Strategic Environment:

Conservation Policies Effectiveness and Strategic Behavior

Angelo Santos [*]

October 5, 2024

Abstract

Conservation area policies are crucial for preserving natural resources. Using a REDD+ project funded by USAID in Eastern Zambia, I investigate the impact of community compensation on environmental outcomes. By analyzing Chiefdom and household data, I explore how forest dependence and social norms influence conservation performance. The protected area boundaries were established with local authorities, potentially leading to strategic site selection. Leveraging geospatial and household data, I assess whether communities strategically selected low-pressure areas for protection. My findings show no reduction in deforestation within protected areas but suggest negative spillover effects in non-protected areas. Non-protected areas increase tree cover loss after the policy, which lead to overall deforestation rates increased in treated chiefdoms post-policy. This research highlights the need to account for incentives and strategic behavior when designing effective conservation policies.

[*]Angelo Santos: Department of Economics, University of Houston, Houston, Texas, USA (afdossan@cougarnet.uh.edu). I want to thank professor Katharine Baldwin for kindly allowing me to use her data on chiefdom boundaries.
I also thank USAID for allowing me to use the CFP project dataset. The views and opinions expressed in this paper are those of the author and not necessarily the view and opinion of the United States Agency for International Development.

1 Introduction

From 2001 to 2022, there was a total of 459 Mha of tree cover loss globally (12% decrease) and 195 Gt of CO emissions. Related to it, deforestation accounts for 12% - 20% of the global Greenhouse gas emissions (GHG) (Watson, Schalatek, and Evéquoz, n.d.). These gases include different types of pollutants such as CO2, NO2, and SO2, contribute to temperature rise and air pollution. Another negative side of these gases is the impact on human development outcomes, such as health, mortality, and cognitive development. To reduce deforestation rates, policymakers have been designing different types of interventions to incentivize pro-environmental behavior, which needs to compensate for forgone income due to deforestation activities (Jayachandran 2022; Jack et al. 2022). This is especially relevant in regions where a considerable share of households are dependent on forest-related activities such as wood product production, charcoal, or cleaning forest for subsistence agriculture (Cisneros et al. 2022; Correa et al. 2020;).

Zambia has the 4th largest forest in Africa and has been facing a worrying increase in deforestation patterns in the last few years. The country has the 12th highest deforestation rate in the world, and 4th in per capita terms (Global Forest Watch). The Zambian tree cover reduced by approximately 9.4 % in the last 20 years, mainly driven by land clearing for agriculture, wood extraction, and charcoal production (USAID 2016). In 2015, USAID launched the CFP program in Eastern Zambia aiming to reduce gas emissions due to deforestation. The project was implemented in partnership with BioCarbon Partners, a firm specializing in REDD+ projects. The project aimed to protect a natural corridor that has an important biodiversity presence and to create a long-term sustainable relationship. To do this, the program delimited protected areas that would be used to quantify protected stored carbon due reduction of deforestation. These saved carbon would be sold as carbon credit in the voluntary market to create revenue to reinvest in the communities. The investments could be implemented in the form of basic infrastructure, and funding of more sustainable income sources such as beekeeping, eco-charcoaling, and cleaner subsistence agriculture. This aims to create income alternatives for the communities to reduce the cost of not deforesting the REDD+ areas.

Even though there is extensive literature analyzing different conservation policies, how to better design these policies is an open question (Balboni et al. 2023; Jayachandran 2022). Important aspects that may influence the effectiveness of these policies still need more evidence.

Take-up is a crucial element for the success of these initiatives, but the various dimensions that might impact this need more research. Some papers have shown how institutional dimensions such, as trust in contract realization, corruption, and political representation can influence the tree cover loss (Burgess et al. 2012; Cisneros and Kis-Katos 2022; Jack et al. 2022; Gulzar, Lal, and Pasquale 2024). Land property rights can also impact tree cover, as agents' incentives to protection may differ (Baragwanath and Bayi 2020, Sze et al. 2022 Baland et al. 2010). The amount of the financial incentives also matters as this needs to compensate for the opportunity costs of not deforesting. This opportunity cost can be explained by different aspects such as forest income dependence, and skill formation cost to change occupations. The sustainability of PES activities is another point of concern. These projects demand considerable funding to compensate for pro-environmental behaviors and research shows that after the end of financial support agents may go back to previous deforestation rates.

Using this intervention, I want to contribute to the literature that investigates how community pay-for-performance programs can reduce harmful environmental activities (Sims 2010; Malan et al. 2024). Specifically, I want to verify if the program reduced deforestation rates, fire alerts, and air pollution levels. In addition, I explore how forest dependence and social norms may impact the effects of the program. Using household survey I aim to construct forest dependence and institution norms indexes to evaluate if the program effectiveness varies depending on how communities are dependent on forest activities to obtain income or nutrition and how local authorities may influence the program effects (Jayachandran 2022; Acemoglu, Reed, and Robinson 2014).

To investigate the impacts of the program in Eastern Zambia I use administrative and geospatial datasets. I use information from USAID on the new protected areas defined by the program and information on treated chiefdoms and communities in the region. Using rich satellite data on tree cover baseline and loss from Hansen et al. (2013), and NASA VIIRS dataset on fire occurrence. I also collected multiple geographical information such as land agriculture productivity, settlement location, altitude, and distance to infrastructure. I merge this information with historical pre-colonial chiefdom boundaries to perform across and within analysis on these chiefdoms. After constructing the cell-level panel dataset I use differences in differences and event studies to identify overall and within chiefdom effects. I identify the overall effect by comparing cells in treated chiefdom with non-treated chiefdom. After that, I split treated cells into protected and non-protected areas to verify if there are within chiefdom differential effects between these two groups.

In the chiefdom level analysis, I found that the overall tree cover loss increased in the chiefdom that received the contract. This can be explained by that non-protected areas face higher deforestation after the program implementation, but no differential effects on the program-protected areas. The results from the event study suggest that there is no effect on protected areas tree cover loss post-policy. However, there is an increase in non-protected areas deforestation after in the post-policy period. This highlights the importance of considering incentives that are created when conservation area policies are established. The fact that the program's pay-for-performance is measured within the protected areas may distort agents' incentives to deforest in non-protected areas. The income received from the carbon market can fund chiefdom development which might be translated into more deforestation activities in non-protected areas.

The rest of this paper is organized as follows. Section 2 provides additional information on the CFP program and institutional background in Zambia. Section 3 describes the datasets used in the analysis and presents summary statistics for the Game Management Areas (GMA) and REDD+ Protected Areas. Section 4 outlines the empirical strategy, while Section 5 presents the results. Finally, the paper concludes with final remarks in Section 6.

2 Institutional Background

In 2015, USAID launched the Community Forests Program (CFP) in Eastern Zambia aiming to reduce gas emissions due to deforestation. The project cost approximately 16 million USD dollars to be implemented and was done in partnership with BioCarbon Partners, a firm specializing in REDD+ projects. The project aimed to protect a natural corridor called the Luangwa Valley, where there is an important biodiversity presence. According to the program report in 2019 (BCP 2019), the intervention successfully institutionalized a minimum of 700,000 protected hectares within the valley. They aimed to create a long-term contract relationship with the chiefdom that committed to protecting areas within the corridor. By estimating the stored carbon and avoiding emissions due to the intervention, the firm sold carbon offsets in the voluntary market. The income from the carbon offsets is reinvested in infrastructure and mitigation activities in communities within the contracted chiefdom.

The project took place in Eastern Zambia, where the majority of the households live in vulnerable situations in terms of income and service access. According to USAID (2016), 87.4

% of the population live in rural areas and do not have access to electricity, public water, or sanitation. Subsistence agriculture is an important component of household income, composing up to 64 % of it. Another important source of income is charcoal production, which is used for electricity, forest product production, and cooking by Zambian families. These families are economically vulnerable with 75 % of them living with less than $1.5 per day, and 60 % live in extreme poverty. This highlights how these communities may be very dependent on forest usage for income and nutrition, a clear challenge to the program's success.

The contract was established at the chiefdom level in partnership with chiefs, government agencies, and local authorities. The chiefdom REDD+ protected areas (Figure 1) were delimited after an extensive interaction with local communities and authorities to determine what areas would be protected and measured in other to measure the carbon offsets. These contracts have a period of 30 years and aim to create a long-term relationship with the chiefdom. During these years, the program committed to selling carbon-verified credits and reinvesting the revenue in projects that would benefit the communities. The carbon verification was established in 2019 and sold carbon-verified credits corresponding to the 5 years from the beginning of the program (2015-2019). After that, the carbon verification will happen in a year manner until the end of the 30-year contract in 2045. This component can potentially lead to a long-term incentive to conserve the protected areas, as the conservation revenue will keep being reinvested into communities.

To incentivize conservation in non-protected areas, the program financed basic infrastructure and livelihood income alternatives. The revenue obtained by the avoided emissions in the protected areas was allocated to the chiefdom responsible for its conservation. The chiefdom authorities then decided how to invest the resources in their communities. The program restricted the investments to the provision of physical infrastructure or financing of mitigation activities that aimed to create alternatives to forest products and non-sustainable activities such as crop burning.

In terms of investments in infrastructure, the program financed multiple types of physical benefits. For example, building schools, water distribution systems, boats for Ecotourism, and trucks for forest monitoring. These resources were provided to the communities conditional on local demands and chiefdom negotiations. For livelihood income alternatives, the project funds non-forest activities aiming to reduce the need for deforestation and forest degradation. The activities include bee-keeping, Eco-charcoal, Ecotourism, and smart agriculture.

For example, the bee-keeping project distributed hives to communities to allow them to produce honey without extracting this from forests. Another type of income alternative, the Eco charcoal technique, allows families to reduce forest degradation by being more selective on the tree type and tools for its extraction. The agriculture initiative includes techniques that reduce land degradation and still allow communities to produce crops but in more sustainable manner.

3 Data

3.1 USAID datasets

In 2015, USAID conducted baseline surveys that included 324 villages and 4,343 households across six different chiefdoms, five of which entered into contracts with the program. The endline dataset was completed in June 2024 and includes additional information on treatment types, such as cash transfer amounts and specific activities provided by the program. These surveys also contain geocoded information regarding the locations of villages and households, allowing me to map these communities.

Additionally, a structured survey interview was conducted with the headperson (traditional leader) of each village in the study area. The current headperson, whether at baseline or endline, was selected for the survey across all 324 communities. Village leaders provided information on whether the program was implemented in their village and if they received any benefits from it. For household heads, the survey also asked whether they received benefits from the program, such as participation in program activities, employment in ecotourism, or infrastructure improvements received by the village. Using this information, I was able to identify villages that received benefits within treated chiefdoms.

3.2 Zambia data

Chiefdom boundaries For defining the chiefdom boundaries, I use the shapefile from Baldwin (2013)[1], which geocodes the limits of the chiefdoms based on historical pre-colonial maps of the region. This shapefile provides an essential geographical foundation, ensuring that the spatial delineation of chiefdoms reflects historically accurate boundaries with cultural and po-

1. More details on the methodology and sources used to create the chiefdom boundaries can be found in Baldwin (2015), which provides a deeper exploration of the pre-colonial era maps and their relevance to modern governance structures.

litical significance. By using these geocoded boundaries, I can align contemporary data with historically rooted geographic divisions.

The geocoded chiefdom boundaries are particularly valuable for examining institutional and environmental outcomes across different regions. These boundaries allow for a more precise analysis of localized governance and resource management, especially in relation to forest dependence and agricultural practices. Additionally, they facilitate the integration of spatial data with household-level survey information, enabling a comprehensive examination of how historical chiefdom boundaries influence present-day outcomes.

Zambia Rural Agricultural Livelihood Survey (RALS) I use the Zambia Rural Agricultural Livelihoods Survey (RALS) to collect information about households and chiefs. This survey provides comprehensive data on small and medium-sized households across Zambia. The dataset is available upon request and requires a confidentiality agreement with the Indaba Agricultural Policy Research Institute (IAPRI).

The survey was conducted in 2012, 2015, and 2019, serving as a critical tool for understanding the dynamics of rural livelihoods, agricultural production, and rural development challenges in Zambia. RALS is a partnership between the Indaba Agricultural Policy Research Institute (IAPRI), the Central Statistical Office of Zambia, and the Ministry of Agriculture.

The study sample includes small and medium-scale farmers, i.e., those cultivating less than 20 hectares, and follows the Zambia 2010 Census sample. The dataset contains 17 sections in 2019, with detailed information on agricultural practices, household demographics, livestock ownership, and economic activities. It covers multiple dimensions of rural livelihoods, such as crop production, income sources, and access to essential services like credit and extension services. The survey's rich demographic and economic data, combined with geographic coordinates for household locations, allows for the mapping and construction of geographical indicators.

The survey includes several sections that gather detailed household information. I will focus on sections covering household demographics, farmland use, crop sales and production, off-farm income and remittances, agricultural inputs and outputs, and agricultural information and advice. These sections will allow me to extract key data on agricultural production, forest-related inputs, and chiefdom characteristics. By analyzing this information, I can construct measures of forest dependence and assess institutional outcomes in the chiefdoms.

3.3 Satellite data

Tree coverage data. To measure deforestation rates, I will be using the Global Forest Watch dataset (Hansen et al. 2013), which provides information for 30m $\times$ 30m cells. I will use two layers from this dataset: Forest tree cover and Tree cover loss.

The first layer is the baseline forest tree cover for 2000, which measures the share of forest canopy at the cell level. The dataset defines a tree as any vegetation taller than 5 meters in height. Following this definition, the author verifies the proportion of a cell covered by forest canopy—the upper layer or "roof" of a forest, formed by the crowns of trees—and assigns this value to the cell. A limitation of this definition is the possibility of including plantations that meet these criteria, such as timber plantations, in my forest cover measurements. For each cell, Hansen et al. (2013) reports the share of the unit covered by forest canopy, ranging from 0 to 100.

The second layer reports tree cover loss at the cell level. According to Hansen et al. (2013), forest loss is defined as a stand-replacement disturbance or the complete removal of tree canopy cover at the Landsat pixel scale. This means I am unable to track the gradual degradation of tree cover in a particular cell. For example, if a cell has 50% tree cover, I cannot track if this decreases to 20% and then to zero. The Hansen dataset only indicates whether the cell was fully deforested in a given year. Tree loss at the cell level is recorded as a numerical variable, indicating the year when the cell lost its tree cover. Values range from 0 to 23, where 0 indicates no loss, and 1 to 23 represents the year of loss starting from 2001. Using the annual data for tree cover loss, I can create yearly deforestation measures.

One limitation of these measurements is the inability to track tree cover gains after 2000. This means that if a cell experiences an increase in tree cover after 2000, this will not be reflected in my measurements. Additionally, a cell that undergoes multiple changes, such as deforestation, reforestation, and deforestation again, will not have these variations captured by my deforestation measure.

Tree cover measurements. Using the tree cover cell information, I can construct two measures of forested areas: total area forested in hectares and share of area forested in hectares. To define a particular cell as forested I use a canopy share threshold α. For example, if α is equal to 10, only cells with more than 10 of their area covered by forest canopy will be considered to calculate the tree cover measurements. Formally, consider a set of locations $i = 1, 2, ..., n$

with n being the total number of locations considered. Each location i has j grids defined by latitude and longitude coordinates $lat_j, long_j$ that are within its location boundaries b_i such that:

$$\begin{gathered} lat_{i,Min} \leqslant lat_j \leqslant lat_{i,Max} \\ \text{and} \\ lon_{i,Min} \leqslant lon_j \leqslant lon_{i,Max} \;. \end{gathered} \tag{3.1}$$

Each of these cells j has a canopy share $cs_{j,2000}$ in the baseline year 2000. With this I can define forest share fs_i in a particular location i conditional on forest canopy threshold α using the following:

$$fs_{i,2000}(\alpha) = \left(\frac{\sum_{j \in b_i} \mathbb{1}\{cs_{j,2000} > \alpha\} \times cs_{j,2000} \times 0.09}{\sum_{j \in b_i} \mathbb{1}\{j \in b_i\} \times 0.09} \right) \tag{3.2}$$

Intuitively, this equation states that forest share of a geographical location can be obtained by a simple ratio. The numerator is the sum of cells share of canopy transformed to hectares for those with canopy share above a specific threshold α and that overlap with location i boundaries. This is also the total forested area in hectares for location i. The denominator is the area cells in hectares that overlap with a geographic location i.

Deforestation measurements. Using the tree cover and deforestation layers, I can obtain my annual deforestation rate for a specific geographical unit. I will measure this in two different ways: total area deforested in hectares and the share of area deforested relative to 2000 baseline forest cover. I will use the same notation as before but add a new element to incorporate the possibility of deforestation per year. Let d_{jy} be the binary variable equal to 1 if a particular cell j faces forest loss in year y, zero otherwise. I will also use a threshold α to limit the cells used in our deforestation measurement. I.e, the relevant cells for the calculation are the ones with canopy share above a specific threshold α. For example, if I want to calculate the deforestation rate considering forest canopy share bigger than 10 %, the cell considered in both the denominator and numerator of the deforestation rate calculation are above this threshold. Using this notation, we can calculate location i's deforestation rate conditional on cell canopy share threshold α as follows:

$$D_{iy}(\alpha) = \left(\frac{\sum_{j \in b_i} \mathbb{1}\{d_{jy} = 1\} \mathbb{1}\{cs_{j,2000} > \alpha\} \times c_{j,2000} \times 0.09}{\sum_{j \in b_i} \mathbb{1}\{cs_{j,2000} > \alpha\} \times cs_{j,2000} \times 0.09} \right) \tag{3.3}$$

In the numerator I use the total area deforested in location i in hectare terms. In the denominator, I use the total forested area in 2000 in hectare terms. Both the numerator and denominator consider a threshold α for cell canopy share.

Figure 2 illustrates the raw measures of tree cover in the baseline for the Eastern Region. Figure 2b maps the tree cover loss in 2023. I also present histograms of measures for 0.1 degree combination grids for different groups (chiefdom and villages) in section 3. Following the literature on tree cover loss (Abman and Lundberg 2024; Cisneros et al. 2022) I will use the inverse hyperbolic sine (IHS) of hectares of tree cover lost. This transformation has been used in the deforestation literature due to the presence of zero in the outcome variable, but the convenience of interpreting the regression in log terms.

Fire data To measure the fire events, I will use the Fire Information for Resource Management System (FIRMS) from NASA. This dataset provides Near Real-Time (NRT) active fire data using images from the Moderate Resolution Imaging Spectroradiometer (MODIS) and the Visible Infrared Imaging Radiometer Suite (VIIRS). The NRT data is available within 3 hours of satellite observations, with the exception of the US and Canada, which have real-time data. The resolution is 375 m, and the data has been available since January 20, 2012.

The relevant information[2] for our research per cell includes the centered coordinates of the fire event in latitude and longitude for the 375 m pixel; the brightness temperature, which reflects the temperature of the fire pixel measured in Kelvin; and the date and time of fire recording.

3.4 Summary Statistics

The summary table 2 provides a comparison of characteristics between two layers in the Eastern region: Game Management Areas (GMA) and REDD+ Protected Areas (REDD+ PA). The analysis begins by looking at the cell canopy share in 2000, which represents the average percentage of tree canopy cover in each region. GMAs had an average canopy share of 22.40%, while REDD+ PAs had a slightly higher share of 24.09%. Similarly, the tree cover in 2000, measured in hectares, shows that GMAs averaged 31.50 hectares of tree cover per cell, whereas REDD+ PAs had 34.03 hectares. A notable difference between the two regions appears in tree cover loss between 2001 and 2014. GMAs experienced an average loss of 0.68 hectares per cell, while REDD+ PAs saw significantly less loss, with an average of just 0.08 hectares.

2. More information on the available data per cell can be found in Appendix A.2

In terms of altitude, GMAs are situated slightly higher, with an average altitude of 697.87 meters, compared to 677.82 meters in REDD+ PAs. Both regions show similar potential for maize yield, with GMAs producing an average of 3,100.74 kg/acre and REDD+ PAs showing a slightly higher potential at 3,117.92 kg/acre.

When considering the remoteness of these regions in relation to human infrastructure, GMAs are closer to both human settlements and roads. GMAs are, on average, 1.30 km away from human settlements and 3.98 km from roads, while REDD+ PAs are located 5.02 km from settlements and 11.91 km from roads, indicating greater remoteness. This pattern is further reflected in the distance to the electrical network, where GMAs are, on average, 45.77 km away compared to 85.85 km for REDD+ PAs.

In summary, REDD+ Protected Areas tend to experience less deforestation and are more remote compared to Game Management Areas. Both regions, however, exhibit similar potential in terms of agricultural productivity.

To understand the dynamics, Table 2 shows the correlation between the post-policy period and three cell groups within treated chiefdoms. First, there is no relationship between overall tree cover loss and the post-policy period. The correlation within the treated chiefdoms suggests possible spillover effects to non-protected areas. There is a negative correlation between the post-policy period and tree cover loss in protected areas. However, there is a positive relationship between the program and tree cover loss in cells outside the protected areas. A naive comparison between these two groups can lead to an overestimation of the program effects, highlighting the importance of finding a proper control for estimating the impacts of the program (West et al. 2020).

4 Estimation Strategy

Selection of protected areas Chiefdoms that received the contract chose the protected areas in partnership with the firm BioCarbon Partners. This decision was made after a validation process in which chiefdom boundaries and land properties in the region were verified. The endogenous nature of this decision-making process can lead to selection bias in the protected areas. Local authorities may have selected locations where deforestation pressure was already low, as the conservation of these areas influences the revenue received from avoided carbon emissions. Following Cisneros et al. (2022), I will run a probit model to understand if the choice of these areas is correlated with cell and chiefdom characteristics. I will estimate the following:

$$PA_{ic} = \alpha + X'_{ic}\beta + \psi_c + \epsilon_{ic} \tag{4.1}$$

In this model, PA_{ic} is a dummy variable that equals 1 if the cell i is included in the protected areas of chiefdom c, as determined by the intervention. X_{ic} is a vector of characteristics of the cell and chiefdom related to the likelihood of being protected. This includes the average deforestation rate before the policy start, tree cover in 2000, altitude, distance to human settlements, distance to rivers, and distance to roads.

To account for the possibility of different decision processes conducted by chiefdoms, I include chiefdom fixed effects ψ_c. This term captures time-fixed observable chiefdom characteristics that may influence the choice of protected areas and their boundaries. For instance, this can include the influence of chiefdom size, political structure, or geographical conditions on the selection of conservation areas.

Additionally, in order to understand if the predictors have heterogeneous effects across different chiefdoms, I can run the same model, adding interactions between chiefdoms and these predictors. If some chiefdom attributes assign different weights to certain cell attributes, this interaction can capture that effect. One hypothesis is that different chiefs may give more weight to more populated areas due to their land tenure situation. One could imagine that more centralized chiefdoms have more enforcement power over land and allocate more land to be protected.protected.

Chiefdom outcomes The program signed contracts with selected chiefdoms located in the Luangwa Valley. After consultations with local authorities, protected areas were delineated within these chiefdom boundaries. The endogenous nature of protected site selection may have led local communities to choose areas with low deforestation pressure. Economic incentive theory suggests that non-protected areas may experience increased deforestation due to the reallocation of deforestation pressure from protected areas. To test this hypothesis, I first examine the overall effects in treated chiefdoms. Next, I differentiate between protected and non-protected areas within the treated chiefdoms. The comparison group for this analysis consists of cell-level data from non-treated chiefdoms located in the Luangwa Valley.

To estimate the program's effect on chiefdom outcomes, I use Difference-in-Differences (DiD) approach that explores cell-level information (Abman and Lundberg 2024; Cisneros et al. 2022). The assumption is that the deforestation rates would maintain the same trend in the absence of the program for chiefdoms that received the USAID contract and those that were not included in the contract. I will run the following regression:

$$y_{ict} = \beta_0 + \beta_1 \mathtt{After}_t \times \mathtt{Treated}_c + \alpha_t + \delta_c + \epsilon_{ict} \tag{4.2}$$

In this equation, $y_i ct$ corresponds to the inverse hyperbolic sine of tree cover loss of cell i in chiefdom c at year t. $\mathtt{After}_y$ is a dummy variable equal to 1 if year y is after 2015, when the program started. $\mathtt{Treated}_c$ indicates if chiefdom c is included in the CFP contract. β_1 captures the overall effect of the program for cells in the treated chiefdoms. I use time (δ_t) and chiefdom (γ_c) fixed effects to control for time-invariant unobservable characteristics that may impact tree cover loss. ϵ_{ict} is an idiosyncratic error term, and I clustered errors at the chiefdom level to account for serial correlation between cells.

To estimate cumulative yearly effects of the program on treated chiefdom, I estimate the following event study:

$$y_{ict} = \alpha + \sum_k \beta_k \mathbb{I}(t = k)\mathrm{CFP}_c + \delta_t + \gamma_c + \epsilon_{ict} \tag{4.3}$$

In this equation, $y_i ct$ corresponds to the inverse hyperbolic sine of tree cover loss of cell i in chiefdom c at year t. I regress this on yearly coefficients (β_k) before and after the program implementation. The coefficients will give me the yearly cumulative effect of the program on deforestation outcomes. I use time (δ_t) and chiefdom (γ_c) fixed effects to control for time-invariant unobservable characteristics that may impact tree cover loss. ϵ_{ict} is an idiosyncratic error term, and I clustered errors at the chiefdom level to account for serial correlation between cells.

After evaluating the effect of the program on the treated chiefdom, I will run a similar regression as before. To split the cells in treated chiefdom in protected areas and non-protected areas I create a dummy variable (PA_i) which is equal to 1 if a cell lies in the protected areas. Similar to that, I also include a dummy (NPA_i) for cells outside the protected areas and within the treated chiefdom. The Difference in Difference equation is the following:

$$y_{ict} = \beta_0 + \beta_1 \mathtt{After}_t \times PA_i + \beta_2 \mathtt{After}_t \times NPA_i + \alpha_t + \delta_c + \epsilon_{ict}\ , \tag{4.4}$$

The change relative to 4.2 is the usage of PA_i and NPA_i. β_1 captures the overall effect of the program in the protected areas, whereas β_2 captures the effect of the program in the non-protected areas.

For the event study, the estimation is the following:

$$y_{ict} = \alpha + \sum_k \beta_k \mathbb{I}(t = k)PA_i + \sum_k \lambda_k \mathbb{I}(t = k)NPA_i + \delta_t + \gamma_c + \epsilon_{ict} \tag{4.5}$$

In this event study, β_k captures the yearly cumulative effects of the program in the non-protected areas, whereas λ_K captures the yearly cumulative effects of the program in the non-protected areas. Splitting the cells within treated chiefdoms allows to identify of differential dynamics within treated chiefdoms.

5 Preliminary results

In this section, I will discuss the results regarding the effects of tree cover loss following the policy intervention. First, I will present the chiefdom-level analysis of overall deforestation for treated chiefdoms. Second, I will investigate the program's effects on both protected and non-protected areas within these treated chiefdoms.

The overall effect of the program is illustrated in Table 3. The first column shows the results for all cells in treated chiefdoms, while columns 2 and 3 separate the estimations for protected and non-protected areas. The results from the first column indicate an increase in overall tree cover loss in chiefdoms after the program's implementation. This effect may be a consequence of the composition of effects shown in columns 2 and 3. The last two columns reveal that protected areas did not experience a reduction in deforestation rates, while non-protected areas saw an increase of 0.6 in tree cover loss.

The event study results from Eq. (4.5) suggest an overall increase, although this increase in chiefdom tree cover loss is not statistically significant. Figure 4 shows that treated and non-treated chiefdoms were following parallel trends before 2015. When the program began, the trend in the cumulative effects of the program was positive but not statistically significant. However, this suggests that chiefdoms implementing protected areas may have experienced an increase in tree cover loss.

When I split the analysis between protected and non-protected areas within treated chiefdoms, the results are similar to those observed in Table 3. The cumulative effect of the program is not systematically different for protected areas, which has one negative and statistically significant coefficient in 2020. However, non-protected areas show an increase in deforestation that persists after the policy implementation.

The event study dynamics decompose the average treatment effects observed in Table 3. These results suggest a compositional effect, indicating that treated chiefdoms deforest less by not systematically changing deforestation rates in protected areas while increasing tree cover loss in non-conservation areas. This corroborates findings in the conservation literature, where spillover effects have been systematically observed (Cisneros et al. 2022; Amin et al. 2019; Giudice et al. 2019).

The results suggest that while the CFP may have had some success in maintaining low levels of deforestation in protected areas, non-protected areas and treated chiefdoms experienced an increase in tree cover loss post-policy. This could indicate leakage effects, where deforestation activities shift to areas outside protected zones, thereby canceling out the overall effectiveness of the program.

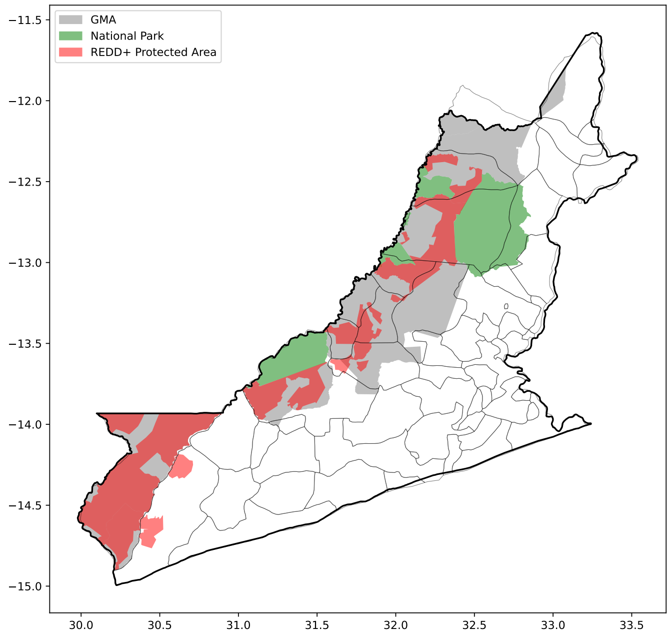
6 Conclusions

This paper investigates the impacts of a REDD+ program in Eastern Zambia on environmental outcomes and examines how local communities may strategically select conservation areas. The results suggest that protected areas were ineffective in reducing tree cover loss trends during the post-policy period. Conversely, non-protected locations within contracted chiefdoms experienced an increase in tree cover loss. Furthermore, I find that treated chiefdoms are deforesting at higher rates following the intervention. This effect may arise from distortions in the incentives faced by stakeholders in non-protected areas, where communities lack motivation to alter deforestation patterns since these actions do not affect revenue from carbon credits. Additionally, investments in these areas may exacerbate deforestation due to development pressures within chiefdoms.

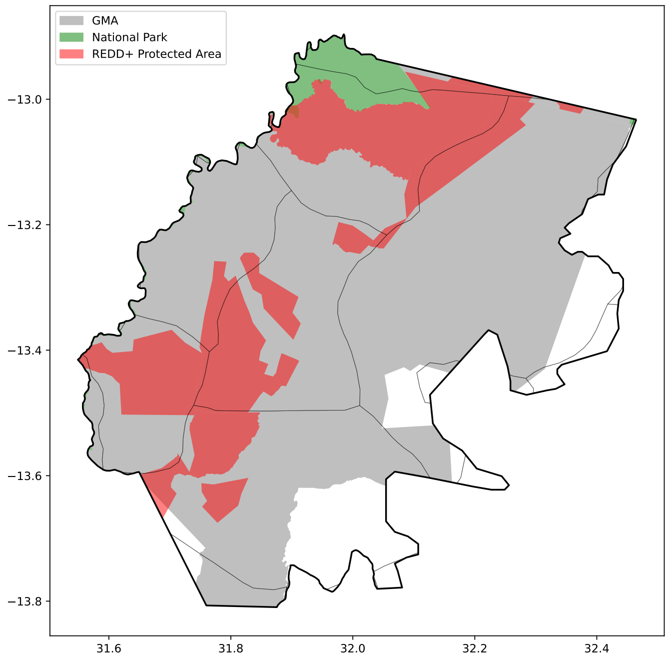
These findings align with existing literature on conservation, which identifies spillover effects of protection initiatives on non-protected areas. This underscores the importance of internalizing stakeholder incentives when delineating conservation areas.

Tables and Figures

Fig. 1. Conservation areas in Eastern Province and Mambwe District



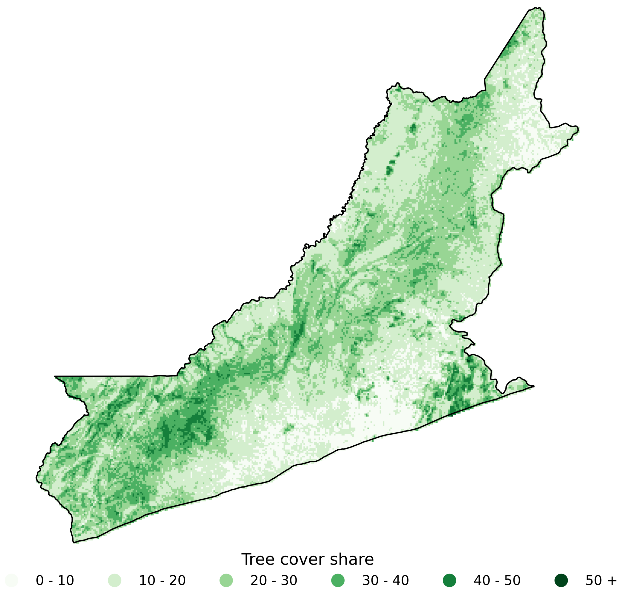
(a) Eastern Province



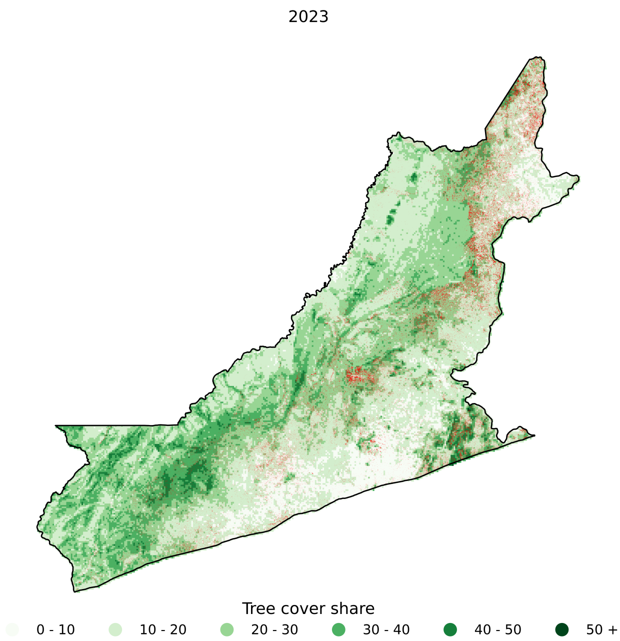
(b) Mambwe District

Note: These maps illustrate the overlapping of different geographical layers in Eastern Province. Black lines delineate the region according to chiefdom boundaries. The green layer represents the Natural Parks, the gray layer shows the Game Management Areas, and the red layer indicates the protected areas defined by the REDD+ program.

Fig. 2. Tree cover and tree cover loss distribution in Eastern Province



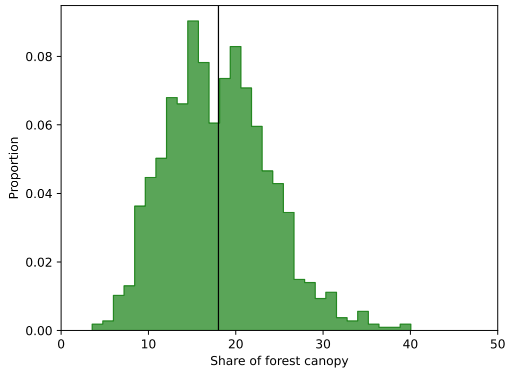
(a) Tree cover in 2000



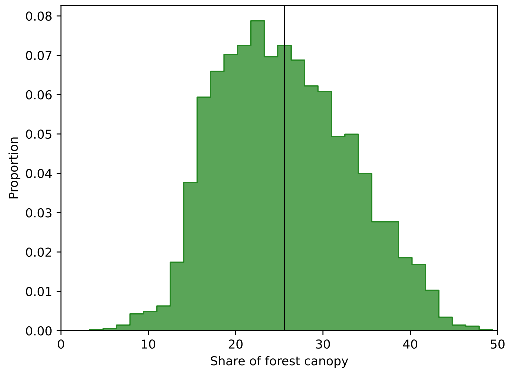
(b) Cell tree cover loss in 2023

Note: The maps plots tree cover in 2000 and tree cover loss in 2023 from **hansen2013<empty citation>**, *using 30m × 30m hansen cells. Figure 2a maps the tree cover in 2000 for each cell, and Figure 2b highlights the deforested cells in 2023, shown in red.*

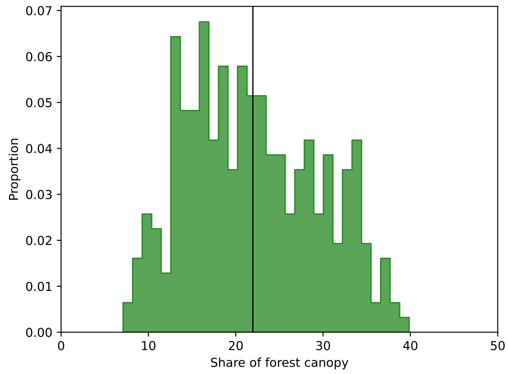
Fig. 3. Chiefdom distribution of 0.1 degree cell share of tree cover



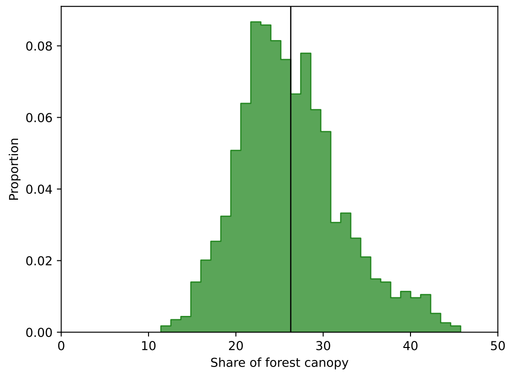
(a) Chikomene



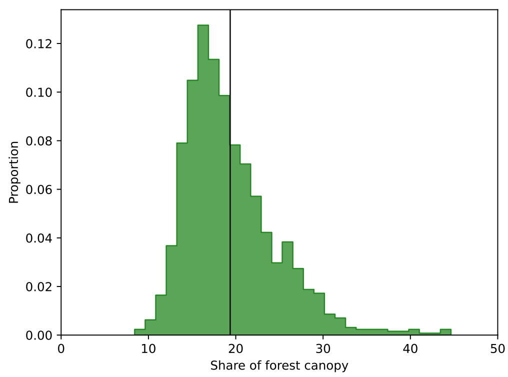
(b) Luembe



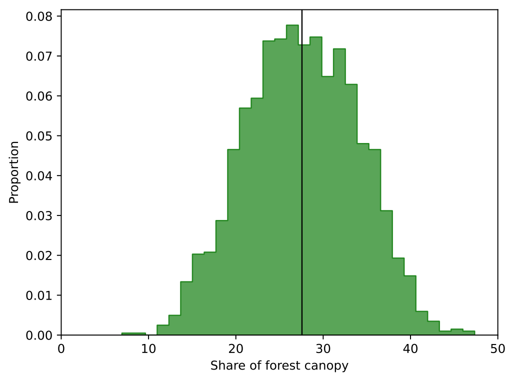
(c) Malama



(d) Msoro



(e) Mwanya



(f) Nyalugwe

Note: This Figure presents the 0.1 degree cell share of tree cover distributions of Chiefdoms surveyed by USAID in 2015 and 2024. The x-axis corresponds to the cell share of forest canopy, i.e., the proportion of the cell which is populated by crowns of trees. The y-axis corresponds to the proportion of cells within a share of canopy bin.

Table 1: Eastern Province GMA and REDD+ PA summary statistics

	GMA	REDD+ PA
Cell canopy share in 2000 (%)	22.40	24.09
	(6.97)	(7.48)
Tree cover in 2000 (ha)	31.50	34.03
	(10.65)	(11.39)
Tree cover loss in ha (2001-2014)	0.68	0.08
	(1.62)	(0.33)
Altitude (m)	697.87	677.82
	(136.81)	(121.03)
Maize Potential Yield (kg/acre)	3100.74	3117.92
	(67.19)	(65.34)
Distance to human settlements (km)	1.30	5.02
	(1.91)	(3.46)
Distance to roads (km)	3.98	11.91
	(5.41)	(9.61)
Distance to Electrical network (km)	45.77	85.85
	(39.71)	(58.79)

Note: This table present summary statistics for Game Management Areas (GMA) and Protected areas (PA) in Eastern Province. The table present the mean value for cells within GMAs and PAs, with exception of Tree cover loss in ha, which corresponds to them sum of tree cover loss in HA for a given 0.01$\times$ 0.01 cells. Standard errors are presented in parenthesis.

Table 2: Correlation between tree cover loss and program offer within treated chiefdoms

	All Chiefdom	Only PA	Only Non PA
After CFP	0.016	-0.246***	0.159**
	(0.050)	(0.040)	(0.067)
R^2	0.34	0.04	0.34
Observations	252,196	89,334	162,862
Chiefdom FE	Yes	Yes	Yes

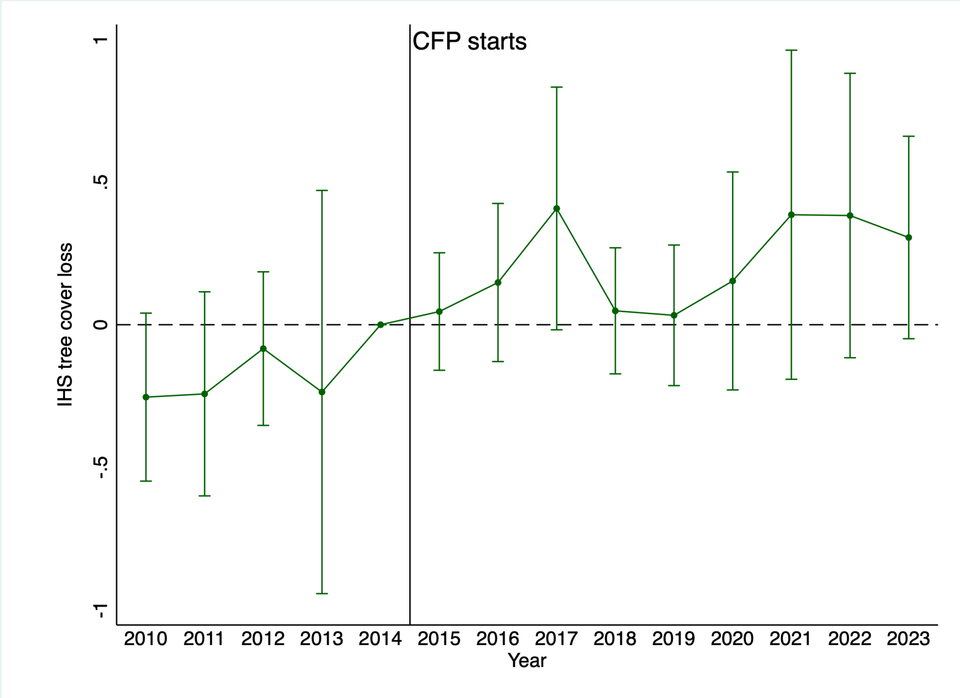
*Note: The dependent variable is the inverse hyperbolic sine of the deforested area in each 0.01$\times$0.01 cell for each year. Column 1 presents the coefficient estimates using all cells in treated chiefdom. Column 2 shows results for cells within protected areas, while Column 3 displays results for cells in non-protected areas within treated chiefdom. Standard errors are reported in parentheses, and errors are clustered at the chiefdom level. *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively*

Table 3: Estimates for the impact of CFP on tree cover loss

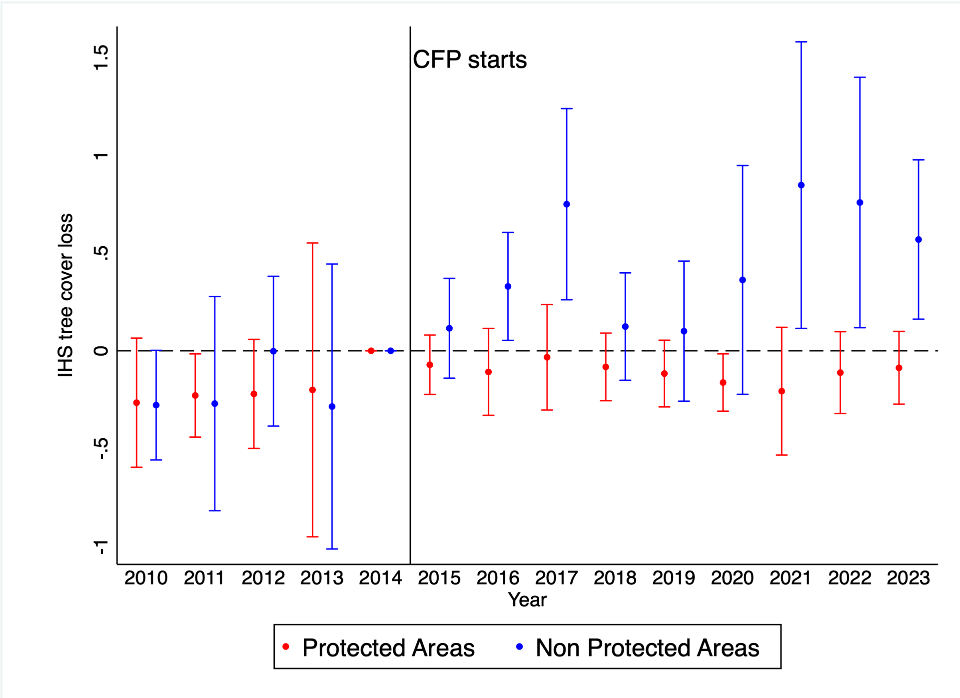
	All Chiefdom	Only PA	Only Non PA
After CFP	0.376***	0.074	0.608***
	(0.068)	(0.094)	(0.182)
R^2	0.09	0.11	0.11
Observations	285,068	285,068	285,068
Chiefdom FE	Yes	Yes	Yes

*Note: The dependent variable is the inverse hyperbolic sine of deforested area in each 0.01×0.01 cell each year. Column 1 shows the regression results for chiefdom cells, not distinguished by protection status. Column 2 and 3 shows the coefficients for protected areas and non protected areas obtained in a regression that separate chiefdom cells into these categories. The comparison group includes cells in non treated chiefdom. Standard deviations are reported in parenthesis and errors are clustered at the chiefdom level. *,**,*** denote statistical significance at the 10%, 5%, and 1% level, respectively.*

Fig. 4. Event study for Treated Chiefdoms and by protection type



(a) Treated Chiefdoms



(b) Protected areas and Non Protected areas

Note: The graphs plot coefficients from treated chiefdom and protection type event study. 4a presents the yearly coefficients for cells within treated chiefdom, and 4b for cells within protected and non protected areas. Yearly coefficients have vertical lines for standard errors. The vertical black line indicates the beginning of the CFP program.

References

Abman, Ryan, and Clark Lundberg. 2024. "Contracting, Market Access and Deforestation." *Journal of Development Economics* 168 (May): 103269. https://doi.org/10.1016/j.jdeveco.2024.103269. https://www.sciencedirect.com/science/article/pii/S030438782400018X.

Acemoglu, Daron, Tristan Reed, and James A. Robinson. 2014. "Chiefs: Economic Development and Elite Control of Civil Society in Sierra Leone." *Journal of Political Economy* 122, no. 2 (April): 319–368. https://doi.org/10.1086/674988. https://www.journals.uchicago.edu/doi/abs/10.1086/674988.

Amin, A., J. Choumert-Nkolo, J. -L. Combes, P. Combes Motel, E. N. Kéré, J. -G. Ongono-Olinga, and S. Schwartz. 2019. "Neighborhood Effects in the Brazilian Amazônia: Protected Areas and Deforestation." *Journal of Environmental Economics and Management* 93 (January): 272–288. https://doi.org/10.1016/j.jeem.2018.11.006. https://www.sciencedirect.com/science/article/pii/S0095069616303163.

Baland, Jean-Marie, Pranab Bardhan, Sanghamitra Das, and Dilip Mookherjee. 2010. "Forests to the People: Decentralization and Forest Degradation in the Indian Himalayas." *World Development* 38, no. 11 (November): 1642–1656. https://doi.org/10.1016/j.worlddev.2010.03.007. https://www.sciencedirect.com/science/article/pii/S0305750X10000896.

Balboni, Clare, Aaron Berman, Robin Burgess, and Benjamin A. Olken. 2023. "The Economics of Tropical Deforestation." *Annual Review of Economics* 15, no. Volume 15, 2023 (September): 723–754. https://doi.org/10.1146/annurev-economics-090622-024705. https://www.annualreviews.org/content/journals/10.1146/annurev-economics-090622-024705.

Baldwin, Kate. 2013. "Why Vote with the Chief? Political Connections and Public Goods Provision in Zambia." *American Journal of Political Science* 57, no. 4 (October): 794–809. https://doi.org/10.1111/ajps.12023. https://onlinelibrary.wiley.com/doi/10.1111/ajps.12023.

———. 2015. *The Paradox of Traditional Chiefs in Democratic Africa.* Cambridge Studies in Comparative Politics. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781316422335. https://www.cambridge.org/core/books/paradox-of-traditional-chiefs-in-democratic-africa/A0A5A0A8B513F792F501C82AF5935457.

Baragwanath, Kathryn, and Ella Bayi. 2020. "Collective Property Rights Reduce Deforestation in the Brazilian Amazon." *Proceedings of the National Academy of Sciences* 117, no. 34 (August): 20495–20502. https://doi.org/10.1073/pnas.1917874117. https://www.pnas.org/doi/full/10.1073/pnas.1917874117.

BCP. 2019. *Community Forests Program, Final Report 2014-2019.* Technical report. Biocarbon Partners.

Burgess, Robin, Matthew Hansen, Benjamin A. Olken, Peter Potapov, and Stefanie Sieber. 2012. "The Political Economy of Deforestation in the Tropics*." *The Quarterly Journal of Economics* 127, no. 4 (November): 1707–1754. https://doi.org/10.1093/qje/qjs034. https://doi.org/10.1093/qje/qjs034.

Cattaneo, Matias D., Nicolás Idrobo, and Rocío Titiunik. 2019. "A Practical Introduction to Regression Discontinuity Designs: Foundations." *Elements in Quantitative and Computational Methods for the Social Sciences* (November). https://doi.org/10.1017/9781108684606.

https://www.cambridge.org/core/elements/practical-introduction-to-regression-discontinuity-designs/F04907129D5C1B823E3DB19C31CAB905.

Cisneros, Elías, Jan Börner, Stefano Pagiola, and Sven Wunder. 2022. "Impacts of Conservation Incentives in Protected Areas: The Case of Bolsa Floresta, Brazil." *Journal of Environmental Economics and Management* 111 (January): 102572. https://doi.org/10.1016/j.jeem.2021.102572. https://www.sciencedirect.com/science/article/pii/S0095069621001200.

Cisneros, Elías, and Krisztina Kis-Katos. 2022. *Unintended Consequences of Anti-Corruption Strategies.* SSRN Scholarly Paper. Rochester, NY, March. https://doi.org/10.2139/ssrn.3899498. https://papers.ssrn.com/abstract=3899498.

Correa, Juliano, Elías Cisneros, Jan Börner, Alexander Pfaff, Marcelo Costa, and Raoni Rajão. 2020. "Evaluating REDD+ at Subnational Level: Amazon Fund Impacts in Alta Floresta, Brazil." *Forest Policy and Economics* 116 (July): 102178. https://doi.org/10.1016/j.forpol.2020.102178. https://www.sciencedirect.com/science/article/pii/S1389934119301170.

Giudice, Renzo, Jan Börner, Sven Wunder, and Elias Cisneros. 2019. "Selection Biases and Spillovers from Collective Conservation Incentives in the Peruvian Amazon." *Environmental Research Letters* 14, no. 4 (April): 045004. https://doi.org/10.1088/1748-9326/aafc83. https://dx.doi.org/10.1088/1748-9326/aafc83.

Gulzar, Saad, Apoorva Lal, and Benjamin Pasquale. 2024. "Representation and Forest Conservation: Evidence from India's Scheduled Areas." *American Political Science Review* 118, no. 2 (May): 764–783. https://doi.org/10.1017/S0003055423000758. https://www.cambridge.org/core/journals/american-political-science-review/article/representation-and-forest-conservation-evidence-from-indias-scheduled-areas/BF8BB4821C97E6E67E682D71345DA0FE.

Hansen, James, Makiko Sato, Gary Russell, and Pushker Kharecha. 2013. "Climate Sensitivity, Sea Level and Atmospheric Carbon Dioxide." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 371, no. 2001 (October): 20120294. https://doi.org/10.1098/rsta.2012.0294. https://royalsocietypublishing.org/doi/10.1098/rsta.2012.0294.

Jack, B. Kelsey, Seema Jayachandran, Namrata Kala, and Rohini Pande. 2022. *Money (Not) to Burn: Payments for Ecosystem Services to Reduce Crop Residue Burning.* Working Paper, November. https://doi.org/10.3386/w30690. https://www.nber.org/papers/w30690.

Jayachandran, Seema. 2022. *The Inherent Trade-Off Between the Environmental and Anti-Poverty Goals of Payments for Ecosystem Services.* Working Paper, April. https://doi.org/10.3386/w29954. https://www.nber.org/papers/w29954.

Malan, Mandy, Rachel Carmenta, Elisabeth Gsottbauer, Paul Hofman, Andreas Kontoleon, Tom Swinfield, and Maarten Voors. 2024. "Evaluating the Impacts of a Large-Scale Voluntary REDD+ Project in Sierra Leone." *Nature Sustainability* 7, no. 2 (February): 120–129. https://doi.org/10.1038/s41893-023-01256-9. https://www.nature.com/articles/s41893-023-01256-9.

Sims, Katharine R. E. 2010. "Conservation and Development: Evidence from Thai Protected Areas." *Journal of Environmental Economics and Management* 60, no. 2 (September): 94–114. https://doi.org/10.1016/j.jeem.2010.05.003. https://www.sciencedirect.com/science/article/pii/S0095069610000586.

Sze, Jocelyne S., L. Roman Carrasco, Dylan Childs, and David P. Edwards. 2022. "Reduced Deforestation and Degradation in Indigenous Lands Pan-Tropically." *Nature Sustainability* 5, no. 2 (February): 123–130. https://doi.org/10.1038/s41893-021-00815-2. https://www.nature.com/articles/s41893-021-00815-2.

USAID. 2016. *Report on Baseline Findings, Community Forest Management Program.* Technical report. United States Agency for International Development (USAID).

Watson, Charlene, Liane Schalatek, and Aurélien Evéquoz. n.d. "Climate Finance Thematic Briefing: REDD+ Finance."

West, Thales A. P., Jan Börner, Erin O. Sills, and Andreas Kontoleon. 2020. "Overstated Carbon Emission Reductions from Voluntary REDD+ Projects in the Brazilian Amazon." *Proceedings of the National Academy of Sciences* 117, no. 39 (September): 24188–24194. https://doi.org/10.1073/pnas.2004334117. https://www.pnas.org/doi/full/10.1073/pnas.2004334117.

ONLINE APPENDIX

Strategic Environment: Conservation Policies Effectiveness and Strategic Behavior

Angelo Santos and Heather Huntington

A Additional Data Details (Online)

A.1 Tree cover data

To measurement tree cover and tree cover loss, I used the Global Forest Change (GFC) (Hansen et al. 2013) from the University of Maryland. This dataset was created based on a set of Landsat images since 2000, providing yearly layers on tree cover change for the all globe. This datasets is publicly available for download and can be used in the Google Engine platform for visualizations and data processing.

The dataset has a spatial resolution of 1 arc-second per pixel, or approximately 30 meters per pixel at the equator. We use two layers of tree cover information. First, I use the tree canopy cover for year 2000. This defines the share of forest canopy for each 30x30m grid, defining tree as any vegetation taller than 5m in height. The information is the percentage per grid cell of forest canopy, ranging from 0-100. This layer is my baseline data and there is no year updates on the forest canopy share per grid, which is a caveat of using the GFC data. The second layer is measure tree cover loss defined as a stand-replacement disturbance, or a change from a forest to non-forest state at the grid level. The cell level information range from 0 to 20, where 0 represents no loss and 1 to 23 representing loss detected primarily in the year 2001-2023, respectively.

In addition to these layers, the GFC have 4 more layers. There is a layer of forest cover gain from 2000- to 2012, which is defined as the inverse of loss, or a non-forest to forest change entirely within the period. The information is encoded as 1 (gain) or 0 (no gain). Another available is the data mask for cells where 0 represents areas with no data, 1 for mapped land surface, and 2 for persistent water bodies based on 2000 to 2012. The last two layers is the Circa year 2000 Landsat 7 cloud-free image composite, and the Circa year 2023 Landsat cloud-free image composite. The first contains reference multispectral imagery from the first available year, typically 2000. The last contains reference multispectral imagery from the last available year, typically 2023.

The dataset is downloaded by 10x10 degree combinations through the website. To process the dataset for my location of interest, I have determined a latitude-longitude box and filtered the 10x10 degree datasets. After filtering the layers, I overlapped locations boundaries (chiefdom, protected areas and village areas) to aggregate 30x30m to the 0.1 degree cell level combinations.

A.2 Fires data

To measure the fires events I will use the Fire Information for Resource Management System (FIRMS) from NASA. This dataset provides Near Real-Time (NRT) active fire data using images from the Moderate Resolution Imaging Spectroradiometer (MODIS) and Visible Infrared Imaging Radiometer Suite (VIIRS). The NRT data is available within 3 hours of satellite observations, with exception of the US and Canada which have real-time data. The resolution is 375m and the data is available since 20 January 2012. This dataset is publicly available for the all globe and can be downloaded through the NASA Earthdata API.

Each 375m pixel has attributes on fire outcomes and technical information. The centroid of the 375m pixel is informed by latitude and longitude coordinates. To measure fire intensity, the dataset informs two scales of temperature in Kelvin reflected by channel brightness, Brightness temperature I-4 and Brightness temperature I-5. The information related to the timing of event is the date, time, day or night category. About the confidence of the measure, there are three groups: Low, Nominal and High confidence. This is determined by the fire intensity and sun glint. Low-confidence detection are typically linked to regions affected by Sun glint and have lower relative temperature anomalies (less than 15 K) in the mid-infrared channel I4. Fire Radiative Power (FRP) is another measurement of fire intensity, that are based on a hybrid approach combining 375 and 750 m data. Nominal-confidence detection are free from potential Sun glint contamination during the day and exhibit strong temperature anomalies (greater than 15 K) in either day or nighttime data. High-confidence detections are associated with saturated pixels, whether during the day or at night. More technical information include the satellite source of observation, which could be Suomi National Polar-orbiting Partnership (Suomi NPP), NOAA-20 (formally JPSS-1), and NOAA-21. Another source of technical information is the version of the data.

A.3 Protected areas

The protected areas information used in the paper comes from the Planet protected data on protected areas and other effective area-based conservation measures (OECMs). This dataset contains the most complete and updated information on terrestrial and marine protected areas around the globe. The database contains shapefiles that informs multiple characteristics of these areas, as location, type of protected area, ownership, forest management plan existence, and other characteristics.

An important information concerned these areas is the designation, which includes National Parks, Game Management Areas (GMAs), and Forest reserves. This is a relevant to know as the rules relative to forest usage differ conditional on the designation type. For instance, Forest reserves do not have strict rules relative to forest usage as National Parks are strictly monitored and prohibit activities that can harm its biodiversity, and human settlements. The GMAs in zambia are more flexible in terms of the rules, allowing Human settlements, and focus on rules for sustainable hunting in the area. According to the dataset, Zambia contains 20 National parks, 36 GMAs, and 464 Forest reserves.

A.4 Zambia Rural Agricultural Livelihoods Survey (RALS)

I use the Zambia Rural Agricultural Livelihoods Survey (RALS), which provides comprehensive information on rural households across Zambia. This dataset is available under request and confidentiality agreement with the Indaba Agricultural Policy Research Institute (IAPRI).

The survey was has been conducted in 2012, 2015, and 2019 and serves as an important tool for understanding the dynamics of rural livelihoods, agricultural production, and rural development challenges in Zambia. RALS is a partnership between the Indaba Agricultural Policy Research Institute (IAPRI), the Central Statistical Office of Zambia, and the Ministry of Agriculture.

The sample of the study include small and medium farmers, i.e. cultivating less than 20 Ha and follows the Zambia 2010 Census sample. The dataset has 17 sections in 2019 and includes detailed data on agricultural practices, household demographics, livestock ownership, and economic activities. This dataset covers multiple dimensions of rural livelihoods, such as crop production, income sources, and access to essential services like credit and extension services. The survey's rich demographic and economic data, combined with its geographic coordinates on households locations, allows to map and construct geographical indicators.

A.5 GRID3 Settlement Mapping

The GRID3 Settlement Mapping for Zambia provides detailed geospatial data identifying the location and extent of human settlements throughout the country. This mapping initiative combines satellite imagery with population data, helping to locate both urban and rural communities with high accuracy. The data captures the physical boundaries of settlements, which include residential areas, infrastructure, and other human-made structures. With the location of these settlement it is possible to construct village level outcomes by allocating cells to the closest settlement according to the data. This is done by creating Voronoi polygons.

A.6 Cell aggregation

My main analysis rely on 0.01×0.01 degrees cell grids, which corresponds to approximately 1km$\times$1km. I created these grids using the function X in python, that rounds a given number to the Y border. I use this function for both latitude and longitude coordinates of a particular cell and create an geoid. The geoid consist of the concatenation of the rounded latitude and longitude strings. For example, a cell located at -z and l will have the following geoid "$-z_l$".

After creating the 0.01x0.01 geoids, I can aggregate cell values using aggregation functions. For example, I can compute the tree cover loss for each geoid in a particular year by using the sum aggregation function for the tree cover loss for smaller grids associate to this id. For other variables I take the mean, as tree cover in 2000.

A.7 Geographical matching

To create multiple cell-level information I used geographical matching. I use an open source package in Python called GeoPandas, which contains functions that perform geospatial operations. To match the multiple layers I use two methods I use the function sjoin which overlap layers and match it geographically. For example, using the shapefile of the Eastern province I can overlap with all the cells within the region and associate to it the province level data.

B Additional Method Details (Online)

In this section, I give more details about additional methods that will be using to obtain more local effects, (RDD) and village level effects (DiD).

REDD+ protected areas effects - Pure RDD To estimate the impact of the protected areas initiated by the CFP program, I will use a Geographical Regression Discontinuity Design (RDD). There are three main elements that define this design: a score s, a cutoff c, and bandwidth choices (formalized below). The identification strategy rely on the fact that cells closer to the cutoff c are similar on observable and unobserved characteristics with the difference of the new protection boundaries. In my context, cells within the protected areas faces a discrete change on the likelihood of being deforested after the conservation boundary implementation. This likelihood can be reflected by the score s of cells, which is a function of the distance to the protected areas borders. Going further from the border of the PAs, the cells tends to be differentiated in multiple dimensions and decrease on the score associated to the likelihood of being deforested. This highlights the importance of bandwidth h choice around the cutoff, as this influences the comparability of cells within the protected areas with contractual outside the boundaries. The optimal bandwidth is estimated using non-parametric methods following Cattaneo, Idrobo, and Titiunik (2019) which minimizes the Mean Squared Error (MSE) influenced by the estimator bias-variance tradeoff.

To formalize the RDD, I define cell i and n_h as the total number of cells observed within the bandwidth h, where $i = 1, 2, ..., n_h$. In addition, $X_i = \{lat_i, long_i\}$ is the centroid of cell i. The cutoff c is the coordinates $\{lat_c, long_c\}$ for the closest PA defined by the program. Treatment is defined by $T_i \in 0, 1$ where 1 indicates that cell i is treated, which happens when score $s = 1$. This includes cell for which the distance between cell i coordinates and the closest PA is positive ($X_i - c \geqslant 0$) and $c - h \leqslant X_i \leqslant c + h$, where h corresponds to the optimal bandwidth. If the distance is negative ($X_i - c \leqslant 0$) and $c - h \leqslant X_i \leqslant c + h$, the cell score is zero ($s = 0$) implying no treatment ($T_i = 0$). This imply that this cell is located outside the protected areas defined by CFP.

To estimate the local average treatment effects (LATE) τ_h of the REDD+ PAs on $cell_i$ deforestation rate conditional on bandwidth choice h, I will follow Cattaneo, Idrobo, and Titiunik (2019) which proposes a non-parametric method to identify the LATE parameter. I will add covariates which should not change the LATE magnitude but add precision to the estimator, the errors are clustered at X by X km grid groups. The reduced form estimation is the following:

$$y_i(h) = \alpha + \tau_h T_i + \beta_{1,h} f(X_i - c) + Z_i'\gamma + \epsilon_i \tag{B.1}$$

Where y_i corresponds to cell i deforestation, τ_h is the average treatment effect of the CFP protected areas, T_i is a dummy variable equal to 1 if the cell is within the borders of the PAs areas. The h parameter indicates that the estimation is performed conditional on cells being located within the optimal bandwidth h. In conclusion, $f(X_i - c)$ is a functional form used to fit the data variation, and Z_i is a vector of cell characteristics as terrain, institutional, and infrastructure information.

The effects of interest (τ) is identified by the following:

$$\tau = \mathbb{E}[Y(1) - Y(0)|X = c] = \lim_{x \downarrow c} \mathbb{E}[Y|X = x] - \lim_{x \uparrow c} \mathbb{E}[Y|X = x] \tag{B.2}$$

This indicates that the local average treatment effect of CFP (τ) can be identified by the difference between deforestation rates of treated cells (protected areas) and control cells (unprotected areas) around the cutoff c. The limit indicates that this parameter is the vertical difference between observations at the cutoff. This rely on a potential outcomes framework that, under the RDD assumptions, implies that $Y(0)$ are proper counterfactual for $Y(1)$ outcomes without the establishment of the new protected areas.

Furthermore, I verify heterogeneous effects estimated by cell subgroups as suggested by Cattaneo, Idrobo, and Titiunik (2019). This consists of estimating different local treatment effects (τ_g) for cells conditional on different subgroups (g) defined by cell characteristics in baseline year 2015. Exploring the USAID surveys, I will test heterogeneous effects for cell proximity to treated villages , and villages/chiefdom characteristics related to forest dependence and social norms. Using supplementary data I will check for effects conditional on road and infrastructure proximity to account for market access and different costs of deforestation. Using remote data on land characteristics as slope, soil quality and agriculture feasibility, I test how these effects differ conditional on terrain characteristics.

To test if the RDD identification assumptions holds, I perform a series of validation and falsification tests proposed by (Cattaneo, Idrobo, and Titiunik 2019). I verify the covariates continuous assumption around the cutoff and some placebo outcome test. This aims to verify if the discontinuity observed around the cutoff only applies to deforestation rates. If there are substantial discontinuities on covariates, this indicates that there are other aspects that can impact the different levels of deforestation observed between areas. For the placebo outcomes test, if there are discrete jumps on outcomes not related to the objectives of the program (for example, soil agriculture feasibility) this would indicate that other aspects are influencing the context during the program period. I also test results sensibility to cutoff and bandwidth choices, and frequency of observations around the optimal cutoff.

Village level outcomes To estimate the program effect on village outcomes, I will use a Difference-in-Differences (DiD) approach. The exogeneity assumption is that the deforestation rates would keep the same trend for treatemnt and control in the absence of the program. The geographic unit of analysis is the surrounding of the villages surveyed by USAID. I choosed a 5 km buffer due the report of the majority of househould on forest access. In the household survey, 90 % of the surveyed hhs report less than 5 km from the forest used to collect or produce goods. Related to agriculture fields, y % report that these fields are h km from them. The same happens in the village headperson survey, where they report the distance to forests used by the community. This suggests that choosing a 5 km buffer around these villages is a reasonable assumption to capture effects due village behavior. However, I used different buffer size to verify if my results are sensitive to its extension. Another concerning about the buffer size is the possibility of having overlapping areas. To deal with this possibility, I assigned weights to cells that are around more than 1 village buffer. The weight is simple the inverse of the number of villages sharing a particular cell, i.e. $1/\#ofvillages$. I show that my results are not sensitive to using weights to measure deforestation rates.

I will leverage the proximity to the protected areas as an intensity of treatment, as the program may. The reduced form is estimated as follows:

$$y_{vy} = \beta_0 + \beta_1 \mathsf{CFP}_{vy} + \alpha_y + \delta_v + \epsilon_{vt} \ , \tag{B.3}$$

where y corresponds to the environmental outcomes around village v in year y, CFP_{vy} is dummy variable equal to 1 if the village is included in one of the treated chiefdom and is located within x km from the protected areas. $alpha_y$ and δ_v are year and village fixed effects included to control for time-invariant village and year characteristics that can impact environmental outcomes. β_1 informs the impact of offering the CFP program to villages closer to the protected areas on environmental outcomes.

I also run an event study for the village level computing yearly coefficients for 5km buffers around treated villages. I estimate the following:

$$y_{ivct} = \alpha + \sum_{k} \beta_k \mathbb{I}(t = k) CFP_v c + \delta_t + \gamma_c + \epsilon_{ict} \ , \tag{B.4}$$

In this equation $y_i ct$ corresponds to inverse hyperbolic sine of tree cover loss within village i 5km buffer in chiefdom c at year t. I regress this on yearly coefficients (β_k) before and after the program implementation. The coefficients will give me the yearly cumulative effect of the program on deforestation outcomes. I use time (δ_t) and chiefdom (γ_c) fixed effects to control for time-invariant unobservables characteristics that may impact tree cover loss. ϵ_{ict} is a idiosyncratic error term, and I clustered errors at the chiefdom level to account for serial correlation between cells.

To test for heterogeneous effects, I will interact the CFP dummy with social norms and forest dependence variables. For social norms, I will estimate the following:

$$y_{vy} = \beta_0 + \beta_1 CFP_{vy} + \beta_2 Social\_Norms + \beta_3 CFP_{vy} \times Social\_Norms + \alpha_y + \epsilon_{vt} \tag{B.5}$$

For forest dependence, I will estimate the following:

$$\begin{aligned} y_{vy} = \beta_0 + \beta_1 CFP_{vy} + \beta_2 Forest\_Dependence + \\ \beta_3 CFP_{vy} \times Forest\_Dependence + \alpha_y + \epsilon_{vt} \end{aligned} \tag{B.6}$$

The standard errors are clustered at the village level to take into account serial correlation between.

C Additional Results (Online)

In this section I will present more results

C.1 Distributional analysis

In this section I will discuss some descriptive facts about the environmental outcomes in Eastern Zambia. Specifically, I discuss deforestation rates and tree cover before the program implementation. I will proceed with the description using different types of geographical units, as chiefdoms, villages and REDD+ protected areas.

Deforestation rates in Zambia Figure **??** shows the distribution of 0.1 degree cell share of forest canopy, by chiefdom. On the x-axis we have the share of forest canopy for 0.1 degree cells, on the y-axis I plot the proportion of cells for each share of forest bin. I also plot the mean share of forest canopy for each location.

These figures clarify how the classification of cells as forested depends on the minimum threshold of forest canopy for considering a cell as forested. The mean share of canopy of these cells is between 18% and 28 %, with the distributions concentrated between 5% and 50%. This suggests that the vegetation in Eastern Zambia is mainly sparse forest.

Contributing to what discussed before, figure D.1 shows the distribution of cell proportion forested conditional on different minimum threshold of tree canopy for forest categorization, by chiefdom. Focusing on the blue distributions, minimum threshold 10% of forest canopy, big proportion of the cell have more than half of the cell considered as forest. However, increasing the threshold shifts the distribution to the left, indicating that the proportion of cells with considerable forested shares almost disappear.

I also plotted the tree cover in 2000 around villages included in the USAID survey. Figure D.3 shows the tree cover distribution around 5km buffers from villages, by distance from the protected areas. The figure suggests that villages 5-10 km from the borders of the protected areas has an slightly different distribution. This villages have more cell with higher shares of canopy.

An important measure related to deforestation is the number of fires alerts captured by satellites. Using the FIRMS NASA dataset I plot the distribution of the fires alerts by chiefdom in figure D.4. The distributions show a seasonal trend concentrating fires alerts between June and November in all the chiefdoms. The same can be seen when plotting a distribution for all the chiefdoms in figure D.5.

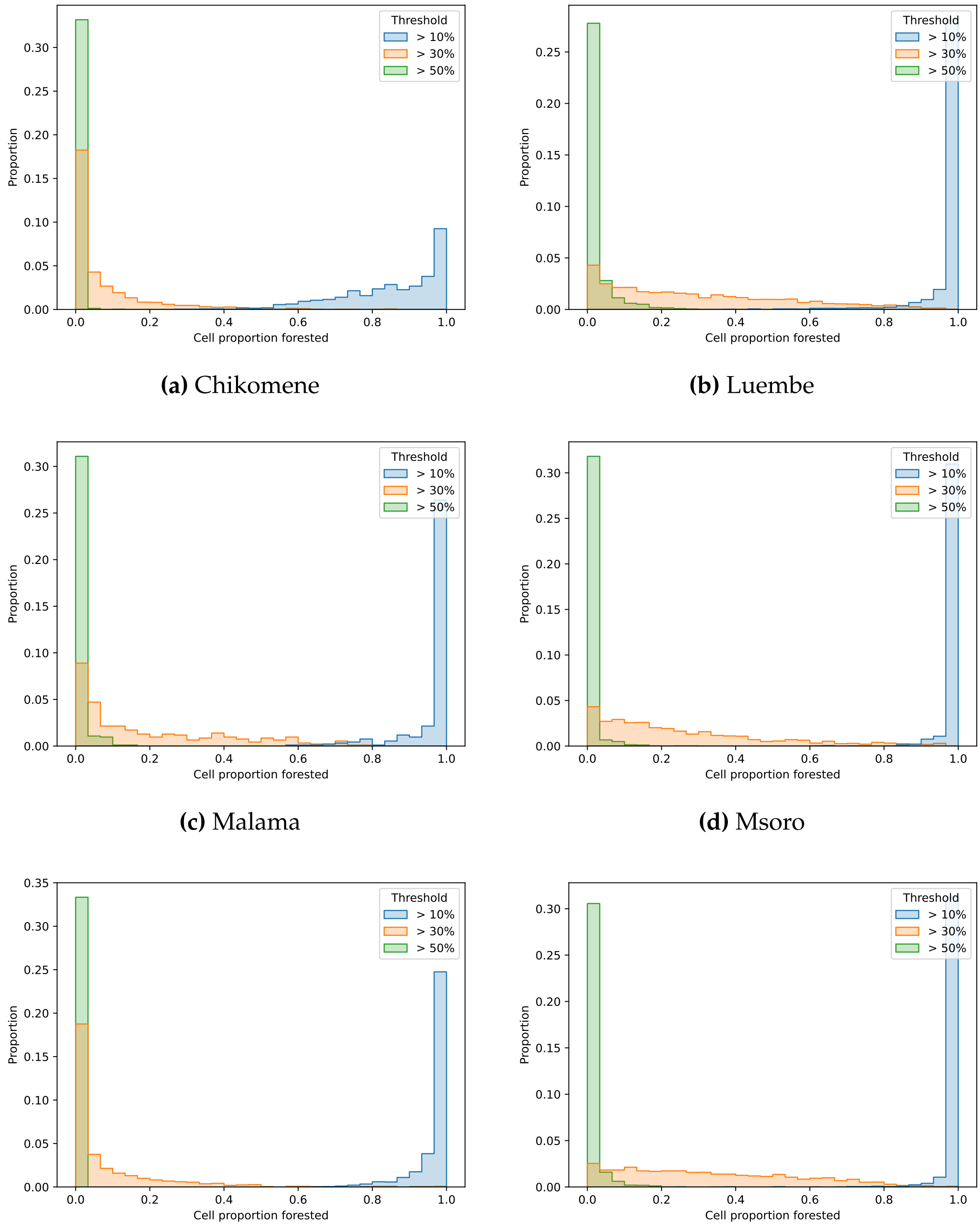
D Appendix Figures and Table (Online)

Table D.1: Summary statistics by Province

	Central	Copperbelt	Eastern	Luapula	Lusaka	Muchinga	North-Western	Northern	Southern	Western	Zambia
Cell canopy share in 2000 (%)	21.319	37.278	20.571	34.473	18.866	21.944	42.132	24.250	11.765	15.723	24.717
	(9.80)	(13.15)	(8.75)	(24.70)	(7.83)	(12.36)	(14.88)	(16.32)	(5.96)	(11.19)	(16.49)
Tree cover in 2000 (ha)	30.198	52.572	29.094	48.634	26.584	31.317	59.973	34.523	16.733	22.393	35.101
	(14.31)	(19.91)	(12.94)	(35.42)	(11.70)	(17.97)	(22.17)	(23.49)	(8.68)	(16.10)	(23.81)
Tree cover loss in ha (2001-2014)	1.085	4.641	0.940	2.529	0.514	0.611	1.684	1.064	0.377	0.742	1.219
	(2.44)	(8.16)	(1.84)	(5.14)	(1.57)	(1.39)	(4.62)	(2.98)	(0.83)	(2.46)	(3.50)
Average rate of tree cover loss (2001-2014)	0.034	0.087	0.034	0.044	0.021	0.018	0.027	0.027	0.024	0.025	0.030
	(0.06)	(0.13)	(0.06)	(0.07)	(0.05)	(0.04)	(0.07)	(0.05)	(0.05)	(0.06)	(0.07)
Altitude (m)	1129.938	1241.392	862.928	1177.494	856.272	1123.954	1223.777	1283.849	1018.425	1071.827	1120.530
	(197.01)	(57.15)	(206.49)	(122.07)	(277.30)	(303.27)	(113.32)	(193.85)	(215.68)	(60.67)	(217.59)
Maize Potential Yield (kg/acre)	2942.332	3062.984	3109.202	2954.047	2993.621	2868.563	3060.152	2776.137	3045.641	3252.409	3017.646
	(406.83)	(82.12)	(143.69)	(132.21)	(138.96)	(357.14)	(228.24)	(410.42)	(161.61)	(181.82)	(310.06)
Observations	92898	26441	43443	42142	22036	73077	105605	64802	57976	108277	636697

Note: This table present summary statistics for Provinces in Zambia. The table present the mean value for cells within GMAs and PAs, with exception of Tree cover loss in ha, which corresponds to them sum of tree cover loss in HA for a given $0.01\times$ 0.01 cells. Standard errors are presented in parenthesis.

Fig. D.1. Chiefdom distribution of 0.1 degree cell proportion forested



(a) Chikomene

(b) Luembe

(c) Malama

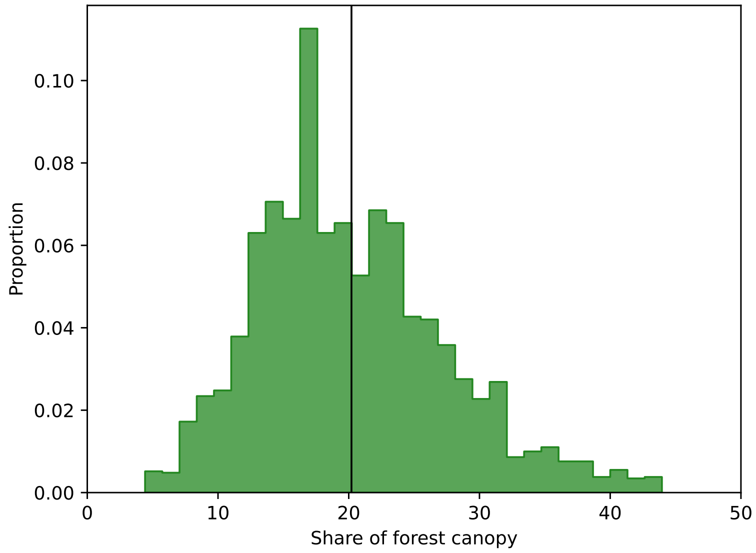
(d) Msoro

(e) Mwanya

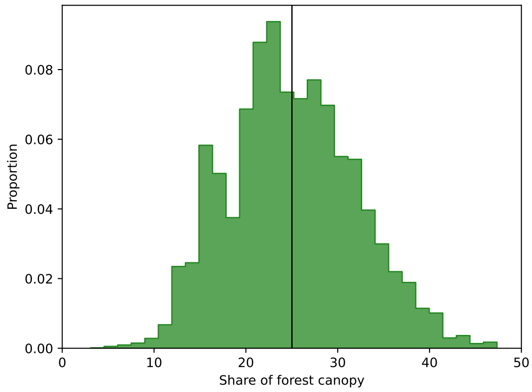
(f) Nyalugwe

Note: This Figure shows the 0.1 degree cell proportion of forested cells conditional on canopy share threshold. The x-axis corresponds to the cell share of forested cells, i.e., the proportion of the cell which is considered forest conditional on different thresholds of forest canopy. The y-axis corresponds to the proportion of cells within a share of forested cell bin.

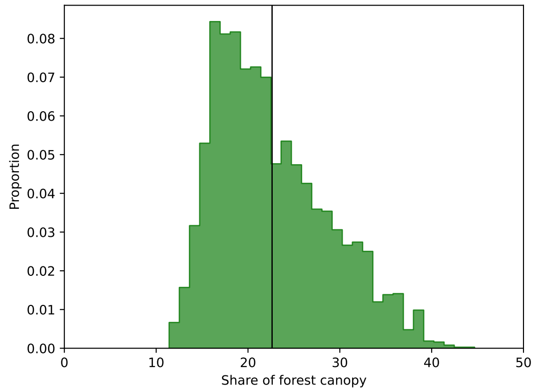
Fig. D.2. Village buffers distribution of 0.1 degree cell share of forest canopy by proximity



(a) 0 to 5 km from REDD+ areas



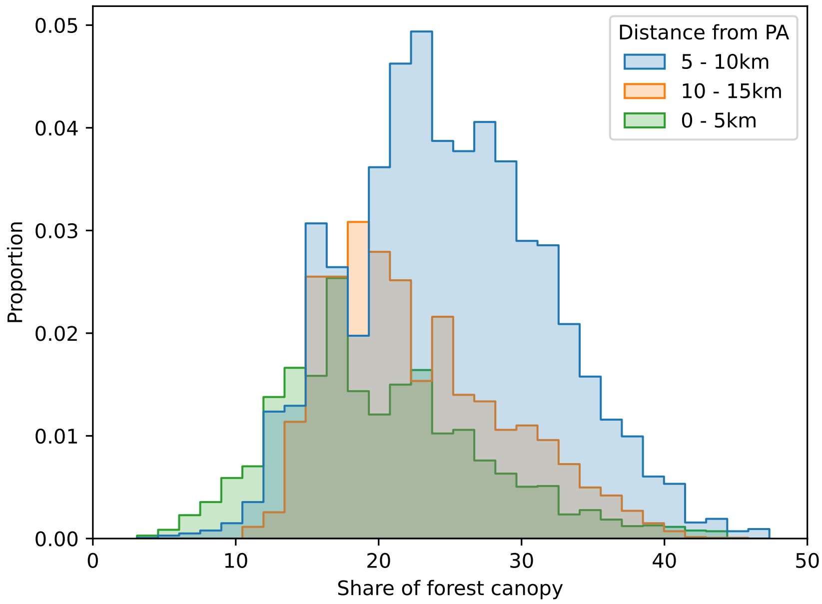
(b) 5 to 10 km from REDD+ areas



(c) 10 to 15 km from REDD+ areas

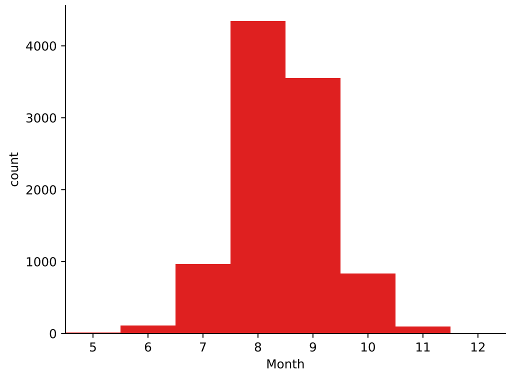
Note: This Figure shows the 0.1 degree cell share of canopy distributions of villages 5km buffer. Figure a plots the distributions for villages within 5 km from the REDD+ protected areas (PA). Figure b and c shows the distributions for 5km to 10km, and 10 - 15km, respectively. The x-axis corresponds to the cell share of forest canopy, i.e., the proportion of the cell which is populated by crowns of trees. The y-axis corresponds to the proportion of cells within a share of canopy bin.

Fig. D.3. Village 5 km buffers distribution of 0.1 degree cell share of forest canopy

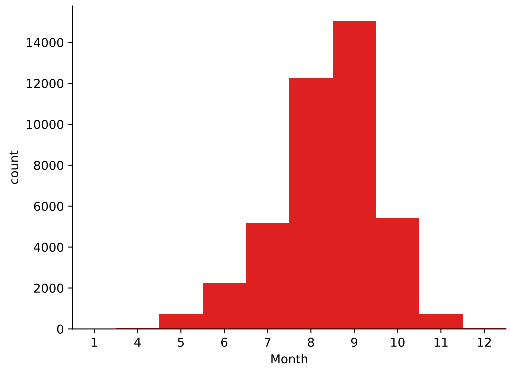


Note: This Figure shows the 0.1 degree cell share of canopy distributions of villages 5km buffer. There are three colors to distinguish villages by proximity to the REDD+ protected areas (PA). Figure D.2 shows these distributions separately. The x-axis corresponds to the cell share of forest canopy, i.e., the proportion of the cell which is populated by crowns of trees. The y-axis corresponds to the proportion of cells within a share of canopy bin.

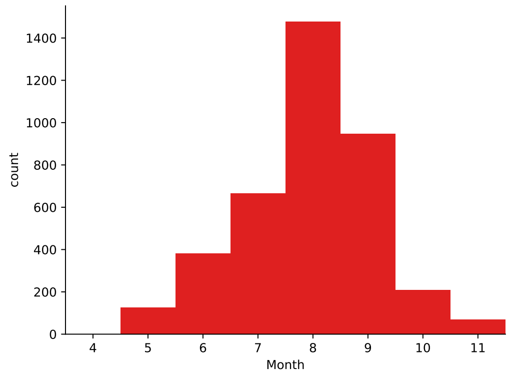
Fig. D.4. Number of fires identified by treated chiefdom.



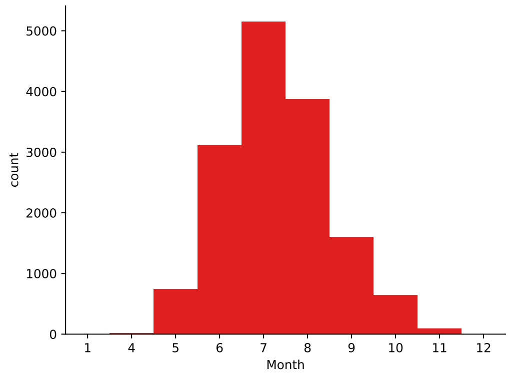
(a) Chikomene



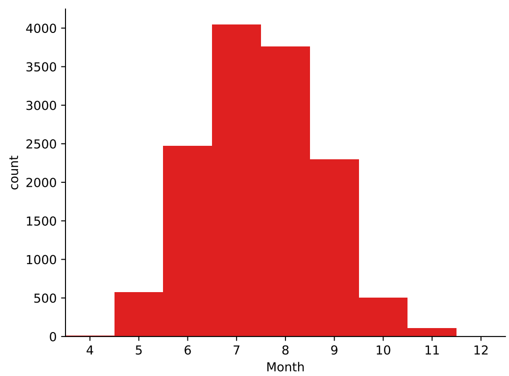
(b) Luembe



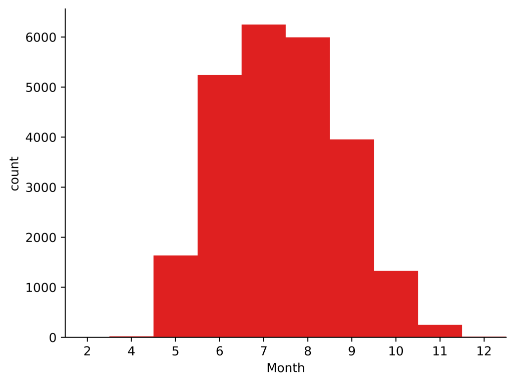
(c) Malama



(d) Msoro



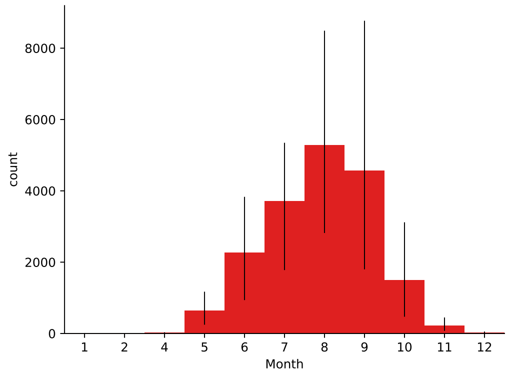
(e) Mwanya



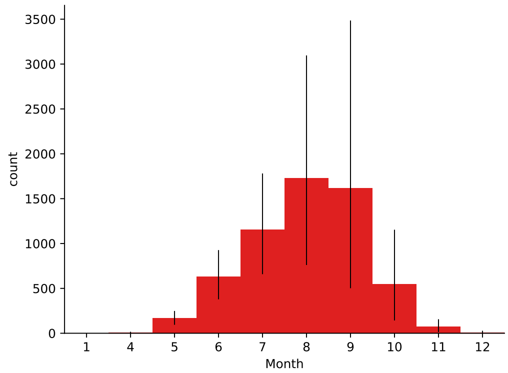
(f) Nyalugwe

Note: This figure shows the number of fires detected within Chiefdom boundaries. The y-axis represents the count of identified fires, while the x-axis displays the month of occurrence.

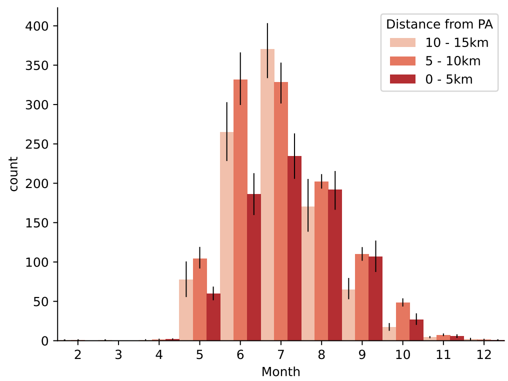
Fig. D.5. Number of fires identified within chiefdom, protected areas and village 5 km buffer.



(a) Chiefdomns



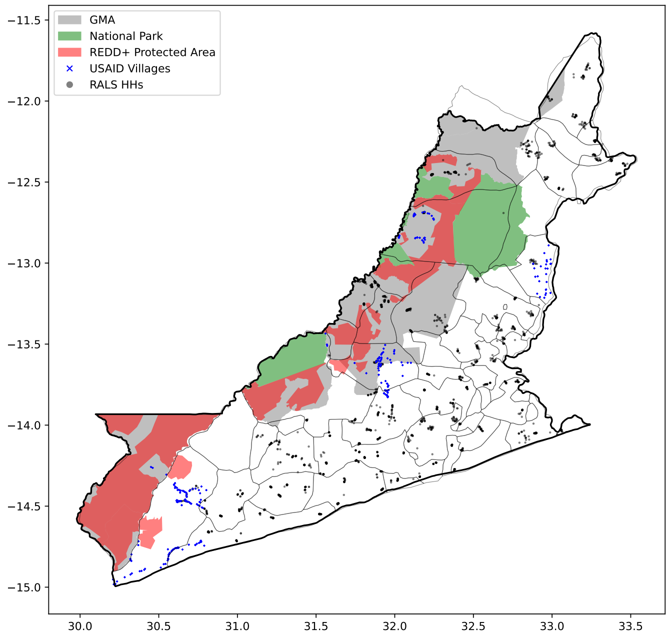
(b) REDD+ protected areas



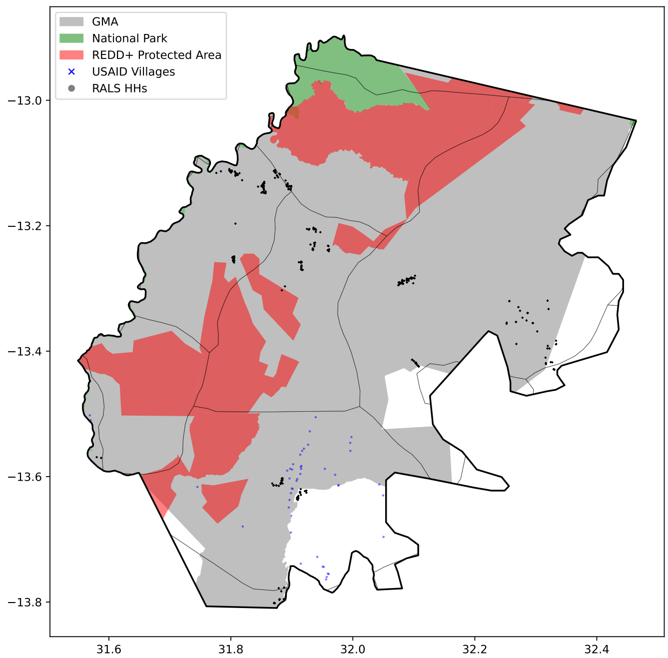
(c) Villages

Note: This figure shows the number of fires detected within treated Chiefdoms, REDD+ protected areas, and 5 km buffers around villages. The y-axis represents the count of identified fires, while the x-axis displays the month of occurrence.

Fig. D.6. Conservation areas and household locations in Eastern Province and Mambwe District



(a) Eastern Province



(b) Mambwe District

Note: These maps illustrate the overlapping of different geographical layers in Eastern Province and location of surveyed Households by USAID and Zambia Statistcal agency. Black lines delineate the region according to chiefdom boundaries. The green layer represents the Natural Parks, the gray layer shows the Game Management Areas, and the red layer indicates the protected areas defined by the REDD+ program. Blue dots indicate the village location according to USAID surveyed Households. Black dots indicate households location from RALS survey.