authorized deforestation tabs and select different years to analyze land use patterns over time. In the property ID section, users have access to a map displaying all properties included in the assessment. Users can easily identify the compliance status of each property and view the amount of unauthorized deforestation associated with them. Additionally, a data table is available for searching and exploring this information.

The **Deforestation and Forest Cover Statistics** page has two sub-sections: one for deforestation statistics and the other for forest cover statistics. These sections offer visual aids like maps and plots to understand the spatial distribution of deforestation and forest cover in the study area. Users can also visualize the data based on political boundaries and explore different years for temporal analysis.

The **Simulations and Prediction Comparisons** page offers users the ability to compare results between Objectives 2 and 3. The user can compare statistics of total forest conserved per land use type from the LUP simulations and deforestation predictions. The user can also explore a map of pixel-wise probability of deforestation based on LUPs generated in the land use simulations.

Results

Objective 1: Assessing LUPs, Quantifying Deforestation & Forest Cover

- This analysis has determined that between 2019 and 2020, 44% of the deforestation within LUPs occurred in protected areas and was considered unauthorized.

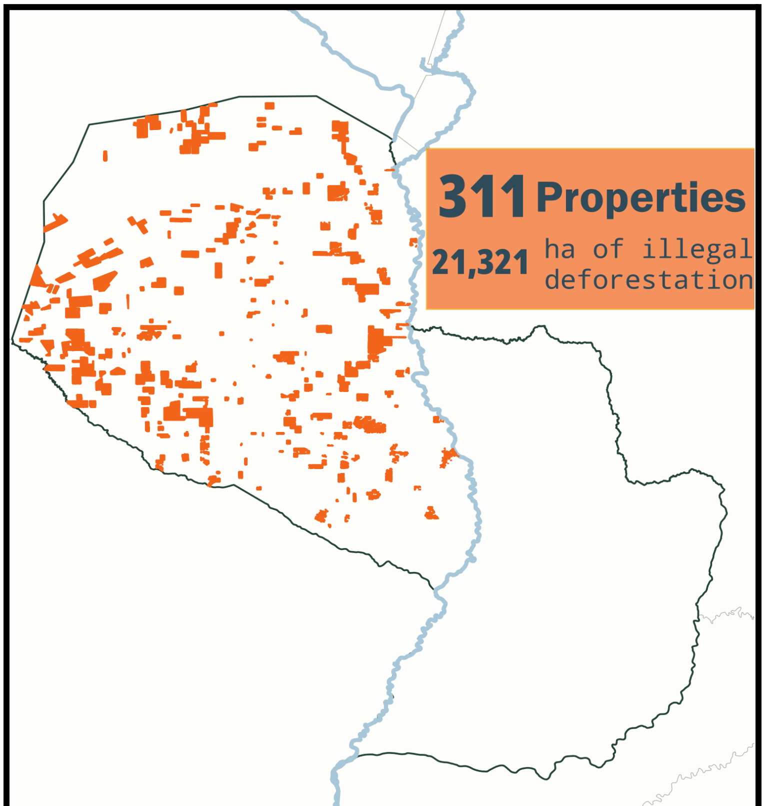


Figure 8 : 2019 Properties out of compliance with their land use plan

- Out of a total of 1800 properties, 311 (or approximately 17%) did not comply with their LUPs. A relatively small proportion of properties primarily drives the high percentage of deforestation in protected areas. This result will allow INFONA to explore the underlying factors or patterns driving this behavior and better determine how to mitigate unauthorized deforestation within LUPs.

- The spatial distribution of properties that committed illegal deforestation in 2019 leans heavily towards the furthest western boundary of the study region in the Boqueron Department. This distribution aligns with this study's analysis of the deforestation of the entire Paraguayan Chaco.

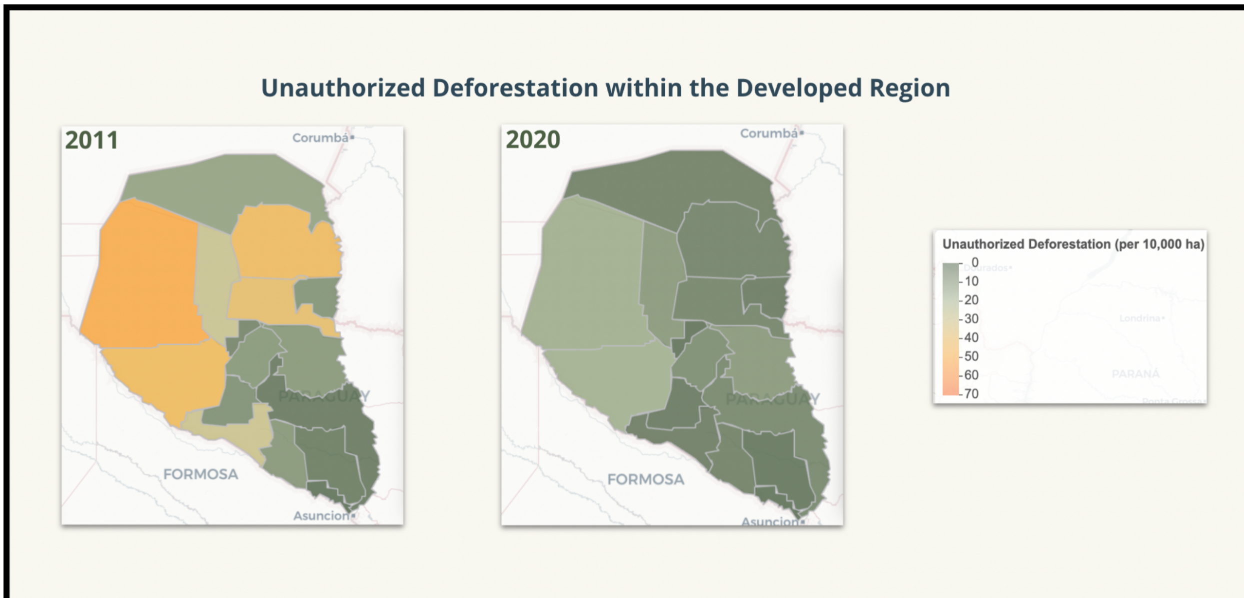


Figure 9: Spatial distribution of unauthorized deforestation within land use plans by district.

- To make comparisons of deforestation across the entire Paraguayan Chaco, the **Forest Loss** data required that they be normalized by year. Normalizing by year suggests that deforestation is constant throughout specific periods but was necessary to avoid spikes in data distribution. **Forest Cover** did not require normalization as the data provided was provided in a yearly format.

- Boqueron experienced an approximately 14% decrease in forest cover from 2011 to 2020, declining from ~67% to ~53%. The yearly percentages of deforestation in Boqueron reflect the decreases observed in unauthorized deforestation within LUPs. In 2011, 2017, and 2020, Boqueron had percentages of area lost by year of 1.88, 2.12, and 0.53, respectively. Though the rate of deforestation decreased over the study period, Boqueron leads in comparison to the departments' Alto Paraguay and Presidente Hayes.

- For the same three years of 2011, 2017, and 2020 Alto Paraguay had comparable decreases in percentages of 1.43, 0.75, 0.2. Forest cover decreased by 10% in the ten year period, 74% down to 64%.

- The values for the percentage of the area lost by year for Presidente Hayes are significantly lower than the other two departments within our study boundary, at 0.76, 0.75, and 0.28 for 2011, 2017, and 2020**,** respectively. An important point concerning the reported low values is that Presidente Hayes had the least forest cover to begin the ten-year period, with an approximate 50% coverage reduced to ~45.87% in 2020.

Objective 2: Estimating Protected Forest in Undeveloped Regions

- The decision was made to use the areas calculated from the rasters created in objective 3 as they are derived from the simulations created in Objective 2. The detailed reasoning for this decision can be found in Appendix A.

Estimates:

Table 2: Land Use Plan Simulation Estimates for Forest Conserved

Policy Scenario/ Simulation	Forest Conserved	Area Authorized For Deforestation
Current Forest Law	3,702,454	5,497,236
Promotes Forest Conservation	5,589,018	3,611,169
Prioritize Cattle Production	2,401,457	6,798,97
Law Ambiguity	2,233,436	6,965,255

- The scenario that prioritizes cattle production decreases forest conserved by approximately **1.30** million hectares.

- When comparing the current law to the one that promotes forest conservation, we see an increase of approximately **1.89** million hectares of forest cover.

- When comparing the current law to the law ambiguity scenario, we estimate a decrease of approximately **1.47** million hectares of forest cover.

- When compared side by side, you can see that the estimated range of forest conserved is **2.2 to 5.6 million hectares**. This is a potential difference of approximately **3.4 million hectares.** Though the difference of forest conserved between the **Law Ambiguity** and **Prioritizing Cattle Production** simulations appears to be minimal, it equates to approximately a 7% difference amounting to ~168,021 hectares which is more than twice the size of New York City (77816.19 hectares) (U.S. Census Bureau, 2022).

- This suggests that the **Law Ambiguity** could have a more detrimental impact on forest conservation than prioritizing cattle production, assuming all other factors are equal. However, it's important to note that these are simplified calculations and the actual impact would depend on a variety of factors, including the specific laws and regulations in place, enforcement of those laws, and the practices of the cattle industry.

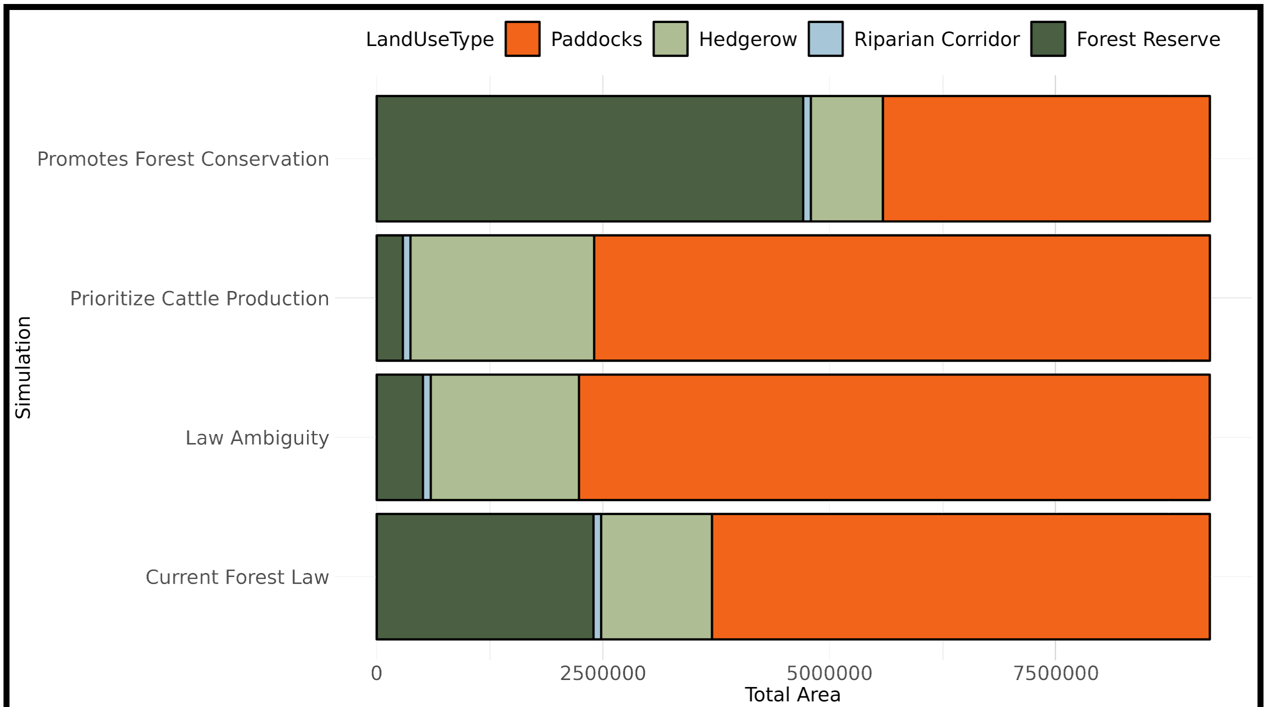


Figure 10: Comparison of estimates of protected forest and authorized areas based on simulations of LUPs.

Objective 3: Predicting Deforestation Patterns Under Alternative Policies

Training and Validation:

In the initial assessment of the predictive power of the LUPs and establishing a baseline for further analysis, the Balanced Random Forest Classifier performed reasonably well, achieving an accuracy of 0.66 and an F1-score of 0.68. A simple parameter space was chosen to increase model run time as performance was expected to improve with the inclusion of additional features.

The final model was a Balanced Random Forest Classifier with the best parameters found by RandomizedSearchCV being 'n_estimators': 50 and 'max_depth': None. With these parameters, the model achieved a best cross-validation score of 0.88, an accuracy of 0.94, and an F1-score of 0.94.

The model demonstrated robust performance on the training data, correctly classifying 2167159 instances of class 0 and 731728 instances of class 1, while misclassifying 10113 instances of class 0 and 2289 instances of class 1. The model's performance on the testing data was also robust, with high precision and recall scores for both classes.

The hyperparameters we tuned were 'n_estimators' and 'max_depth', with the options being [50, 100, 200] for 'n_estimators' and [None, 5, 10, 20] for 'max_depth'. The RandomizedSearchCV results indicated that the model with 'n_estimators': 50 and 'max_depth': None achieved the highest mean test ROC AUC of 0.975002, ranking first among all models.

Predictions:

For each of the simulations created in Objective 2 a corresponding map of pixel-wise probabilities (Fig.11), forecasting deforestation patterns over a 10 year period under current and alternative laws was created. The pixel-wise probabilities determine the amount of area predicted to be deforested and the remaining probability for the pixel is designated as non-deforested.

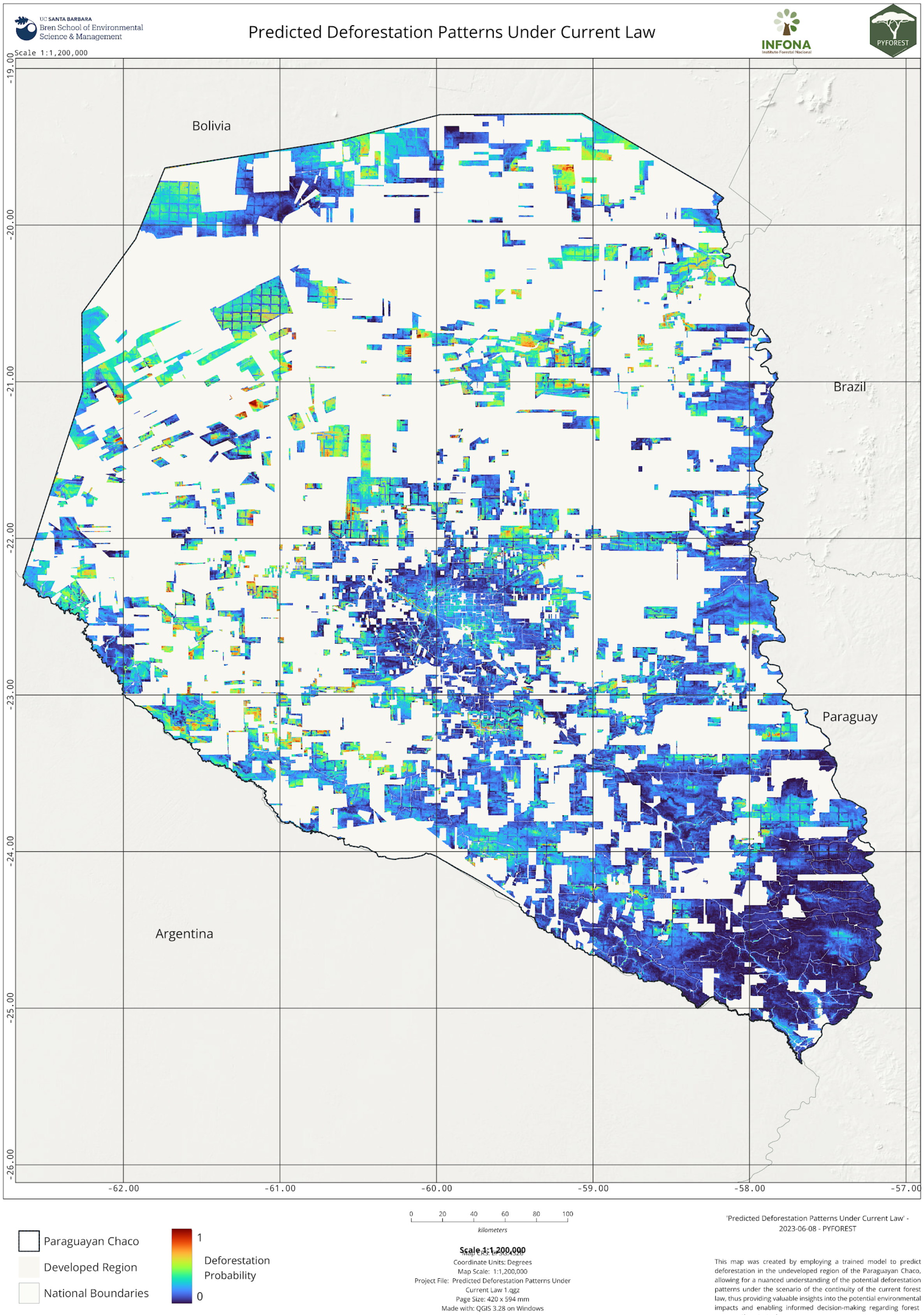


Figure 11: Pixel-wise probability of deforestation in the undeveloped chaco region,

The bins created for deforested and non-deforested create eight divisions, as opposed to simply the four categorical divisions of the simulations. As an example, the one categorical variable of paddocks in the simulation has two division for the prediction, the area designated as paddocks that we predict as deforested and the remaining area of paddocks designated as non-deforested.

The eight divisions of the prediction are illustrated in the figure 12, organized in a left to right fashion of non-deforested to deforested. The organization of the charts allows for distributions of estimated and predicted totals to be compared. The attached table to each illustration applies a two bin approach to the simulations of non-deforested and deforested where paddocks are the only variable within the deforested bin. The two bin approach allows for a different perspective when considering the implications of possible scenario.

The deforestation rate is calculated as the ratio of the deforested area to the total area (i.e., the sum of the deforested and non-deforested areas) multiplied by 100.

Current Forest Law: The total deforested area is 1,696,318 ha, the non-deforested area is 7,503,372 ha, and the deforestation rate is 18.43%. This means that under the Current Forest Law scenario, about 18.43% of the total area is expected to be deforested.

Law Ambiguity: The total deforested area is 1,841,201 ha, the non-deforested area is 7,357,490 ha, and the deforestation rate is 20.01%. This means that under the Law Ambiguity scenario, about 20.01% of the total area is expected to be deforested.

Prioritize Cattle Production: The total deforested area is 1,841,214 ha, the non-deforested area is 7,359,220 ha, and the deforestation rate is 20.01%. This means that under the Prioritize Cattle Production scenario, about 20.01% of the total area is expected to be deforested.

Promotes Forest Conservation: The total deforested area is 1,507,326 ha, the non-deforested area is 7,692,862 ha, and the deforestation rate is 16.38%. This means that under the Promotes Forest Conservation scenario, about 16.38% of the total area is expected to be deforested.

These results provide a comparative overview of the expected deforestation rates under different policy scenarios. They can help policymakers understand the potential impacts of different policies on deforestation and make informed decisions.

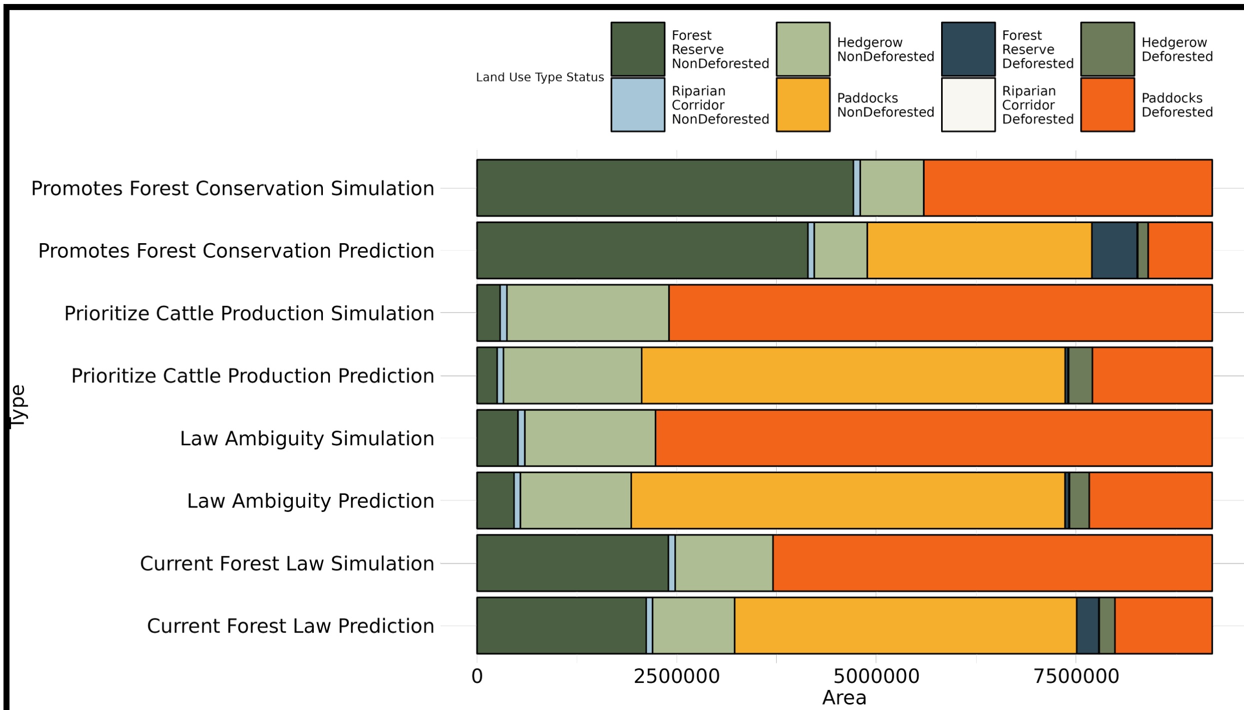


Figure 12: Comparison of Deforestation and Areas under Different Policy Scenarios by Land Use Types. Riparian Corridor Deforestation is insignificant, Non-deforested Riparian Corridor remains relatively constant.

Table 3: Comparison of Deforestation Rates and Areas under Different Policy Scenarios

Policy Scenario	Simulation Non-Deforested	Predicted Non-Deforested	Simulation Deforested	Predicted Deforested	Total Area	Deforestation Rate
Current Forest Law	3,702,454	7,503,372	5,497,236	1,696,318	9,199,690	18.43
Promotes Forest Conservation	5,589,018	7,692,862	3,611,169	1,507,326	9,200,187	16.38
Prioritize Cattle Production	2,401,457	7,359,220	6,798,978	1,841,214	9,200,435	20.01
Law Ambiguity	2,233,436	7,357,490	6,965,255	1,841,201	9,198,691	20.01

Visual Interpretation:

The results of the LUP assessment for Objective 1 reveal that properties not in compliance with their LUPs are predominantly located in the western region of the study boundary. The analysis of forest cover and deforestation rates throughout the Paraguayan Chaco confirms both reduced forest cover and lower rates of deforestation in the Department of Presidente Hayes. This study's machine learning model takes these facts into account, aiming to understand the lack of LUPs in regions at higher risk of deforestation and how policy changes could potentially shape the landscape.

Examining the pixel-wise probabilities of deforestation in the western region (see Figure 13, left panel), the risk of deforestation escalates significantly, with various pockets indicating near-certain probability of occurrence. Conversely, in the right panel of Figure 13, which depicts the southern region of the Department of Presidente Hayes, we see the trends of deforestation and forest cover mirrored in the low to near-zero probabilities of deforestation.

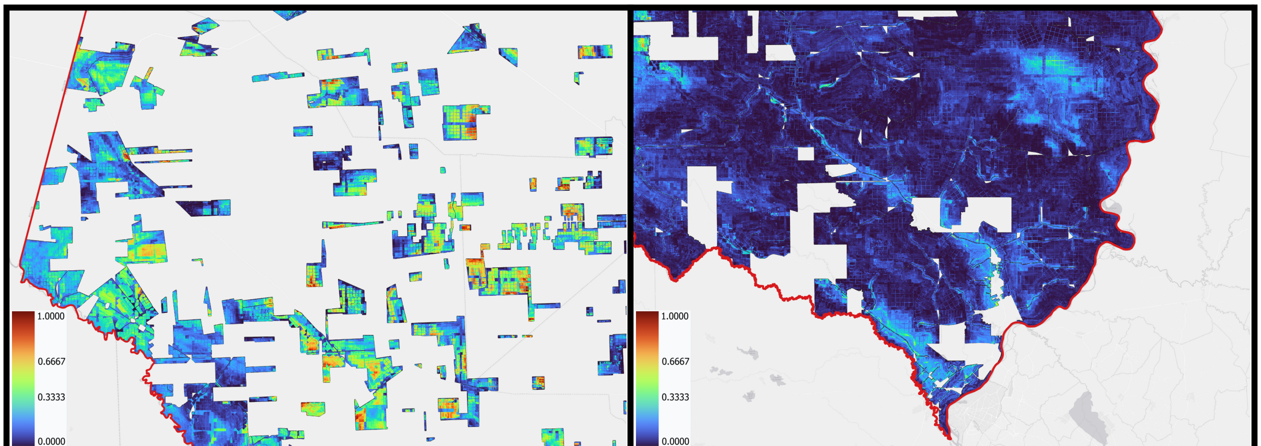
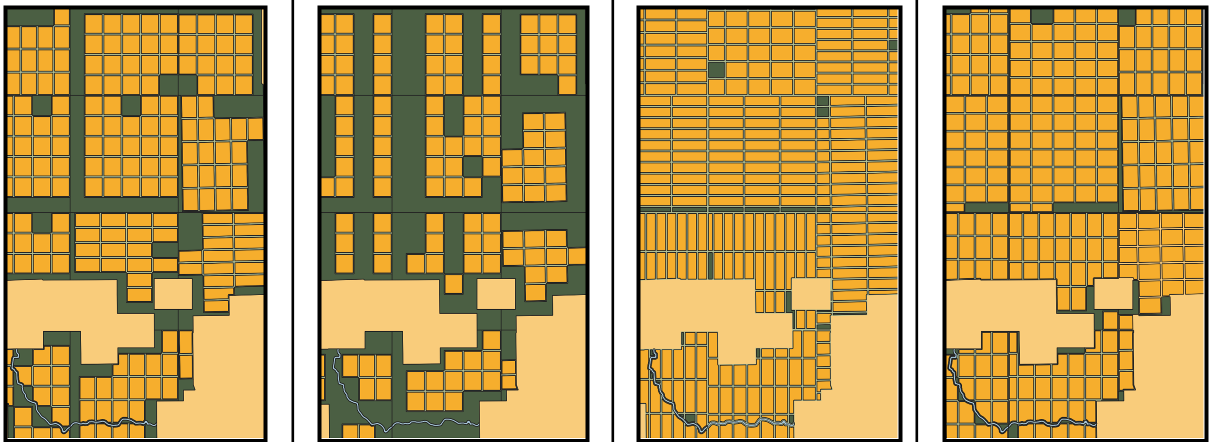
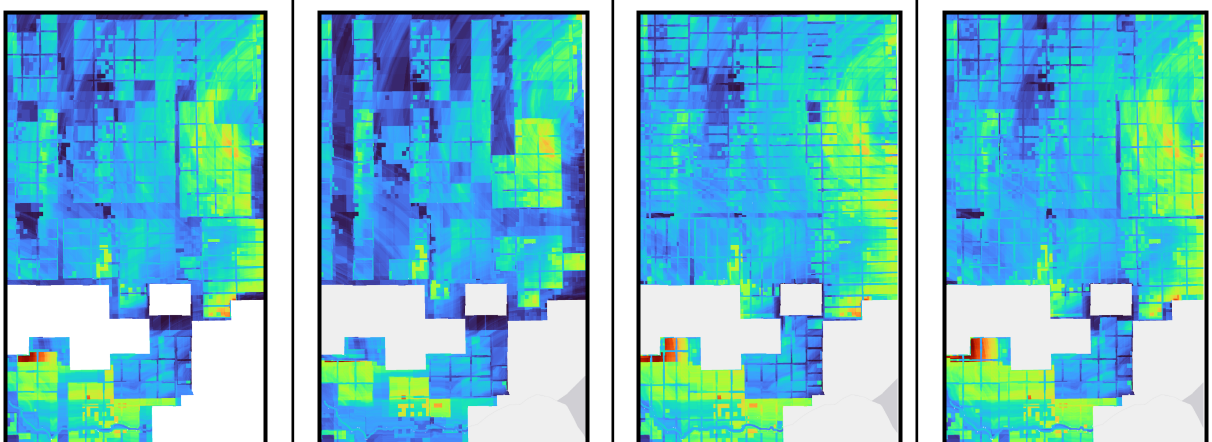


Figure 13: Left, western edge of study boundary of Department Boqueron with higher probabilities of deforestation. Right, southern tip of study boundary of Department Presidente Hayes with low to zero probabilities of deforestation.

The patterns observed suggest a potential high spatial autocorrelation within the trained model. Feature importance analysis on the trained model identifies distance to rivers and roads as the most influential features, accounting for approximately 20-25% of the model's predictive power, followed by precipitation. This is strongly reflected in the southern region, which boasts a dense network of rivers and experiences higher annual average precipitation. To further understand the influence of LUPs on the model's predictive power, refer to Figure 14 a-d, which contrasts against the pixel-wise probability raster of the corresponding simulation.



a. Current Forest Law Simulation | b. Promotes Forest Conservation Simulation | c. Prioritize Cattle Production Simulation | d. Law Ambiguity Simulation



e. Current Forest Law Prediction | f. Promotes Forest Conservation Prediction | g. Prioritize Cattle Production Prediction | h. Law Ambiguity Prediction

Figure 14: Contrasting pixel-wise probability rasters of deforestation against corresponding simulations to illustrate the influence of land use plans on the model's predictive power. Each panel (a-d) represents a different scenario, highlighting the variability in deforestation probabilities across different land use plan categories."

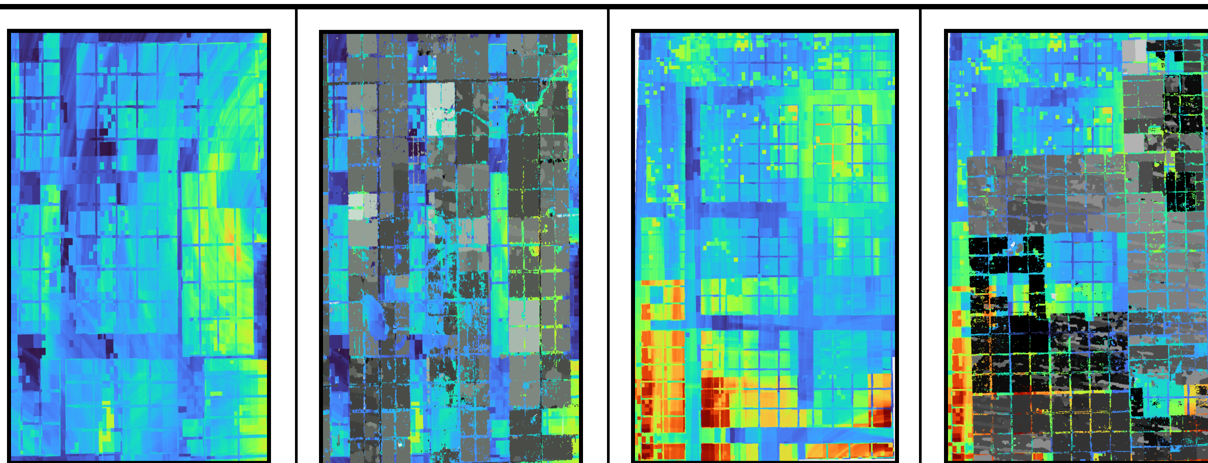
In a more detailed examination of the influence of the LUP, we can observe distinct patterns across different categories. In all scenarios, the outlines of the various categories within the LUP are discernible, and they generally exhibit higher probabilities in areas designated for deforestation. However, these areas also display a range of probabilities, with some spots nearing a probability of 1, indicating almost certain deforestation, and others hovering around 0.33, suggesting a lower but still significant risk.

Interestingly, lower probabilities are observed in protected areas, reflecting the effectiveness of these zones in mitigating deforestation. Additionally, we notice patterns that suggest the influence of geographical and environmental factors. For instance, there appear to be buffers that could represent a certain distance or travel time from key features such as roads or rivers. This could imply that accessibility plays a role in deforestation risk. Similarly, patterns in soil pixels are distinguishable, particularly in areas designated for deforestation. These patterns could be indicative of certain soil types or conditions that are more conducive to deforestation, although further investigation would be required to confirm this. This nuanced view underscores the complexity of the factors influencing deforestation and the importance of considering these factors in land use planning and policy development.

In conclusion, we present a pair of compelling examples where, by chance, we generated a simulated LUP in an area previously without one where the pattern closely mirrored the actual deforestation that occurred in the area. In some instances, the simulated plan even predicted very high probabilities of deforestation.

However, it's crucial to clarify that our model is not designed to predict specific locations where deforestation will occur. Instead, its strength lies in its predictive power to understand the potential outcomes of simulated LUPs based on different policy scenarios.

This underscores the potential of our model as a tool for policy simulation and planning. By providing insights into the potential impacts of different land use policies, our model can support more informed and effective policy development.



a. Current Forest Law Prediction	b. Current Forest Law Prediction + Hansen 2000 - 2021 Deforestation	d. Current Forest Law Prediction	e. Current Forest Law Prediction + Hansen 2000 - 2021 Deforestation

Figure 15: Two representative examples demonstrating the model's predictive power in simulating land use plans. These simulations, created in an area previously devoid of any land use plan, align closely with the actual deforestation patterns from Hansen (2013), underscoring the model's potential in forecasting outcomes under various policy scenarios."

Products and Deliverables

Table 4 summarizes the products and deliverables associated with the project's objectives, describing the specific output for each objective.

Table 4. Products and Deliverables

Objective	Product/Deliverable Description	Format
LUP Compliance Analysis	A dataset of spatial analysis assessing property owners' execution of their approved LUPs, quantifying historical deforestation rates, and identifying unauthorized deforestation.	lup-assessment.gpkg, forest_cover.gpkg, forest_loss.gpkg
Estimating Forest Reserve in Undeveloped Regions	A dataset of simulations dividing the undeveloped region into mock properties, applying a custom function to simulate LUP configurations under current and alternate laws, and estimating the area designated as forest conserved in each scenario.	mock-properties.gpkg
Predicting Deforestation Patterns	A raster of the pixel-wise probability of deforestation in the undeveloped region of Paraguay's Chaco.	final-prediction.tiff
Interactive Dashboard	An R Shiny dashboard providing an accessible and user-friendly interface for policymakers and the general public to explore the results and data visualizations generated from Objectives 1-3.	HTML

Summary of Testing

The following steps outline the various testing strategies used throughout the project. The strategies test for the data functionality, interaction testing, user testing, and performance testing conducted throughout the objectives outlined in the project.

Team Code Review

The team conducted peer team code reviews before merging the code into the main GitHub repository branch. Code review provided a quality check and caught any potential issues or bugs. The coding team member submitted a merge request via GitHub, and assigned team members reviewed, provided feedback on the code, and approved the merge request once it had met the standards to be merged into the main repository branch.

Unit Testing

Unit testing consisted of testing individual dashboard components, such as the data inputs and outputs, to verify that they function as intended. Testing was conducted using the R package testthat. The unit testing also includes standardizing all data sets to address any data that may be inaccurate or incomplete. Our analysis created data outputs consisting of dense mathematical calculations; built-in check-ins are incorporated throughout the code scripts to provide continuous feedback and to test for accuracy. A visualization and further upscaled checks and tests with larger datasets were completed to assess the visuals created.

Interaction Testing

In developing the RShiny dashboard, we took a targeted approach to ensure its effectiveness for specific user groups, such as internal INFONA users and academic researchers. To achieve this, a testing document was created with specific tasks and questions geared toward interaction and user testing, enabling these groups to understand the variables and variable types.

During interaction testing, we thoroughly tested all inputs and outputs to ensure they interacted seamlessly, enabling users to interact easily with the app. This involved carefully considering all variables and their types when creating interactions and testing them across different dashboard inputs to ensure seamless integration.

User Testing

To assess the RShiny dashboard's usability, functionality, and user experience, we tested with our target audience: policymakers and the general public, who may not have a technical background. During the test, users navigated the dashboard and interacted with its outputs to provide feedback on areas that require further development and ensure that the dashboard met their needs. We used the feedback gathered from the test to enhance the design and development of the app, thereby improving its overall quality and user satisfaction.

Performance Testing

Performance testing evaluated the Shiny dashboard's responsiveness, speed, and stability under various workloads and conditions to identify performance issues and potential improvements in the app. Testing the app included load time, response time, and processing time tests. This step tested the app's performance under different scenarios, such as a high volume of filters and selections, testing the responsiveness and efficiency of the app. The performance testing results were used to optimize the dashboard's design and improve its overall performance.

User Documentation

Intended Audience

This section will provide an overview of the user documentation available for the project, designed to assist INFONA, researchers, and other stakeholders in understanding, reproducing, and evaluating the project. The Capstone Team's code and documentation are version controlled and saved on GitHub at the following organizational link: cp-PYFOREST.

GitHub Documentation

The cp-PYFOREST organization consists of four main repositories that each contain code and documentation, including:

- *README.txt:* A file that outlines the workflow of each repository to reproduce the results
- *Code Files:* The repository hosts all code files (.R, .qmd, etc.) used in the project, ensuring accessibility and reproducibility for clients and researchers.

Links have been created within each repository to facilitate connections between repositories, providing improved access to the entire scope of the project.

Project Repositories and Descriptions

The tables below provide an overview of the file structure for each repository, along with descriptions of the purpose of each file.

1. Land-Use-Assessment

Table 5: Land Use Assessment Repository

Directory/ Subdirectories	Subdirectories/ Files	Description
01-preprocessing	active-inactive.qmd active-lup.qmd	Preprocessing required to determine the active properties and LUPs for each year between 2011 to 2020.
02-analysis	**compliance** lup_assessment.qmd **forest_loss_forest_cover** forest_cover_area_df.qmd forest_loss_area_df.qmd	**compliance** Compliance analysis quantifying illegal deforestation for each active property **forest_loss_forest_cover** Quantifying forest cover and forest loss for the Chaco
03-visualizations	leafdown_visualizations.qmd forest_cover_forest_loss_visualizations.qmd visualizations.qmd	Visualizations for the compliance, forest cover, and forest loss analyses
capstone.Rproj		Project directory
README.md		Markdown file providing information about the project

2. Land-Use-Plan-Simulation

Table 6: Land Use Plan Simulation Repository

Directory/ Subdirectories	Subdirectories/ Files	Description
01-preprocessing	active_inactive.qmd clip_buffer_river.qmd lup_simulator_hedges.qmd lup_simulator.qmd make_developed_region.qmd make_mock_properties.qmd map_demo_hedges.qmd	Custom functions, scripts for determining the undeveloped region, making mock properties, and simulating LUPs
02-analysis	create_maps_with_optimal_dimensions_25.qmd create_maps_with_optimal_dimensions_50.qmd create_maps_with_optimal_dimensions_5.qmd create_maps_with_optimal_dimensions_hedges.qmd create_optimal_dimensions_25.qmd create_optimal_dimensions_50.qmd create_optimal_dimensions_5.qmd create_optimal_dimensions_hedges.qmd lup_simulator_hedges.R lup_simulator.R	Using mock properties, simulation of LUPs dimension, and statistics. From dimensions, create a visual representation of LUPs
03-visualizations	plots.qmd visualizations_data_wrangling.qmd	Visualizations of the estimated protected forest area for various simulations
Land-Use-Plan-Simulation.Rproj		Project directory
LICENSE		Terms and conditions for use, reproduction, and distribution.
README.md		Markdown file providing information about the project

3. PYFOREST-ML

Table 7: PYFOREST-ML Repository

Directory/ Subdirectories	Subdirectories/ Files	Description
01-requirements	environment.txt environment.yml README.md	README.md provides steps for recreating the necessary Python environment, while the files environment.txt

Directory/ Subdirectories	Subdirectories/ Files	Description
		and environment.yml specify the required packages.
02-data-loading	hansen-ee.ipynb precipitation-ee.ipynb	workflows to retrieve the Hansen et al. (2013) and CHIRPS datasets from Google Earth Engine.
03-pre-processing	active-inactive.qmd active-lup.qmd extract-pixels-hansen.ipynb process-functions.py process-line.ipynb process-shapefiles.ipynb process-road-river.ipynb **sim-features-preprocessing** crop-mask-fores.-sim.ipynb crop-mask-lup-sim.ipynb process-road-river-sim.ipynb	These files contain code for pre-processing steps, including data extraction, conversions from vector to raster, and mosaicking and cropping data.
04-models	**notebooks** random-forest_validate-lup.ipynb BalancedRandomForestClassifier.ipynb BRFC-all-features.ipynb BRFC-all-features2.ipynb BRFC-sim-lup-prediction.ipynb BRFC-sim-lup-prediction-all-features.ipynb **scripts** BalancedRandomForestClassifier.py BRFC-all-features.py BRFC-all-features2-log.py BRFC-all-features2-nolog.py joblib-test.py	These notebooks contain the code for training, testing, evaluating, and finally predicting with the Balanced Random Forest model.
05-outputs	unique-values.txt **reports** BalancedRandomForestClassifier.pdf model_report-features.qmd model_report-log.qmd model_report-lup-only.qmd model_report-no-log.qmd plots.qmd random-forest_validate-lup.pdf	Model reports after training, testing and validating. Final statistical files

Directory/ Subdirectories	Subdirectories/ Files	Description
	predictions-log-lut-areas sim-25-log-lut-area-hectares.csv sim-5-log-lut-area-hectares.csv sim-50-log-lut-area-hectares.csv sim-hedges-log-lut-area-hectares.csv	
constants.py		This file defines constants used across the project.
__init__.py		Special Python file used to define the directory as a package, allowing import of its modules and sub-packages.
LICENSE		Terms and conditions for use, reproduction, and distribution.
README.md		Markdown file providing information about the repository

4. PYFOREST-Shiny

Table 8: PYFOREST-Shiny Repository

Directory/ Subdirectories	Subdirectories/ Files	Description
data	combined_auth_df_by_dist.rds combined_illegal_df_by_dist.rds compliance_updated.gpkg chaco_forest_cover.gpkg chaco_forest_loss.gpkg	Data required for dashboard
rsconnect/shinyapps.io/	**hendatasrl** pyforest-dashboard.dcf	A configuration file to manage the hosting and synchronization of the Shiny dashboard hosted on shinyapps.io

Directory/ Subdirectories	Subdirectories/ Files	Description
text	about_the_dashboard.md citation.md data_source.md disclaimer.md data_source.md Data_source.md land_use_assessment.md land_use_simulation.md project_information.md home_page_footer.md technical_documentation.md	Markdown and HTML files that contain text to be included in the dashboard.
www	Predicted_Deforestation_Patterns_Under_Current_Law.png auth_deforestation_yr_dpt.gif custom.css river_trees.JPG forest_conservation_lup_example.png	Directory in Shiny for images, logos, and styling.
global.R		This is the primary file for loading stored data object vectors used in the dashboard.
server.R		Defines the back end of the Shiny dashboard.
ui.R		Defines the front end of the Shiny dashboard.
pyforest-shiny.Rproj		Project directory
README.md		Markdown file providing information about the repository.

Archive Access

Data sharing and access

The National Cadastre Service, Federation for the Self-determination of Indigenous Peoples, and Ministry of the Environment and Sustainable Development datasets are all publicly available open-access datasets. The raw datasets provided by INFONA do contain personal identifying information. As such, data sharing and access are strictly managed and controlled by INFONA, ensuring compliance with privacy regulations and ethical standards. Any use or access to these datasets must be granted by the organization INFONA, respecting the privacy and confidentiality of the individuals concerned. Intermediate and final datasets are approved for public distribution, excluding identifying information to address privacy concerns.

Intellectual property and re-use

The datasets we utilized from the National Cadastre Service, Federation for the Self-determination of Indigenous Peoples, Ministry of the Environment and Sustainable Development, and INFONA are public sources, operating under the License of Use of Information and Public Open Data owned by the Paraguayan State. This license grants free, perpetual, and non-exclusive authorization for the use and/or transformation of public open data and information owned by the Paraguayan State to any natural or legal person who makes use of them.

Data archiving and preservation

The raw data for this project is stored in the shared *PYFOREST* Google Drive, accessible to all team members. The non-sensitive results data is stored on Zenodo, an open repository developed under the European OpenAIRE program and operated by CERN. Clear data documentation is provided in the cp-PYOFREST GitHub organization. The documentation outlines how to access and use the data, including any restrictions or usage requirements to help ensure that future users can reproduce our work. Furthermore, our GitHub repositories serve as documentation to note transformations or manipulations performed on the data that can help future users understand how it was used in our applications.

References

ABC Color. (2022, September 27). Deforestación en Paraguay: 756.967 hectáreas de bosque se perdieron en tres años, según Infona. ABC Color. https://www.abc.com.py/nacionales/2022/09/27/deforestacion-en-paraguay-756967-hectareas-de-bosque-se-perdieron-en-tres-anos-segun-infona/

de la Sancha, N. U., Boyle, S. A., McIntyre, N. E., et al. (2021). The disappearing Dry Chaco, one of the last dry forest systems on earth. Landscape Ecology, 36, 2997-3012. https://doi.org/10.1007/s10980-021-01291-x

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., & Michaelsen, J. (2015). The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. Scientific Data, 2, 150066. https://doi.org/10.1038/sdata.2015.66

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., Kommareddy, A., Egorov, A., Chini, L., Justice, C. O., & Townshend, J. R. G. (2013). High-Resolution Global Maps of 21st-Century Forest Cover Change. Science, 342, 850-853. https://doi.org/10.1126/science.1244693. Data available on-line at: https://glad.earthengine.app/view/global-forest-change.

Instituto Forestal Nacional (INFONA). (2001). “Resolucion SFN INT. N 224/2001, por la cual se reglamenta la elaboración y presentación de los planes de uso de la tierra. Dezember 7. https://www.infona.gov.py/application/files/6214/6108/8408/2001_Resol_SFN_224_Plan_de_Uso_de_la_Tierra.pdf

Instituto Forestal Nacional (INFONA). (2022). Nuestros bosques: Reporte de la cobertura forestal y cambios de uso de la tierra 2017 a 2020 (p. 78). https://drive.google.com/file/d/1aUCJTvdYenjAP52P5Owh4osyaeG4Gv4X/view?usp=sharing

ISRIC - World Soil Information. (2023). SoilGrids: A system for global soil information. Retrieved from https://soilgrids.org/

Milán, M. J., & González, E. (2022). Beef–cattle ranching in the Paraguayan Chaco: typological approach to a livestock frontier. Environment, Development and Sustainability, 25, 5185-5210. https://doi.org/10.1007/s10668-022-02261-2

National Congress. (2009, November 26). Proyecto S. 095374 "Que modifica el artículo 43 de la Ley 422/73 Forestal". http://silpy.congreso.gov.py/expediente/3118

National Congress. (2017, November 23). Proyecto S. 177745 “Que modifica el artículo 43 de la Ley 422/73 Ley Forestal”. http://silpy.congreso.gov.py/expediente/111372

Nelson, A. (2019). Travel time to cities and ports in the year 2015 [Data set]. Figshare. https://doi.org/10.6084/m9.figshare.7638134.v4

Paraguay. (1986). Decreto Nº 18.831/86: Por el cual se establecen normas de protección del medio ambiente. Retrieved May 19, 2023, from https://www.edydsi.com/index.php?option=com_content&view=article&id=12118:legislasys-decreto-1986-18831-d-18-831-86&catid=636&Itemid=1339

Paraguay. (1995). Ley N° 542 De Los Recursos Forestales. Retrieved May 19, 2023, from http://www.icasa.com.py/web/LEYES/AMBIENTE/LEY_N_542._DE_LOS_RECURSOS_FORESTALES.pdf

Rojas Villagra, L., & Areco, A. (2017). Las colonias campesinas en el Paraguay. BASE-IS. https://www.baseis.org.py/wp-content/uploads/2018/03/2017Dic_Las-Colonias-del-Indert.pdf

U.S. Census Bureau. (2022). QuickFacts: New York City, New York. https://www.census.gov/quickfacts/fact/table/newyorkcitynewyork/LND110220#LND110220

World Wildlife Fund. (2016). Gran Chaco. https://www.worldwildlife.org/places/gran-chaco

Appendix A

Table 1: Dataset Information

Dataset	Year(s)	Source	Data Type	Spatial Reference/Resolution
Political Boundaries	2000-2023	National Cadastre Service	Polygons	CRS: WGS 84 / UTM zone 21S
Forest Cover	2000-2020	INFONA	Rasters	CRS: WGS 84 / UTM zone 21S Resolution: 30m x 30m
Forest Loss	2000-2020	INFONA	Rasters	CRS: WGS 84 / UTM zone 21S Resolution: 30m x 30m
Permitted land use	1994-2023	INFONA	Polygons	CRS: WGS 84 / UTM zone 21S
Wildlife Protection Areas	2023	Ministry of the Environment and Sustainable Development	Polygons	CRS: WGS 84 / UTM zone 21S
Indigenous Land	2023	Federation for the Self-determination of Indigenous Peoples	Polygons	CRS: WGS 84 / UTM zone 21S
Forest Biome	2023	INFONA	Polygons	CRS: WGS 84 / UTM zone 21S
Hydrography and paths	2023	National Cadastre Service	Polygons	CRS: WGS 84 / UTM zone 21S

Area Error:

The following details the thought process in deciding to use the derived values for each of the categorical areas described in this Objective 2, from the raster created in Objective 3. This decision was made as the rasters are derived from the simulations created in Objective 2 results.

- For each of the four simulations, a shapefile was produced containing the geometry for the four categorical land use types to estimate both protected forest and authorized areas in the undeveloped Chaco region.

- During the generation of the final four simulations, statistics for each scenario were calculated before committing shapefiles to disk. As part of the development of the functions used to create the simulations, checks were conducted to ensure that the sum of the individual categorical types equaled the total property area. During this iterative process, it was observed that the sum of the categorical areas would occasionally be slightly lower than the total property area by approximately less than 1%. This discrepancy was consistent and infrequent enough to be considered negligible.

- However, when the final statistics of the completed four simulations were assessed, the error of total area for some of the individual properties carried over into the final total area of the undeveloped region. Each simulation produced a different total area, but the percent difference between the largest and smallest was approximately 1%. This issue arises from invalid geometry errors during the generation process.

- An external check using QGIS was conducted to determine the true area of the undeveloped region. When the smallest value for total area from the simulation was compared against the true value, the percent difference was calculated to be 2.66%.

- The need to delve into these details arises from the ultimate goal of comparing the values for the area of individual categorical land use types from the simulations against the predicted output of Objective 3.

- The issue of inconsistent areas re-emerged when the total area was calculated in Objective 3. Objective 3 produced total areas for the entire undeveloped region completely different for the same scenario when compared to the output of Objective 2.

- In Objective 3, we converted these vector shapefiles into rasters. This process was able to fill in the missing geometry with the appropriate encoded pixel for the categorical type, as manually determined in QGIS.

- The percent difference was 6.6X closer to the true value of the total area for the undeveloped region using the values determined from the rasters. This eliminates the variation across simulations and predictions, and reduces the variation between simulations.

- Comparing total area between smallest and largest rasterized simulation the percent difference was ~0.02% and comparing the smallest against the true value, the percent difference was ~0.4%. When comparing the raster against the vector for the same simulation the difference was 2.24% for the simulation with the largest gap between the two.