FOREST CONSERVATION THROUGH VOLUNTARY CARBON OFFSETTING INTERVENTIONS

Alejandro Guizar Coutiño

Wolfson College

University of Cambridge

December 2022

This dissertation is submitted for the degree of Doctor of Philosophy.

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any work that has already been submitted before for any degree or other qualification except as declared in the preface and specified in the text.

It does not exceed the prescribed word limit for the 60,000 Degree Committee

ABSTRACT

Forest conservation through voluntary carbon offsetting interventions

Alejandro Guizar Coutiño

Ongoing deforestation in the tropics is destabilising planetary carbon cycles, impacting biodiversity and affecting societies the world over. Multiple strategies have been put in place to reduce threats to tropical forests through integrated conservation and sustainable development projects. Reducing Emissions from Deforestation and Forest Degradation (REDD+) refers to a diversity of mechanisms to help conserve and restore tropical forests, involving local, subnational or national initiatives to support on-the-ground sustainable management and conservation. REDD+ represents a move to support conservation by rewarding forest owners and users to reduce emissions and enhance forest-based carbon stocks, through bundles of interventions including conditional and non-conditional livelihood enhancement activities together with enabling measures and policies. REDD+ has generated debate among policy and academic circles around its lack of (or negative) impacts on social and environmental outcomes. A decade of research on REDD+ has yielded insights about its performance on multiple fronts, particularly on the social impacts, yet little is still known about its effectiveness on preventing land-cover changes and associated carbon emissions.

This thesis addresses the question of whether REDD+ projects have helped to prevent deforestation. I approached the problem by producing three pieces of research that range from the applied evaluation of REDD+ to interrogating methodological approaches used to evaluate the impact of conservation interventions, using a representative sample of REDD+ projects across multiple countries operating within the voluntary carbon market. The research provides evidence of the impact of local-based REDD+ interventions in reducing deforestation while examining some of the challenges involved in using quasi-experimental methods to evaluate forest-based interventions. **Chapter 2** presents one of the first global assessments of the impact of REDD+ in reducing deforestation using a standard quasi-experimental approach. I found reductions in forest loss due to REDD+, particularly in areas of higher threats of deforestation. **Chapter 3** provides a comparison of matching approaches using state-of-the-art robust evaluation metrics, I found consistent average reductions in deforestation due to REDD+ across multiple matched models, however alterations in matching parameters led to considerable variation of baseline

levels in some projects, thus impacting estimates of treatment effect. **Chapter 4** extends the analyses of REDD+ by measuring local leakage effects and the potential implication on carbon stocks. I found higher concentration of biomass in REDD+ and leakage belts compared to controls and evidence of positive and negative leakage effects within the vicinity of interventions.

The study shows that despite the multiple hurdles to delivering REDD+ carbon and non-carbon benefits, there is promising evidence that these interventions can help to reduce deforestation. As the current Cop27 in Egypt takes place and actions to scale up commitments made in the Glasgow Leaders' Declaration on Forests and Land Use are discussed, this thesis emphasise the role of forest-based interventions as natural climate solutions and highlights some of the challenges that need to be tackled to ensure its successful implementation.

ACKNOWLEDGEMENTS

Thank you to my supervisor, Dr. David Coomes, for giving me the opportunity to pursue my research interests and your valuable advise over these years. Thank you for your support along this journey into uncharted territories. This work wouldn't have been possible without my amazing collaborators. Thank you, Prof. Andrew Balmford and Dr. Rachel Carmenta, for your guidance and sharing your expertise to make this research more impactful and interesting. To Prof. Julia Jones, I feel privileged to have counted you as a collaborator during the PhD journey. Thank you so much for all the energy, guidance and mentorship.

I was lucky to be surrounded by amazing colleagues during PhD. To my friends from the DAB's Forest Ecology corner: in no particular order, thank you Aland, Amelia, Anne, Boris, Edgar, Erika, Florian, James, Jon, O'Neil, Matheus, Ruben, Sacha, Tom, Yi and Yuhie. Many thanks for the kindness, insight, laughter, good food, and interesting conversation and debates.

Thank you to Consejo Nacional de Ciencia y Tecnología Mexico (CONACYT) and Wolfson College to provide me with the financial support to pursue my PhD. Many thanks as well to the Cambridge Centre for Carbon Credits (4C) for assisting me financially to take this research to completion.

To the friends whom I now consider family: Noor, Ara, Lucia, Steff, Sol, Cota, Mido. Of all the things I am grateful about my life in Cambridge, you are the reason why I feel the deepest gratitude.

And to my family back home; Quique, Carlos: Thank you for being a bedrock of happiness and support, even from afar. Agustin, I wish you were here to see this. Maluye, no hay forma de agradecerte con palabras, gracias por ser ese ejemplo de vida.

Publications and collaborations

The analytical chapters presented in this thesis have been formatted as scientific papers to facilitate publication in academic journals. Co-authorship is attributed, reflecting contributions of collaborators to my research, in the form of supervisory support, logistic support, or contributions to data collection. The pronoun “we” is used in the introduction to the thesis and throughout analytical chapters to reflect these contributions, which are specified for all chapters in the following text. The pronoun “I” is used in the conclusion as it contains my personal reflections and opinions.

Chapter 2

Guizar-Coutiño, Alejandro, Julia PG Jones, Andrew Balmford, Rachel Carmenta, and David A. Coomes. "A global evaluation of the effectiveness of voluntary REDD+ projects at reducing deforestation and degradation in the moist tropics." *Conservation Biology* (2022): e13970.

Chapter 3

Guizar-Coutiño, Alejandro, Julia PG Jones and David A. Coomes. "Sensitivity of avoided deforestation estimates to matching specifications”. *In preparation.*

Chapter 4

Guizar-Coutiño, Alejandro, Julia PG Jones and David A. Coomes. " Prevailing spillover effects in REDD+ projects: Implications for forests and carbon”. *In preparation.*

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

CONSERVING TROPICAL FORESTS

Tropical forests are remarkable ecosystems without which life as we know it on Earth would not be possible. They cover approximately 7% of the Earth's land surface (DeFries et al. 2005) and are home to an enormous amount of global biodiversity (Watson et al. 2018; Barlow et al. 2018). These forests influence key global environmental processes including the carbon and water cycles (Gibson et al. 2011; Avitabile et al. 2016) and contribute directly to national economies and local livelihoods around the world (Oldekop et al. 2020). Despite awareness of their critical importance and efforts made to prevent further decline, tropical deforestation has risen in the last couple of decades, with trends rising as recently as 2020 (Weisse & Goldman 2020), generating multiple environmental problems, including greenhouse gas (GHG) emissions, biodiversity loss and a decline in the capacity of tropical ecosystems to sustain global environmental processes (Lewis et al. 2015; Baccini et al. 2017; Barlow et al. 2018; Symes et al. 2018).

Multiple factors contribute to the accelerated decline of tropical forests. Tropical deforestation has been primarily driven by commercial and industrialised agriculture linked to commodity production, often for international markets (Curtis et al. 2018; De Sy et al. 2019; Pendrill et al. 2022). Agricultural production was the major driver behind tropical forest loss in the Brazilian Amazon and South-East Asia over the last two decades (Seymour & Harris 2019). The expansion of these agricultural frontiers into previously undisturbed forests opens the way to a new host of resource extraction activities, such as mining, logging, and unplanned settlement (Leblois et al. 2017), consolidating the conversion of forestlands into permanent human-modified landscapes. Furthermore, as areas affected by deforestation and degradation become vulnerable to fires and extreme weather events such as windstorms and droughts, long-term impacts to carbon stocks and forest structure can be expected after initial waves of disturbances (Rappaport et al. 2018).

The relationship between tropical forests conservation and climate change mitigation is complex. Forests act as carbon sinks by absorbing and fixing carbon in leaf tissues, where it remains locked over short time scales, produce lignin and cellulose to produce wood (which persists years to centuries) and help push carbon further into soils where it accumulates over longer periods of time. Overall, tropical forests account for approximately one-third of the Earth's terrestrial gross primary productivity, containing about half of the terrestrial carbon stocks (Lewis et al. 2015). Such is the extent of this productivity that about 15% of excess anthropogenic emissions have been absorbed by tropical forests since the onset of the industrial period (Pan et al. 2011), perhaps due to "carbon dioxide (CO_2) fertilisation" processes resulting from the enhanced availability of CO_2 in the atmosphere (Tagesson et al. 2020).

While forests in their normal functioning help to mitigate the effects of anthropogenic emissions, tropical deforestation is now a major contributor to climate change, which results in the ecosystem carbon accumulated over millennia being emitted as GHG over a brief period of time (Barlow et al. 2018; Turubanova et al. 2018). Until recently, fluctuations of stored carbon in tropical forests have been in balance due to counteracting regrowth and loss processes (leading to a 'sinking' and 'sourcing' of carbon) (Mitchard 2018). However, acceleration in tropical deforestation rates have rendered these ecosystems a net source of GHG (Baccini et al. 2017; Hubau et al. 2020; Gatti et al. 2021), with deforestation accounting for the second largest source of emissions after the energy sector, estimated at ~11% of global emissions (Shukla et al. 2019). Unless strong measures to conserve forests are taken, this trajectory is expected to continue due to increasing anthropogenic pressures, exacerbated by climate change impacts (Withey et al. 2018; Brando et al. 2019; Hansen et al. 2020).

Forest conservation

Effectively conserving and managing tropical forest is critical for meeting international climate, biodiversity and development goals, and is now recognised as a core component of several multilateral agendas (Pascual et al. 2017; Hansen et al. 2020), including the Sustainable Development Goals (SDGs), Aichi Biodiversity Targets and the Paris Climate Agreement. Particular attention has been paid to conserving tropical forest in recent years given its potential to mitigate emissions in a cost-effective way, helping to meet net zero targets in the decade ahead while delivering social and environmental benefits (Griscom et al. 2017; Seddon et al. 2020)

Calls to preserve tropical forests have prompted responses from a variety of actors. Protected areas (PAs) are one of the earliest and most common approaches for conserving wilderness areas, many of which have been established in tropical forests (Geldmann et al. 2013; Watson et al. 2014). The impact of these interventions on conserving tropical forests has received much attention: research shows PAs have been relatively successful at mitigating deforestation within their boundaries (Fuller et al. 2019; Geldmann et al. 2019), although impacts vary considerably among projects, and appear to be determined by multiple factors, including the country governance and law enforcement (Schleicher et al. 2019b), the type of protected area designation (e.g. setting aside strict conserved zones v.s. allowing for sustainable practices) (Nolte et al. 2010; Schleicher et al. 2017), as well as the management and resourcing of these interventions (Eklund et al. 2019; Schleicher et al. 2019b; West et al. 2022).

Addressing the needs of communities that are dependent on forests to sustain their livelihoods has galvanised alternative approaches to PAs. Pushes to recognise the contribution that indigenous communities make to forest conservation have had a long history (Dawson et al. 2021), and recently came into focus with research showing their effectiveness on stopping forest disturbances within their demarcations (Sze et al. 2022; Prioli et al. 2023). These calls stem from a broader agenda on raising issues about the social impacts of conservation (Adams et al. 2004), and are centred on recognising indigenous communities claims to their territories, which have been traditionally shun away from formal decision-making processes (Dawson et al. 2021). This rights-based approach to conservation is not only essential to assert the rights of indigenous communities, but also necessary to counteract the risks they face to the livelihoods and well-being through commodity frontier expansions (Bille Larsen et al. 2021).

The challenge of achieving environmental gains without compromising social objectives has also motivated incentive-based approaches to conservation. These stem from the assumption that incentives may be more effective for achieving conservation outcomes than restrictions alone (e.g. , protected areas). Perhaps the most prominent example of such approaches is payment for ecosystem services (PES) (Wunder et al. 2020a). PES generated interest across multiple sectors given its potential to scale up conservation efforts while addressing needs for poverty alleviation and improving livelihoods (so called ‘win-win’ situations). PES became a valuable tool in the conservation arsenal in the 2000s concerning tropical forest conservation, with programmes implemented mostly in Costa Rica, Mexico, and Colombia (Ferraro 2017), where forest conservation activities were primarily motivated by aims to ensure the flow of ecosystem services, such as sheltering from extreme weather, ensuring water availability and conserving carbon stocks.

PES has been subject to criticism due to mixed outcomes. Although limited robust empirical evidence on the social and environmental impacts are available (Ferraro 2017), PES programmes have shown modest and varied effectiveness at reducing deforestation, as highlighted by case studies in Bolivia (Wiik et al. 2019), Costa Rica (Arriagada et al. 2012), Colombia (Pagiola et al. 2016), Mexico (Alix-Garcia et al. 2012) and Uganda (Jayachandran et al. 2017). A body of work indicates key challenges that need to be addressed for PES to deliver social and environmental benefits, including better spatial targeting to maximise the intended benefits (Bottazzi et al. 2018), and the generation of sufficient demand to secure revenues for the provision of ecosystem services (Jack & Jayachandran 2018; Wunder et al. 2020a).

Reducing Emissions from Deforestation and Forest Degradation (REDD+)

Reducing Emissions from Deforestation and Forest Degradation (REDD+) refers to a diversity of mechanisms to help conserve and restore tropical forests, involving local, subnational, or national initiatives to support on-the-ground conservation. Much in the spirit of PES programmes, REDD+ started as a simple idea when originally proposed at the UNFCCC negotiations in 2007 at COP13: mitigate GHG emissions by supporting activities to reduce forest loss and enhance carbon stocks in tropical countries, financed primarily by industrialised countries, in exchange of carbon offsets (Agrawal et al. 2011). This concept has been adapting, as the challenges of implementing the original vision came to the fore, with REDD+ now encompassing an array of activities, from direct area-based interventions (with conditional and non-conditional financial incentives) to policy-level interventions aimed at strengthening institutional coordination and capacity to implement forest conservation policies (Angelsen et al. 2018; Wunder et al. 2020b).

REDD+ was originally conceptualised as a climate change mitigation strategy operating within a global carbon-market focused on forest conservation. Questions about focusing on a single ecosystem service in complex socio-environmental contexts (Corbera et al. 2010) led to an expansion of REDD+ objectives, to include the delivery of livelihood (e.g. securing indigenous rights) and biodiversity (e.g. conserving biodiversity and ecosystem services) benefits (Agrawal et al. 2011). A lack of international support eventually led to stagnation of the global carbon-market, yet the concept of REDD+ continued to develop and became adopted by parties of the UNFCCC. Around 50 countries are in the process of developing national REDD+ strategies with the aim of developing conditions to tackle deforestation domestically (Seymour et al. 2018) and four of these (Brazil, Colombia, Ecuador, and Malaysia) have met all the criteria established by the UNFCCC to access results-based payments (Duchelle et al. 2019). Meanwhile, a number of local-scale REDD+ projects were kicked-off in the early days of REDD+ as part of 'demonstration activities'. These projects are the earliest, longest-lasting on-the-ground efforts to conserve forest under REDD+ and will be discussed in more detail in ensuing sections.

Approaches to REDD+ at the subnational jurisdictional level have been gaining traction at international negotiations. A milestone in international REDD+ negotiations was the agreement of the Warsaw Framework in 2013, which puts a strong emphasis on developing REDD+ initiatives at the national and jurisdictional scales (Seymour 2020). This approach stems from the recognition that the complex and diverse range deforestation drivers, which manifest differently across the tropics (Curtis et al. 2018; Seymour & Harris 2019), need to be tackled at multiple scales. Formally endorsed in the Paris Agreement, a jurisdictional approach to REDD+ seeks to align actions within a politically defined areas, through institutional coordination and enforcement of national REDD+ policies. Jurisdictional programmes are expected to complement local-scale interventions in addressing issues such as deforestation leakage (i.e., displacements of deforestation to areas outside intervention zones) and reversals (e.g loss of protected forest as the result of natural disasters such as fires), as well as help streamlining monitoring and verification efforts (Seymour 2020). Many countries have pledged to developing jurisdictional approaches for reducing deforestation and some of these commitments are being implemented (Stickler et al. 2018).

Additionality

For REDD+ to be considered additional, it should demonstrate reductions in forest-based emissions that would not have happened in the absence of interventions. The additionality is based on an intervention's reference level, or business-as-usual (BAU) scenario, i.e., the expected biomass loss and resulting emissions that would have occurred without REDD+, which are calculated on historical forest loss rates and associated emission factors by vegetation type (Goetz et al. 2015).

Methods to estimate additionality depend on the scale and type of project. Local-scale projects have typically derived BAU scenarios from historical land-use trajectories, which are then projected to provide an estimate of the expected land-use conversion, or by spatially modelling deforestation risk factors and projecting their likely trajectory into the future (Wilebore 2015). For instance, projects registered in the voluntary carbon standards (VCS) typically use a 10-years reference period to model forest disturbance trends and produce an estimate of avoided emissions. Additionality is then measured by comparing reference biomass values of the replacing land-use cover after disturbance (e.g., agricultural lands) with project-level biomass estimates collected thorough field inventory (Shoch et al. 2011). Other approaches used at the local-scale include a "control-region" method, in which a reference forest area (i.e., a control) is monitored post-implementation and the subsequent changes in carbon stocks are taken as the BAU reference level (Wilebore 2015).

Additionality at the jurisdictional scale is established by country-level reference levels. The UNFCCC provides flexible guidelines to produce forest reference emission levels (FRELs), derived mostly from historical forest loss trajectories, allowing for the inclusion of domestic deforestation risk factors. Reference levels at the subnational/jurisdictional scale follow criteria established in the national FRELs. Countries have submitted jurisdictional commitments to the UNFCCC and some of them have reported on their performance (Stickler et al. 2018). There is a move to ‘nest’ local-scale baselines with jurisdictional reference levels to ensure harmonised accounting (TSVCM 2021), and there is a recognition too that FRELs methodologies will need to be standardised to ensure fairness, effectiveness and efficiency (Stickler et al. 2018).

CONSERVATION IMPACT EVALUATION

The conservation sector is undergoing a paradigm shift with respect to how the effectiveness of projects and programmes are evaluated. Impact evaluation techniques developed fields such as health, education, and development economics (Baylis et al. 2016) are gaining traction among conservation practitioners (Ferraro & Hanauer 2014), following calls for better impact evaluations standards in conservation (Ferraro & Pattanayak 2006; Baylis et al. 2016). The earnest pursuit of improved evaluation methods is warranted: with interventions growing in complexity and scope (Baylis et al. 2016), usually in the context of limited resources (Reed et al. 2020). Understanding whether conservation programs are succeeding is essential to address pressing socio-environmental challenges (Ferraro & Pattanayak 2006; Baylis et al. 2016).

A major breakthrough was the incorporation of causal thinking in the evaluation of conservation programmes. Conventional approaches to programme evaluation have typically limited to reporting indicators and outcomes over time (Miteva et al. 2012; Samii et al. 2014), mostly lacking any form of controls, so it is not possible to attribute changes due to interventions empirically. Current approaches put a special emphasis on identifying causal relationships between the intervention and the outcome of interest, including the effect *caused* by the intervention on the outcomes, as well as the confounding factors that affect both: the location of where interventions take place and the outcomes of interest.

Examples of confounding factors abound in real-world conservation settings. Interventions such as protected areas have been typically established through a combination of opportunistic factors and land use preferences, which makes them prone to be established in remote areas with lower agricultural or commercial value (Joppa & Pfaff 2009, 2010). Such remoteness is also a key determinant of exposure to deforestation risks (Busch & Ferretti-Gallon 2017), which has a direct impact on the expected rates of deforestation, regardless of whether the intervention takes place. In this way, remoteness becomes a confounder in the evaluation of protected area effectiveness: affecting both the location of where a protected area is established, and the likelihood of lands becoming eventually deforested. Therefore, an accurate assessment of protected area effectiveness should factor remoteness when seeking for control units. Failing to do so may result in the selection of comparison units that are fundamentally different,providing an unreliable reference for estimating impacts.

Experimental approaches help deal with confounding variables, but their application in conservation is limited. Experimental designs such as Randomised Control Trials (RCTs) enable researchers to eliminate confounding bias *ex-ante* due to the random nature of treatment assignment, ensuring no systematic differences in characteristics between treated and control units. Although deemed the 'gold standard' for causal analysis in fields like medicine and public policy, RTCs and experimental approaches are unlikely to play a prominent role in conservation due practical and ethical considerations (Ferraro & Hanauer 2014; Pynegar et al. 2019), however prominent examples exist that have used randomised designs to evaluate the impact of PES programmes (Jayachandran et al. 2017; Jack & Jayachandran 2018; Wiik et al. 2019). Conservation impact evaluation is more likely to rely on observational designs, suited for *ex-post* analysis where treatment was not allocated at random and confounding bias is expected (Baylis et al. 2016).

Central to the notion of causality is the concept of *counterfactuals*, i.e., knowing what outcomes could have happened in the absence of an intervention. Counterfactuals can credibly be constructed by eliminating potential rival explanations, to the extent that the observed effect could only be attributed to the intervention alone (Ferraro 2009). Robust impact analyses therefore rely on research designs that either eliminate, or account for, confounding factors to elucidate the causal effect of the treatment. If unaccounted, confounding factors can systematically bias our analyses, as it is not possible otherwise to isolate treatment effects from the effect that confounders project on both the treatment and the outcome.

Counterfactuals can be constructed by experimental or observational approaches. Experimental approaches like RCTs lead naturally to a credible counterfactual, as units allocated randomly to the control group represent a probable alternative outcome had the intervention not taken place. In observational studies, where units are not randomly allocated to treatment, finding a credible counterfactual requires the identification of untreated units that are as similar as possible to the treated units (except for the treatment condition). Achieving this necessitates strong theoretical understanding of how the intervention is likely to alter the outcome, as well as solid grasp of the potential related variables at play in the causal chain, including observable (i.e., for which data is available) and unobservable (i.e., where data is missing or cannot be measured) confounders. In situations where data on sufficient confounders are available, observational methods such as those based on conditioning can be used to reduce or eliminate confounding bias, rendering the treatment effects estimable.

MATCHING

Matching refers to a suite of statistical techniques for selecting a comparable set of observations for a group of elements within a larger population and is a commonly used approach in observational studies to reduce or eliminate confounding. This is achieved by pairing treated observations with untreated units that are similar in observable confounder characteristics (i.e., covariates), such that no systematic bias exists between treatment groups. Matching thus helps isolate treatment effects by conforming sets of treatment and control units with similar observable characteristics and probability of outcomes.

Multiple matching methods exist to construct sets of observations with similar observable characteristics. The most rudimentary technique will search for an untreated element with an equal covariate value than the reference treated unit. Since 'exact matching' with continuous covariates is rarely feasible, most matching applications involve a scalar classification to perform approximate matches within scale values (Iacus et al. 2019). Two of the most used metrics in conservation (Schleicher et al. 2019a) include propensity score matching, where units are classified based on their probability of receiving treatment given their covariate distributions (King & Nielsen 2019), and Mahalanobis distance matching, where units are classified based on their distance to each other on a multivariate space. Matching methods are an active area of research and practical guidance for the implementation of these methods in applied settings is still emerging (Schleicher et al. 2019a; Rasolofoson 2022). Recent developments of particular interest to conservation practitioners include synthetic control matching (Abadie et al. 2010), which optimises covariate balance and outcome trajectories for units observed over for determined periods of time.

After matching, one must give extensive proof that the selected control group represents a credible counterfactual. This must be backed up theoretically and empirically, through diagnostics of the matched sets. Matching will be valid under the assumption that the selected covariates account for both observed and unobservable confounding factors (in the latter case, indirectly through their correlation with observable variables). The matched set must also show 'common support' between treated and control groups, with substantial overlap in covariate densities across treatment groups. If matching is combined with difference-in-differences to analyse treatment effects in a before-after design, the counterfactual should provide evidence of exhibiting "parallel trends", i.e. similar outcomes trajectories in the pretreatment period (Angrist & Pischke 2008). Methods such as synthetic control matching are optimised for producing matched sets that exhibit similar outcome trajectories. Since the possibility of hidden confounding can never be ruled out in observational studies, sensitivity analyses to tests for unobserved confounders should be included in post-matching routines (Jones et al. 2022). Hidden confounding are all too common in conservation , where unobserved confounders can originate from factors that are known to be correlated with selection to treatment but remain unobserved nonetheless because are hard to quantify (e.g., institutional factors that correlate with programme rollout) (Usmani et al. 2022).

The use of matching method for assessing conservation impact evaluation is on the rise (Börner et al. 2020). These studies have enhanced our understanding of the strengths and limitations of forest conservation strategies such as protected areas (Andam et al. 2008; Joppa & Pfaff 2011; Carranza et al. 2013; Schleicher et al. 2017), and community forest management (Rasolofoson et al. 2015; Eklund et al. 2016; Oldekop et al. 2019). Approaches have been also used to determine the effect of cash-transfer programs (Ferraro & Simorangkir 2020).

EARTH OBSERVATION DATA FOR EVALUATING FOREST CONSERVATION

The arrival of observational methods to forest impact evaluation is in part fueled by the availability of data sources to track forest cover globally. Moreover, the availability of Earth observation data is opening new opportunities for evaluating the impact of on-the-ground interventions aimed at reducing deforestation and forest degradation (De Sy et al. 2019), and is particularly instrumental for monitoring initiatives that apply conditional rewards for forest conservation (Goetz et al. 2015).

Remote sensing (RS) technologies in the form optical, thermal and radar sensors detect information encoded in the electromagnetic spectrum and uses these data to infer characteristics of abiotic and biotic surfaces, such as temperature, chemical processes, moisture content and texture (Pettorelli et al. 2014). RS is a key tool in the ecology and conservation biology fields, and has become a leading method for measuring and assessing ecosystem functions due to the provision of repeatable, comparable, and cost-effective (at least to the user) quantification of environmental features at multiple spatial, and temporal, scales (Pettorelli et al. 2014; Lausch et al. 2017)

The detection of forest cover change has been facilitated by satellite-based RS. The NASA's series of Landsat Earth satellites and Moderate Resolution Imaging Spectroradiometer (MODIS) sensors, commencing in 1972 and 1999 respectively, have been utilised in countless studies to generate regional and global estimates of forest cover change. Advances in processing power over the last decade has prompted the development of global-scale forest-tracking products at Landsat pixel-resolution (~30m) (Sexton et al. 2013; Hansen et al. 2013; Vancutsem et al. 2020). Optical satellite imagery is also routinely used by countries in combination with field-based data to generate national forest inventories for tracking changes in forest cover, and has been key for the development of alert systems to track deforestation events at sub-monthly intervals in Brazil, Peru, the Republic of Congo and Indonesia (Hansen et al. 2016; Wheeler et al. 2018). Progress made in remote sensing technology to date represents a major step towards advancing knowledge on global deforestation patterns, and has galvanised responses from the private and public sectors to address deforestation (Seymour & Busch 2016). However, uncertainties remain about the estimates of forest change, which are largely defined by the choice of methods and definitions for land cover classification (Lui & Coomes 2015; Melo et al. 2018).

The contribution of tropical forest loss to carbon emissions remains uncertain (Baccini et al. 2012; Harris et al. 2012; Tyukavina et al. 2015). Reducing these uncertainties is important to enhance understanding on the role of forests in the global climate system, and to support forest conservation initiatives important in the context of carbon markets and REDD+ (Mitchard 2018). Field-based forest inventories are essential for developing a baseline estimation of carbon stocks at national and sub-national scales, however, the collection of inventory field-data is often a resource intensive activity, and the rigour of sampling can be constrained by access, budgets and technical capacity. Combining synthetic aperture radar (SAR) measurements with field-based inventories enables to generate aboveground biomass estimates for extended areas (Mitchard et al. 2013). SAR sensors operate by emitting wavelength pulses that allows the characterisation of surface properties, and have recently been used to quantify vegetation metrics such as forest cover, canopy volume and forest height at regional scales (Mitchard et al. 2013), as well as at the global scale (Shimada et al. 2014).

Light Detection and Ranging (LiDAR) is an active RS sensor that can be mounted in satellites, aircrafts or unmanned aerial vehicles, and operates by emitting laser pulses from which metrics of forest structure can be derived. Landscape level assessments involving airborne LiDAR provide fine-detailed estimates of aboveground biomass at a sub-meter resolution (Asner et al. 2013; Asner & Mascaro 2014). Repeat LiDAR measurements further enhances carbon flux estimates although deploying this technology is costly, preventing the application of this technique for systematic assessments. Integrating LiDAR with field-based measures has seen an increasing interest for monitoring global forest carbon (Duncanson et al. 2021).

Monitoring forest degradation and resulting carbon emissions is essential for evaluating the performance of forest conservation strategies but remains technically challenging. Forest degradation refers to partial loss of canopy cover due to anthropogenic or natural disturbances, resulting in the diminished capacity of forests for sustaining biodiversity and supporting ecosystems function (Simula 2009). Forest degradation is commonplace in tropical forests, with estimates suggesting that it affects an area similar to that affected by deforestation (Pearson et al. 2014; Vancutsem et al. 2020). Tracking forest degradation can be an effective tool for conserving forests as, among other things, it enables early detection of can be effective in preventing further deforestation (Joshi et al. 2015), but this is still an ongoing area of research.

Uncertainties in the estimation of biomass using RS approaches remains one of the major challenges that to be addressed (Mitchard 2018). The Global Ecosystem Dynamics Investigation (GEDI) LiDAR instrument has been added to the constellation of active sensors in space and provides consistent measurements of forest structure across the global tropics. Likewise, the European Space Agency (ESA) Sentinel (Berger et al. 2012) and BIOMASS missions (Le Toan et al. 2011) will support innovative ways to generate high resolution global biomass estimates. Developing improved methods for quantifying forest degradation based on open-access imagery will allow to generate better estimates of intervention performance for conserving carbon, increasing our ability to track forest degradation will help to guide monitoring and restoration efforts.

LOCAL INTERVENTIONS AS CASE STUDY FOR EVALUATING THE IMPACT OF REDD+ IN REDUCING FOREST LOSS

Local interventions conform the earliest, long-lasting on-the-ground efforts to conserve forest under REDD+. Around 350 landscape-based REDD+ interventions have been implemented since the first wave of projects initiated more than a decade ago (Simonet et al. 2020). Although branded as REDD+, and having overlapping goals for achieving social and economic benefits with biodiversity and environmental gains, these interventions in practice are hugely varied in scope, social objectives, and approaches for achieving conservation (Sills et al. 2014; Carmenta et al. 2020), involving a variety of stakeholders and forest users, from indigenous communities seeking to halt deforestation from encroaching their territories, to private actors setting aside conservation lands that would otherwise have been allocated to commercial activities (Simonet et al. 2020).

The development of REDD+ into a heterogeneous collection of interventions has in part been determined by the evolution of its funding landscape. REDD+ was originally envisioned as a mechanism whereby industrialised economies would fund forest conservation activities in tropical forest countries in exchange for carbon offsets. While negotiations about the global carbon market architecture never consolidated, discussions around the implementation of national and subnational REDD+ frameworks continued in the multi-lateral fora, opening streams of funds for local REDD+ initiatives from multilateral and international development agencies . This contributed to the 'aidification' of REDD+ in the last decade (Angelsen 2017), with a variety of players involved in the implementation of REDD+, including public, private, multilateral, and nongovernmental organizations (NGOs). These projects entail a mixture of aid-based, for-profit, or mixed approaches to project financing, which may or may not involve results-based payments in the latter case (Wunder et al. 2018). For some of these interventions, carbon offsetting in the voluntary market is only one of the different revenue streams that projects pull to support on the ground activities, but the reliance on carbon market is expected to grow.

Rigorous evaluations of local-scale REDD+ are only starting to emerge with mixed results across different fronts (Duchelle et al. 2018b). An important body of work has focused on the implications of REDD+ on the local communities, showing that REDD+ thus far has had negligible impact on issues such as securing land tenure (Sunderlin et al. 2018a) and on improving the livelihoods of local communities (Duchelle et al. 2018a). Moreover, while progress was made in some countries on integrating different stakeholders in REDD+ land-use decision-making processes, there is also evidence of insufficient local involvement through free, prior, and informed consent processes (Sunderlin et al. 2018b), with some pointing to biased gendered participation in early REDD+ projects (Kariuki & Birner 2016). Furthermore, empirical impact evaluations about the ability of local-scale interventions to conserve forests is still scarce, with recent analysis showing a range of outcomes, from projects achieving reductions in deforestation (Jayachandran et al. 2017; Simonet et al. 2019; Guizar-Coutiño et al. 2022), to null or negative results (West et al. 2020; Correa et al. 2020).

Interest in the voluntary market has been rekindled, but important challenges remain for REDD+ to benefit from this uptake. Demand for carbon offsets has increased in recent years as more organisations seek to meet their decarbonisation plans using offsetting, which for forest-based carbon credits meant an increase in the volume of traded offsets in 2020 and 2021, of 30% and 40% respectively (Donofrio et al. 2021). Some estimates suggest that the voluntary market will have to scale-up 15-fold by 2030 to meet expected demands, with nature-based credits playing a significant role in the portfolio of offsets (TSVCM 2021). While this may indicate favourable market signals, important challenges remain to mobilise enough resources for nature-based carbon credits to meet this demand, including increased access to financing (TSVCM 2021), and a fair carbon pricing that better accounts for the opportunity costs of forgone land-use (UN-REDD 2022). Furthermore, substantial harmonisation between REDD+ local and subnational approaches, with clear pathways for integrating results-based approaches to REDD+ (Duchelle et al. 2019) will be required to set the conditions to scale-up REDD+ within a market-based framework. Another important challenge determining the future of REDD+ concerns the integrity of REDD+ carbon credits. Recent independent evaluations using quasi-experimental methods to compare deforestation rates in VCS REDD+ sites relative to counterfactuals have found consistently less environmental gains than forecasted by VCS (West et al. 2020; Guizar-Coutiño et al. 2022), potentially leading to an over-crediting of projects. Others have found similar inconsistencies for forest-based carbon offsets from the California Air Resources Board standards (Badgley et al. 2022).

Initial examination of REDD+ projects

This PhD is centred on understanding whether REDD+ projects have delivered on reducing deforestation, forest degradation and associated reductions in carbon. Throughout my PhD I focused on projects registered in the voluntary carbon standards (VCS) as they were one of the registries that provided easier access to project data in 2019. The examination commenced by mapping all available data for AFOLU type of projects established in the tropics. The database at the time of examination obtained data for Afforestation, Reforestation and Revegetation (ARR, n=111), Reducing Emission from Deforestation and forest Degradation (REDD, n=81), Improved Forest Management (IFM, n=7), Wetland Rewetting and Conservation (WRC, n=1), established in 41 countries (Fig. 1.1).

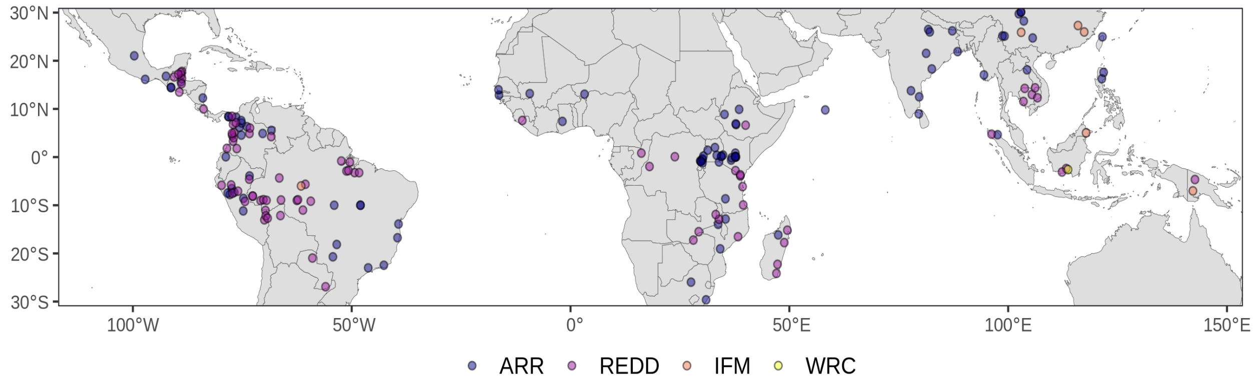

Figure 1.1: Location of the 200 projects in the VCS database (circa 2019). Afforestation, Reforestation and Revegetation (ARR, blue), Reducing Emission from Deforestation and forest Degradation (REDD, purple), Improved Forest Management (IFM, orange), Wetland Re-wetting and Conservation (WRC, yellow). Location points obtained from the VCS website.

I proceeded to obtain all the data available for REDD+ projects in the VCS database. In specific cases I contacted project proponents to request project boundary files, particularly when the boundaries in the database were not available, or did not correspond to the maps depicted in project documents. I amassed a database of interventions in a PostGIS database. Prior to compiling files in the database, I cleaned geometries and removed overlapping features, following best practices for handling protected area data (Bingham et al. 2019). This was a necessary step as many of the files available were produced with the aim of showing where the interventions took place but were not adequately formatted for geospatial analysis. The data collection process resulted in 71 REDD+ interventions mapped, which included projects established in Central and South America (n=45), Africa (n=18) and Asia-Pacific (n=8), with a median project area estimated at 73620 Ha (inter-quartile range = 39770 to 193580 Ha). Most of these projects (n=60) were certified under the dual VCS Carbon Climate and Biodiversity standard, which stands for projects that seek to produce social and biodiversity co-benefits, in addition to climate gains. These projects were established between 1994 and 2017, used a total of 10 different methodologies to define emissions baselines.

Observational approaches for data analysis were required as REDD+ projects tended to be located where conservation and development efforts were already in place. Many REDD+ sites were established adjacent to PAs, in cases overlapping partially or wholly with their boundaries, in an attempt to generate a buffering barrier between deforestation activities and PAs.

Moreover, many projects currently enlisted as REDD+ were originally developed as PES or integrated development and conservation projects (Lin et al. 2012), effectively making REDD+ area selection more likely where NGO engagement was present prior to the REDD+ designation (Usmani et al. 2018).

Structure of the thesis

In the remaining chapters of this thesis I will develop three pieces of research that range from the applied evaluation of REDD+ to interrogating methodological approaches used to evaluate the impact of conservation interventions. The research provides evidence of the impact of local-based REDD+ interventions in reducing deforestation while examining some of the challenges involved in using quasi-experimental methods to evaluate forest-based interventions. **Chapter 2** presents one of the first global assessments of the impact of REDD+ in reducing deforestation using a standard quasi-experimental approach. I found reductions in forest loss due to REDD+, particularly in areas of higher threats of deforestation. **Chapter 3** provides a comparison of matching approaches using state-of-the-art robust evaluation metrics, I found consistent average reductions in deforestation due to REDD+ across multiple matched models, however alterations in matching parameters led to considerable variation of baseline levels in some projects, thus impacting estimates of treatment effect. **Chapter 4** extends the analyses of REDD+ by measuring local leakage effects and the potential implication on carbon stocks. I found higher concentration of biomass in REDD+ and leakage belts compared to controls and evidence of positive and negative leakage effects within the vicinity of interventions.

Chapter 2: A global evaluation of the effectiveness of voluntary REDD+ projects at reducing deforestation and degradation in the moist tropics

Abstract

Maintaining tropical forests is vital for biodiversity conservation, and keeping carbon locked-up in forests is key to tackling climate change. Reducing Emissions from Deforestation and forest Degradation (REDD+) projects aim to contribute to climate change mitigation by protecting and enhancing carbon stocks in tropical forests, but there are no systematic global evaluations of their impact. Using a new data set for tropical humid forests, we used a standardised evaluation approach to quantify the performance of a representative sample of 40 voluntary REDD+ certified under the Verified Carbon Standard, located in nine countries. In the first five years of implementation, deforestation within project areas was reduced by 47% (95% CI = 24-68%) compared with matched counterfactual pixels, while degradation rates were 58% lower (95% CI = 49-63%). Reductions were small in absolute terms but greater in sites located in high deforestation settings and did not appear to be substantially undermined by leakage of activities into a 10-km zone surrounding projects. At COP26 the international community renewed its commitment to tackling tropical deforestation as a nature-based solution to climate change. Our results indicate that incentivising forest conservation through voluntary site-based projects can slow tropical deforestation; they also highlight the particular importance of targeting financing to areas with higher deforestation threat.

Introduction

Rapid decarbonisation of economies is essential to avert the worst impacts of human-induced climate change, but protecting natural ecosystems and the carbon they store could contribute significantly to meeting net zero targets, particularly in the decade ahead (Griscom et al. 2017; Seddon et al. 2020; TSVCM 2021). If delivered at scale, keeping carbon stored in forests by avoiding deforestation could be one of the most effective nature-based climate solutions (Stern 2007; Houghton et al. 2015; Lewis et al. 2019). Given that tropical forest ecosystems support the majority of terrestrial biodiversity (Lewis et al. 2015; Barlow et al. 2018), slowing the loss of these vital habitats would also have substantial co-benefits for biodiversity. The 26th Conference of the Parties of the United Nations Framework Convention on Climate Change (COP26) saw leaders of over 100 counties, containing over 85% of the world's forests, make a commitment to bring deforestation and degradation to an end by 2030, backed by almost USD 20 billions of investments from public and private funds (UNFCCC 2021). There is justified scepticism, however, as previous international commitments have failed to materialise, including most recently the New York Declaration on Forests, which aimed to half deforestation by 2021 (Partners 2020).

Since the early 2000s, REDD+ programmes have sought to reduce deforestation by creating financial and institutional mechanisms to deliver genuine emission reductions while benefitting local livelihoods and biodiversity (Holloway & Giandomenico 2009; Agrawal et al. 2011). Around 50 countries have national and subnational REDD+ programs at various stages of development and more than 350 REDD+ projects have been initiated to date (Simonet et al. 2020). These are likely to vary in effectiveness, for a variety of reasons: they are exposed to different drivers of deforestation (Simonet et al. 2020), have differing social objectives (Sills et al. 2014; Carmenta et al. 2020), engage in different activities to reduce deforestation, and operate under varying degrees of conditionality (Wunder et al. 2018). In practice, REDD+ project implementation has faced many difficulties (Duchelle et al. 2018a; Milne et al. 2019). Nevertheless, if the world is going to meet its renewed deforestation commitments it is important to learn from REDD+ initiatives to date, and in particular from in-depth analysis of their effectiveness at reducing deforestation and forest degradation.

Several recent studies have evaluated the impact of REDD+ and similar interventions on tropical deforestation. Randomised control trials in Africa have found that paying households to reduce deforestation were effective (Jayachandran et al. 2017), while unconditional payments were not (Wilebore et al. 2019). Similarly, a study of REDD+ interventions along Brazil's Trans-Amazon Highway reported a 50% reduction of deforestation rate compared with matched control sites (Simonet et al. 2019), whereas two other studies of Amazonian REDD+ projects concluded that voluntary REDD+ projects had little impact (West et al. 2020; Correa et al. 2020). Deforestation rates in Guyana remained below those recorded in a counterfactual region while a Norway-supported jurisdictional REDD+ program was active (Roopsind et al. 2019). Given this heterogeneity in results it seems timely to quantify impacts across a large sample of REDD+ projects. Projects need to calculate emissions reductions and have them verified in order to sell carbon credits, but unfortunately the monitoring reports produced by the projects assess outcomes in inconsistent ways (Wilebore 2015), and cannot be used to make global comparisons of effectiveness. To address this gap, here we apply a standardised approach to examine how far these site-based REDD+ projects have slowed deforestation across a global sample of projects.

Specifically, we quantified the impact of site-based REDD+ projects on deforestation by examining projects certified by Verified Carbon Standards (VCS) developed by Verra, one of the leading accreditation registries for voluntary REDD+ projects (Donofrio et al. 2019). We used a new dataset that tracks annual deforestation and degradation rates across the moist tropics at ~30-m resolution using multispectral imagery collected by earth observation satellites (Vancutsem et al. 2021) and estimate effect sizes by comparing rates after REDD+ projects were implemented with rates in matched control pixels. Our analysis covers all VCS projects in the humid tropics for which project location maps were provided by Verra, and sufficiently long time-series of satellite imagery were available and appropriate counterfactuals could be identified. We compared REDD+ project areas with matched pixels in the wider landscape, with and without the inclusion of areas included in protected areas (see Appendix A Fig. 1.3). Matching pixels in this way provides a “quasi-experimental” analysis of the effectiveness of REDD+, by eliminating from our analyses, to the best of our ability, other confounding factors that might have influenced deforestation rates. Leakage, which occurs when deforestation activities in project areas are shifted elsewhere upon project implementation - is a widely acknowledged risk of REDD+ and other forest-based interventions (Pfaff & Robalino 2017). We therefore also evaluate evidence of local leakage by estimating changes in forest loss in a 10-km zone immediately outside project boundaries (Ewers & Rodrigues 2008). Additionally, we explore how REDD+ effectiveness varies with deforestation threat, defined as the background deforestation rate in the host country during the implementation period, and whether effectiveness varied through time.

Methods

Selection of an initial set of REDD+ projects

REDD+ projects earn carbon credits for independently verified emission reductions relative to a business-as-usual scenario (e.g. an estimation of emissions in the absence of the project); these reductions may arise by avoiding deforestation, reducing degradation or increasing forest cover through restoration or afforestation activities. We selected REDD+ sites listed in the Verified Carbon Standards (VCS), currently the largest registry for voluntary REDD+ projects (Donofrio et al. 2019), and one of the only registries that had boundary data on REDD+ projects publicly available at the time of the study. Between January and March 2019 we gathered project design documents, validation reports and geospatial datasets depicting project area boundaries from the VCS registry (http://www.vcsprojectdatabase.org). Our analysis focused exclusively on projects categorised as “Reducing Deforestation and Degradation” and established in the tropics (Africa, South-East Asia, Latin America and Oceania) of which 81 were found. We contacted project proponents and the VCS registry to request source boundary files if project boundary maps were not available. For the 71 projects for which we were able to obtain boundary files, we performed geospatial standardisation techniques typically recommended for protected areas polygons (Bingham et al. 2019). Specifically, we normalised overlapping polygons so that each overlap was contained by a unique geometry and re-projected the database to a Mollweide equal-area projection (Bingham et al. 2019) using PostGIS (2.3).

The VCS methodology constrained how we could analyse the effects of the projects. Specifically, the avoided deforestation protocols require that a project’s spatial extent (i.e. its “accounting zone”) comprises a parcel of land that has maintained 100% forest cover for at least 10 years prior to the project starting date. Thus any deforested areas adjacent to, or within, REDD+ boundaries were systematically excluded from the intervention zones provided to VCS for monitoring, reporting and verification purposes (Shoch et al. 2011). We had no choice other than to adopt a similar approach, defining our basic unit of analysis as a pixel that was observed to have remained as undisturbed forest from 1990 until the project starting year. This meant we could not employ a difference-in-difference approach to isolate the effect of a project because deforestation in the project area was, by definition, zero prior to project commencement. Nevertheless, after-only analysis is widely used to evaluate the impacts of conservation interventions on environmental outcomes including deforestation (Rasolofoson et al. 2015; Eklund et al. 2016; Geldmann et al. 2019). Arguably our approach can be regarded as equivalent to a difference-in-difference analysis, in which the pre-implementation treatment and control sites have zero deforestation.

We diverged from VCS protocols in the approach used to estimate project additionality. Under VCS, a project must select a counterfactual area of forest that has similar deforestation threats to the project area. We instead adopted a pixel-based matching approach, which lead to us having pixels scattered over many sites, instead of a single area. A benefit of this approach is that it ensures that the control set of pixels are exposed to the same geographic drivers of deforestation as the pixels in the REDD+ project sites (Schleicher et al. 2019a). It would be valuable to compare the two approaches, and critically evaluate whether VCS counterfactual areas are selected robustly, but this is not feasible because shapefiles of counterfactual areas are not typically provided by projects.

Yearly maps of forest cover, deforestation and forest degradation

Annual maps of forest cover, deforestation and forest degradation were taken from the recently published Tropical Moist Forests (TMF) database (Vancutsem et al. 2021), which was derived from time-series of multispectral imagery collected by Landsat, with pixels of about 30 m resolution. This database provides a long-term characterisation of forest disturbances including degradation and deforestation, on an annual basis from 1990 to 2019. We focused our analyses on quantifying temporal changes between three forest classes: the *undisturbed* class, which represents closed evergreen or semi-evergreen forest areas that have not been disturbed over the entire period examined; the *degraded* class, which is characterised as forest disturbances that last over a period of up to 2.5 years, resulting from natural causes such as wind storms or anthropogenic causes such selective logging; and the *deforested* class, representing long-term forest disturbances (> 2.5 years) and the complete removal of forest cover. The 71 REDD+ sites for which we had boundary maps were established in moist and seasonally dry tropical biomes, but as the TMF database does not monitor drier regions, we limited our analysis to sites that were densely forests, i.e. had at least 80% forest cover at the project start date, bringing the total down to 54 projects. Furthermore, since our estimation approach involves analysing forest cover change over a period of at least five years after project implementation (see *Impacts of REDD+ projects on deforestation and forest degradation rates*) we excluded three projects that had been active for a shorter period. We also excluded one site that started operations before the year 2000 (since this is prior to the operationalisation of REDD+) leaving us with 50 projects with which to search for counterfactual observations for comparison (see *Matching*).

Sampling design within project areas and surrounding landscapes

We used a pixel sampling approach to characterise project areas (i.e. treatment areas) and the regions where these were located, from which we identified our control groups to evaluate treatment effects. Pixels in the treatment areas were sampled by creating a regular pattern of sampling points, each separated by 250 meters, within the boundaries of REDD+ projects, using the project boundary files. To generate observations from which we generated control pixels by matching (see below), a large number of pixels (up to seven times the number located in the project area) located within the same country and biome as treatment pixels were sampled at random. We retained pixels if they remained undisturbed for at least 10 years prior to the project starting date (i.e. mirroring the VCS methodology for project areas). Following this approach, all the control and treatment pixels used in our analyses had zero deforestation and degradation rates in the ten years up to project implementation.

To account for local leakage effects in our design, we defined 10-km buffers around the REDD+ interventions (i.e. leakage belts) from which we did not collect control pixels. Leakage belts were adjusted to 1) exclude any overlaps with protected areas and other nearby REDD+ projects, and 2) exclude any overlap between buffer zones of REDD+ projects that were close together (n=7) (Appendix A Fig. 1.9). We then assessed the extent to which leakage activities took place (e.g. significant differences in deforestation rates) by examining deforestation patterns before and after project implementation within leakage belts (see *Quantifying local leakage*).

Matching

We performed statistical matching to identify sets of control pixels for each project area that were similar in observable confounders associated with forest loss, thus ensuring that selected controls were exposed to the same drivers of deforestation as project area pixels. To implement a standardised method we sought a single set of covariates across the full set of sites, acknowledging that drivers of deforestation vary across the countries included in this study (Curtis et al. 2018). We collected pixel-level data on socio-demographic and biophysical characteristics that are typically associated with deforestation (Angelsen & Kaimowitz 2001; Busch & Ferretti-Gallon 2017): elevation and slope (Jarvis et al. 2008), distance to the nearest urban centre in 2015 (Weiss et al. 2018) and distance to forest edge (Laurance et al. 2011; Ewers et al. 2011). To account for temporal changes in distance to forest edge, we constructed annual time-series of the mean distance to the nearest deforested pixels based on the TMF map. For each sampled pixel, we calculated the distance to the closest pixel that had changed its status from undisturbed to deforested during the observed year (for 2000-2019). We then produced a rolling average estimate of the mean distance to the closest deforested pixel in the previous five years, for the period 2005-2019.

Matching was performed with the R *MatchIt* package (Ho et al. 2011), measuring the similarity between treatment and control pixels using the Mahalanobis distance metric (Legendre & Legendre 1988), which has been shown to result in balanced comparison groups when the number of matching covariates is relatively low (Stuart 2010). We performed 1:1 nearest-neighbour matching with replacement using elevation, slope, mean distance to population centres and mean distance to deforested areas over the five years prior to project commencement (see Appendix A Table A.1). We exact-matched on country and biome (Dinerstein et al. 2017), and for 10 sites which intersected more than one biome we subdivided

REDD+ sites to generate sets from the same biome to match against controls. By matching within the same tropical moist forests in the same biomes and countries we ensured comparability of bioclimatic conditions for agricultural development. We considered an absolute standardised mean difference of <0.25 between treated and control samples across all covariates as acceptable (Stuart 2010). Only those REDD+ projects that met this criterion for at least 90% of pixels (across all subgroups) were included in further analyses. Ten sites were dropped after matching as these were not matched across all covariates (see Fig. 2.1, Appendix A Fig. 1.7). To evaluate whether the resulting subset of 40 sites was representative of the environmental conditions found in the original set of 71 sites, we ran a logistic regression to predict the probability of being included in this analysis as a function of accessibility, distance to deforestation by project staring date, elevation, Human Development Index (HDI) and project area size. If environmental conditions in the filtered dataset were different from those in the original dataset, we would expect significant effects in the model (having applied a Bonferroni correction for multiple comparisons). Furthermore, we ran a post-hoc analysis of the importance of the covariates we used for matching confirmed that they did indeed predict forest loss, with the final parsimonious model describing a moderate proportion of the observed variation in deforestation across the examined landscapes (Nagelkerke's r^2 0.47) (Appendix A, Predicting deforestation with the matching variables).

Selecting an appropriate control to evaluate the impact of conservation interventions can be complicated by the presence of other interventions occurring in the examined landscapes (Schleicher et al. 2019a). To account for the presence of protected areas we ran a separate set of analyses in which we excluded pixels located within areas protected area polygons (see Appendix A Fig. 1.3), based on the World Database on Protected Areas (UNEP-WCMC & IUCN 2019). We standardised the protected area database by removing areas categorised as "not designated" or "inscribed", as well as UNESCO Biosphere Reserves (Bingham et al. 2019) and reprojected the geometries to a Mollweide equal-area projection.

IMPACTS OF REDD+ PROJECTS ON DEFORESTATION AND FOREST DEGRADATION RATES

Annual deforestation rates were calculated for each REDD+ site and its control pixels. To do this, we built a time series of annual rates of deforestation events spanning 2001 to 2019, for all our groups of treatment and matched control pixels. We then estimated annual deforestation rates within a REDD+ project area as: $r_t = \Delta p_t / p_d$, where Δp_t is the total number of pixels deforested in year t and p_d is the number of forested pixels at the start of that project. Annual deforestation rates within control pixels were calculated the same way. We then calculated degradation rates using the same approach in control and treatment sites. These time-series provided the information needed to calculate annual deforestation and forest degradation rates from the project starting date for up to 10 years after implementation (where enough data was available).

Absolute differences in deforestation rates between treatments and controls were calculated as $\overline{r}_t - \overline{r}_c$ for each REDD+ site, where $\overline{r}$ refers to the mean deforestation rate within the first five years of implementation, and the *t* and *c* subscripts refer to treatment and control groups, respectively. We used the 40 site-level estimates to derive the global mean change in deforestation and estimated 95% confidence intervals by non-parametric bootstrapping. The same approach was used to calculate site-level differences and global mean change in forest degradation rates. Note that, although we did not incorporate spatial autocorrelation into our analyses of effect sizes, we did reduce autocorrelation in our datasets by taking a systematic sample of pixels within treated areas (each 30 m pixel separated by 250 m), and by drawing random candidate control pixels from other areas of tropical moist forest within the same country, prior to matching. Subsequent analyses to evaluate changes in effect size over time were performed with subsets of sites with enough observations to estimate the treatment effect after eight and ten years of implementation, where we used all the temporal observations available at each time subset to estimate the treatment effect.

Site-level proportional differences in forest disturbances were calculated by dividing the mean disturbance rate (e.g. deforestation or forest degradation) at treated sites within the first five years of implementation, by the mean disturbance rate at control groups over the same period, i.e. $\frac{\overline{r}_t}{\overline{r}_c}$. The overall mean in deforestation and forest degradation across the 40 sites was calculated to estimate the proportional reduction in disturbance rates associated with REDD+ projects globally; 95% confidence intervals were estimated by bootstrapping.

Effectiveness of REDD+ in relation to background deforestation rates

Given that the sampled REDD+ projects were located in countries and periods with different rates of deforestation, we explored how the REDD+ treatment effect varied with background deforestation. Background deforestation rates was estimated from the mean rates of tropical moist forests loss for the country in which a project was located, during the first five years of project operation (Appendix A Fig. 1.4). These rates were used to classify projects into “high threat” and “low threat” categories depending on whether the rates were above or below the mean annual deforestation rate observed across the humid tropics over the last three decades, (i.e. 0.57% yr^{-1}) (Vancutsem et al. 2021). To determine changes in annual forest loss rates between high and low deforestation groups, we grouped site-level mean differences and derived group-level mean estimates with 95% confidence intervals by bootstrapping. We further conducted a Wilcoxon’s rank-sum test to compare differences in reductions between high and low threat groups. To determine threat level proportional differences between treatment and controls, we grouped site-level proportional changes and high and low deforestation groups and derived mean estimates with 95% confidence intervals by bootstrapping. We repeated the same analyses to derive absolute and proportional changes in forest degradation between high and low deforestation groups.

Quantifying local leakage

We evaluated local leakage by testing whether there was a significant increase in annual deforestation rates in the leakage belts (e.g. buffer zones), following project implementation. Annual deforestation rates within leakage belts were estimated by dividing the area deforested for each year (for 1991-2019), by the extent of undisturbed forests in 1990. We extracted five years of data before and after projects started and tested whether there was a significant increase in rate following implementation using site-level bootstrapped t-tests. We examined whether leakage grew worse or lessened over time by performing bootstrapped t-tests on subsets of sites with enough observations to examine changes over periods of eight and ten years before and after project implementation.

Results

Project selection

The 40 REDD+ projects selected for this study follow systematic filtering of the initial database were located in nine countries and together encompassed 8.38 million ha of humid tropical forest, with a median area of 92,353 ha (interquartile range IQR = 46,192 to 190,660 ha). Thirty-three were in the Americas, five in Africa, one in Asia and one in Oceania (Fig. 2.1); note that several projects in Africa and Asia were excluded because they were situated in dry forest and savanna regions into which the TMF deforestation maps do not extend. Although the analysed sites represented a subset of the 71 projects initially obtained from the VCS database examination, they were similar to the wider sample in most characteristics (elevation, slope, Human Development Index, project area and background deforestation rate) but were significantly closer to populated centres (Appendix A Table A.2). This suggests that our analysis may be indicative of the performance of REDD+ projects in sites that are more exposed to deforestation (e.g. due to the expansion of infrastructure or agricultural activities; Geist & Lambin 2002; Busch & Ferretti-Gallon 2017).

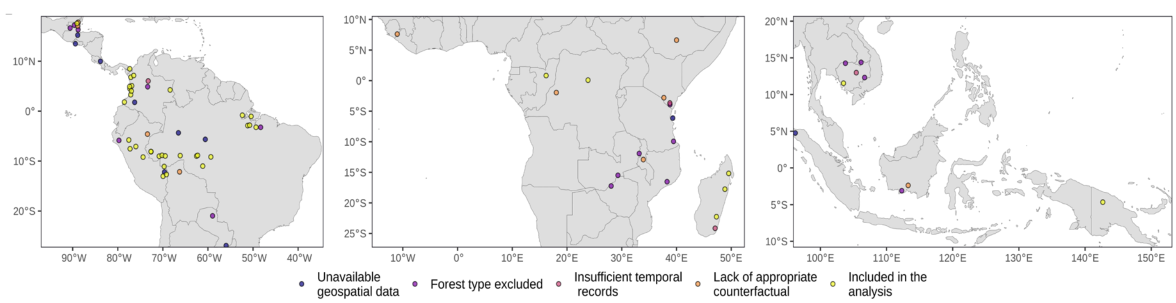

Figure 2.1: Location of the REDD+ projects included in the analysis. Of 81 tropical REDD+ projects certified by VCS by 2018, 10 did not provide detailed maps of project locations (blue dots), 17 had less than 80% evergreen forest cover at the start of projects (purple), 4 had been operating for fewer than five years or had commenced before the year 2000, and 10 could not be matched with appropriate control pixels (orange), leaving 40 projects in the final dataset (yellow).

The average effectiveness of REDD+ projects

REDD+ project implementation was associated with significant reductions in both deforestation and forest degradation over their first five years of operation, compared to matched control pixels in the wider landscape (Fig. 2.2). Reductions in deforestation rates were observed in all but three sites (Fig. 2.2 *a*), which together amount to a mean reduction of 0.22% yr^{-1} (95% CI= 0.13-0.36% yr^{-1}; Fig. 2.2 *b*). To put this in perspective, the average deforestation rate in the moist tropics between 1990 and 2019 was 0.57% per year (Vancutsem et al. 2021). Reductions in degradation rates were observed in all but five sites (Fig. 2.2 *c*), with an average reduction of 0.41% yr^{-1} (CI= 0.24-0.65% yr^{-1}); we lack a pan-tropical degradation rate to compare with this figure. Expressing these absolute reductions in the rate of deforestation or degradation as relative reductions (i.e., as a percentage of rates observed in controls), we found that REDD+ projects reduced deforestation by 47% (CI = 24-68%) and degradation by 58% (CI = 49-63%) in the first five years (Fig. 2.2-time). These annual reductions in deforestation rates amounted to a total of 66,754 ha of avoided forest loss across all 40 project sites within the first five years of project implementation, which equates to ~0.8% of the combined area of these REDD+ projects. Rates of deforestation and degradation were closely correlated among projects (Spearman's rho= 0.82, $p<0.0001$).

When examining the subset of projects that had been operating for at least eight and ten years (n=24 and 14, respectively) we found no evidence of varying effect sizes through time, as we observed similar estimates of reductions in deforestation (Fig. 2.2-time *a*) and degradation (Fig. 2.2-time *c*) throughout these periods. Moreover, estimates of avoided deforestation and degradation were similar when control pixels excluded protected areas (Appendix S2-S4).

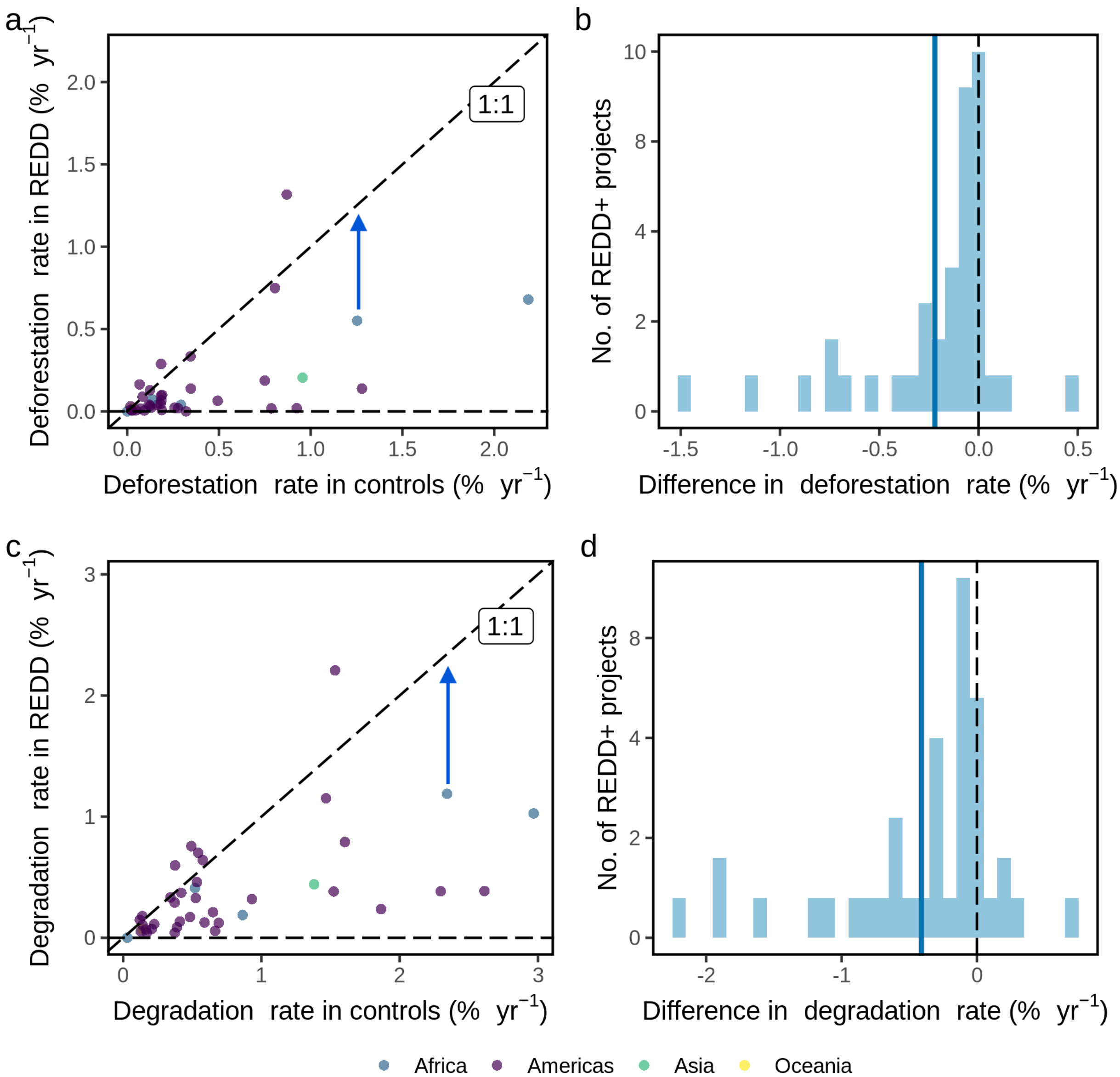

Figure 2.2: Changes in deforestation and degradation rates resulting from REDD+ projects over their first five years of operation. (**a**) and (**c**) scatterplots of deforestation and degradation rates in REDD+ projects versus matched control pixels; the reduction in deforestation or degradation resulting from a project is given by the vertical distance between the datapoint and the 1:1 line (blue arrows); (**b**) and (**d**) histograms of the differences in deforestation and degradation rates (relative to controls), with the mean shown as a blue line.

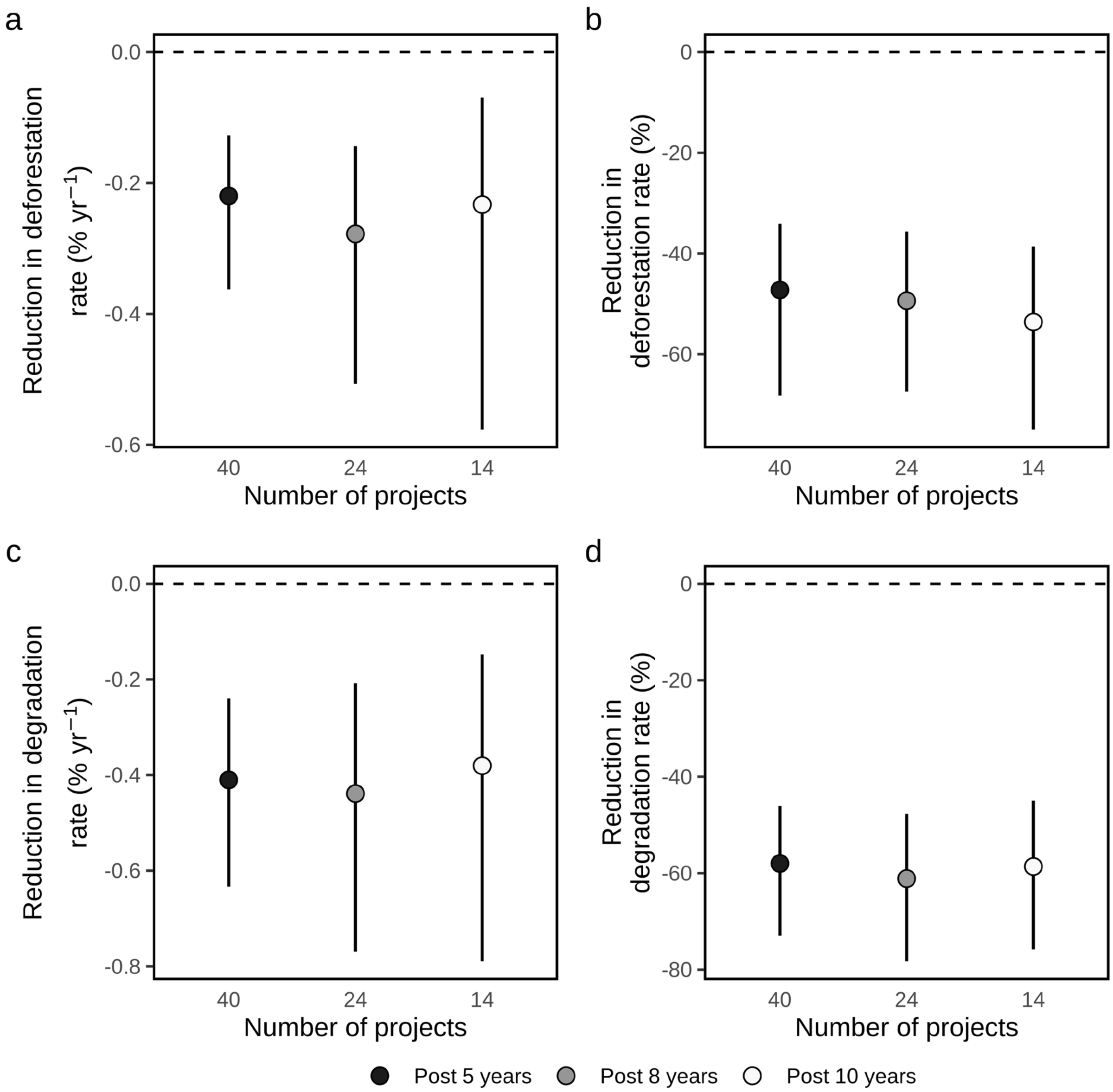

Figure 2.3: Avoided deforestation and degradation associated with 40 REDD+ projects, for three post implementation periods. (**a**) and (**c**) reductions in annual deforestation and degradation rates (means, with 95% confidence intervals); (**b**) and (**d**) percent reductions in deforestation and degradation rates (means, with 95% confidence intervals).

VARIATION IN REDD+ EFFECTIVENESS IN RELATION TO BACKGROUND DEFORESTATION RATES

We observed a moderate correlation between country-level background deforestation rates and reductions in deforestation (Spearman's rho= 0.42, p=0.006; Fig. 2.4 *b*) and reductions in forest degradation (Spearman's rho= 0.39, p=0.013). REDD+ projects in the low threat group showed small reductions in deforestation (mean= 0.16% yr^{-1}; CI = 0.07-0.28% y^{-1}) and degradation (mean= 0.33% yr^{-1}; CI = 0.16-0.58% yr^{-1}) rates. Significantly greater effect sizes were observed for the seven projects in the high threat group: deforestation was reduced by 0.52% yr^{-1} (CI = 0.25-1.0% yr^{-1}) and degradation by 0.79% yr^{-1} (CI = 0.42 - 1.32% yr^{-1}) (Fig. 2.4 **c**). We calculate that 49,197 ha of forest saved by REDD+ projects were in regions of high threat (i.e. 74% of the total saved) even though these forests only represented 20.5% of the total area of the 40 projects investigated. Therefore, in the high threat group, ~2.9% of the area of REDD+ projects were saved over the first five years. When measuring relative reductions in forest disturbances, we observed larger effect sizes in low threat groups compared to high threat groups, with mean reductions in deforestation of 52% (95% CI= 36% to 76%) and 25% (95% CI= 13% to 39%), and mean reductions in degradation of 61% (95% CI= 47% to 79%) and 43% (95% CI= 33% to 60%), in the low and high threat groups, respectively (Fig. 2.4 **d**).

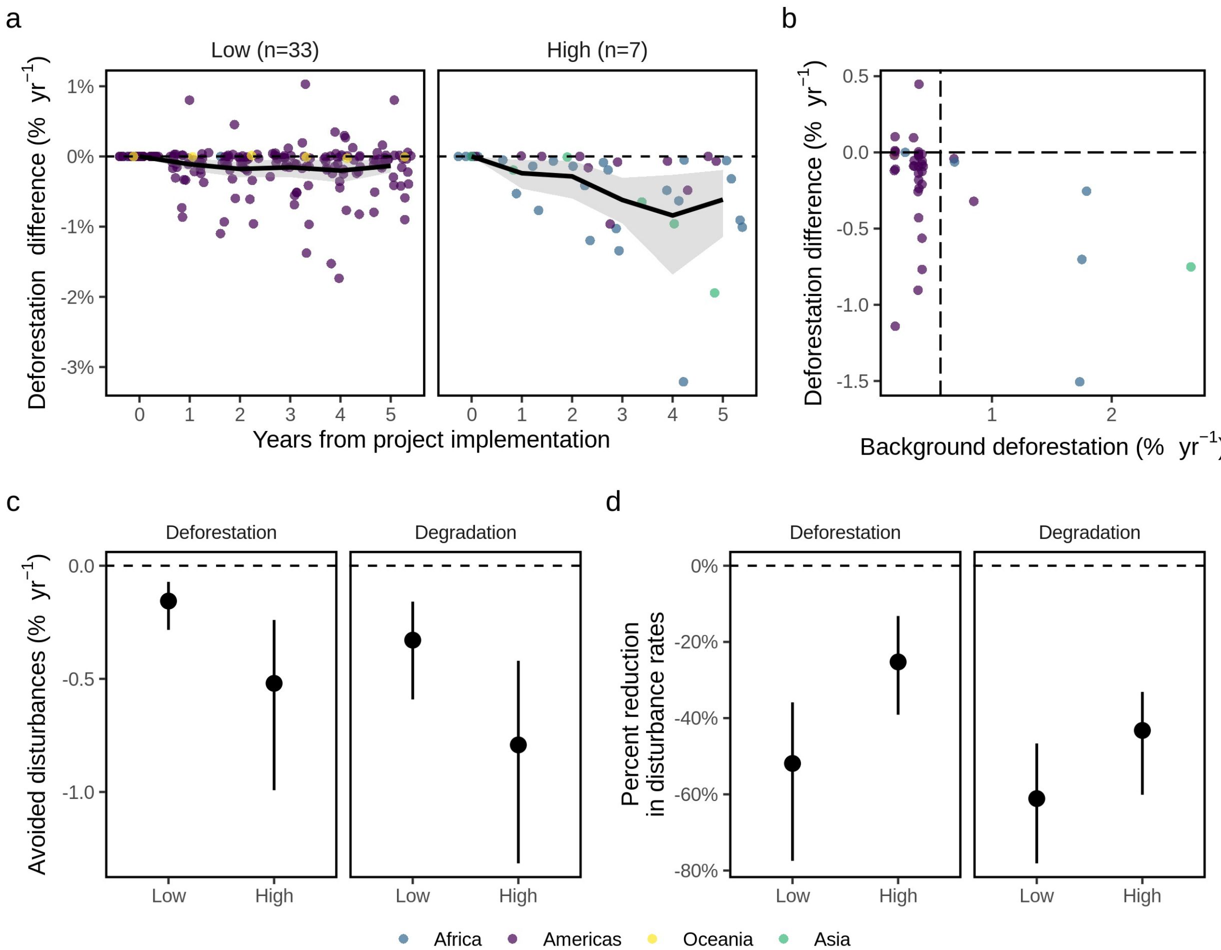

Figure 2.4: REDD+ project effectiveness in relation to background deforestation rates, for 40 sites in humid tropical forests. (**a**) annual differences in deforestation (with jitter; % yr^{-1}) between project areas and matched controls over five years after project implementation, with a black line showing the mean annual differences and 95% CI shaded in grey; (**b**) mean differences in deforestation rates (**%** yr^{-1}) against country-level background deforestation rates within the humid tropics (calculated for the project implementation period), with a vertical line showing the pan-tropical mean rate of deforestation (0.57% yr^{-1}); (**c**) mean differences in deforestation and degradation rates within regions categorised as having low (< 0.57% yr^{-1}) or high (> 0.57% yr^{-1}) deforestation rates, based on the average deforestation rate across the entire humid tropics; (**d**) mean percent reductions in deforestation and degradation rates relative to controls, within regions of high and low deforestation rate. The 95% CIs displayed at **a**, **c** and **d** were estimated using non-parametric bootstrapping.

Evidence of local leakage

Our tests provide no evidence of systematic local leakage of deforestation activities from project areas within the 10-km leakage belts, following project implementation. Three sites had significantly higher rates of deforestation in the leakage belts after project implementation, while two sites had significantly lower rates (bootstrapped t-tests, $p < 0.05$; Fig. 2.5). When examining the variation in leakage effects on subset of projects that had been operating for at least eight years we found one project with significantly higher rates of deforestation, while four sites showed a significant reduction in deforestation (Appendix A Fig. 1.10 **a**). We found no significant increase or decrease in deforestation rates for the subset of projects that had been operating for at least 10 years (Appendix A Fig. 1.10 **b**).

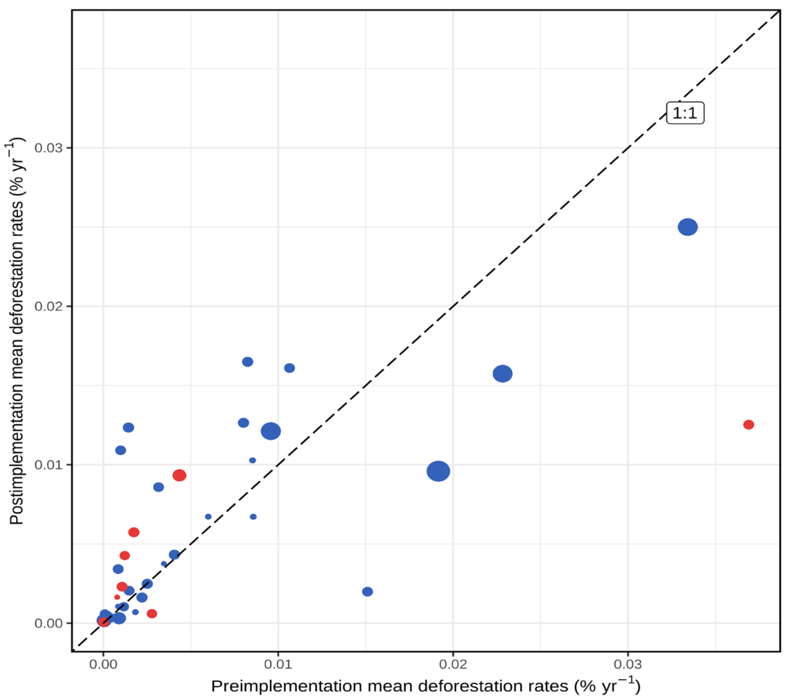

Figure 2.5: Evidence of deforestation leakage. Filled circles depict the mean rates of deforestation (% yr^{-1}) in the 10-km leakage belt in the five years before (x-axis), and after (y-axis) the commencement of projects. Red circles indicate significant differences in the post-implementation deforestation rates relative to the pre-implementation period (bootstrapped t-tests, $p < 0.05$), while blue circles indicate differences were not statistically significant. Circle sizes were scaled to reflect the background deforestation rates observed at the host country within the first 5 years of project implementation.

Discussion

Across 40 voluntary REDD+ projects in nine countries, on average REDD+ interventions reduced deforestation and degradation relative to control pixels over the first five years of operation. REDD+ projects achieved greater reductions in deforestation and forest degradation where the threat of deforestation was greatest (which in our sample included three projects in Madagascar, one in Belize, one in Brazil, one in Cambodia, and one in Democratic Republic of the Congo). Similarly, our analyses showed that REDD+ projects were also effective at reducing degradation. The absolute reductions in these rates were modest in most projects, but in relative terms both rates were roughly halved by the projects we investigated. To the best of our knowledge, this is the first study that has use remotely sensed degradation and deforestation data to show that voluntary REDD+ projects were effective at reducing small-scale temporary disturbances, alongside long-term deforestation, across a sample of geographically dispersed projects that differ in deforestation drivers and social objectives. In theory, protecting and restoring natural forests could be a nature-based climate solution that is cost effective, if the many hurdles to implementation can be overcome (Duchelle et al. 2018a; Milne et al. 2019; TSVCM 2021). This paper provides some room for optimism: despite the many challenges to just and economically sustainable implementation, the initial wave of REDD+ projects are effective at reducing forest loss. Pressures on biodiverse tropical forests are expected to increase in the future (Laurance et al. 2014; Barlow et al. 2018), making the evidence that REDD+ has reduced deforestation and degradation where threat is high particularly important.

Estimating the impact of an intervention, such as REDD+, from observational data is inherently difficult as it relies on estimating what would have happened in the absence of the intervention (Ferraro 2009; Ferraro & Hanauer 2014; Baylis et al. 2016). We have matched our REDD+ and control pixels on appropriate drivers of deforestation, but there will inevitably be unobserved covariates. For example, VCS REDD+ projects have often been implemented in areas where conservation NGOs have been operating for some time (Sunderlin & Sills 2012; Usmani et al. 2018) and are likely associated with certain land tenure conditions (Wunder 2013; Wunder et al. 2020b). Characterising these social and institutional dimensions at the site level can be challenging, let alone at the wider landscape level from which control observations were selected. This has implications for the selection of appropriate controls and therefore our results. For example, where REDD+ projects are the most recent manifestation of longer-running conservation efforts at sites (Lin et al. 2012; Sunderlin & Sills 2012), it is not possible to say how much reductions in deforestation is due to the REDD+ specifically, given the long-term engagement at these landscapes.

Different methods can be used to derive impacts of forest interventions when temporal observations are available for treatment and control groups, before and after project implementation, such as combining matching with difference-in-differences (Arriagada et al. 2012; Costedoat et al. 2015; Santika et al. 2021), or using synthetic control methods to ensure similar pre-treatment deforestation rates between treatment and control groups (West et al. 2020; Correa et al. 2020). We use a simpler approach: matching units with similar modelled deforestation risk without any pre-treatment comparisons (e.g. Rasolofoson et al. 2015; Eklund et al. 2016; Geldmann et al. 2019). While this approach has limitations (Schleicher et al. 2019a), it was most appropriate for our context: VCS requires projects, when delimiting their boundaries, to exclude from the accounting area any locations where deforestation has taken place in the 10 years prior to the project start date (Shoch et al. 2011). Therefore, the rates of deforestation in the treatment groups are by definition zero due to active exclusion of deforested pixels, and thus we applied the same constrains when selecting candidate control observations. For the same reason, while we could have combined matching with difference-in-differences, we are restricted to an after-only analysis because deforestation rates in the before period is zero, for both treatment and control groups. Nevertheless, we account for pre-treatment deforestation rates and related deforestation risks by selecting pixels with similar distance to recent forest clearings, which is the strongest predictor of deforestation outcomes in the landscapes we examined. Moverover, while acknowledging these limitations our approach to estimating the impact of REDD+ differs fundamentally from that of VERRA’s, which lacks counterfactual designs (Balmford et al., 2023). VERRA's methodologies involve a before-after comparison, in which the observed changes in forest cover are compared to an extrapolation of forest loss trajectories based on historical trends. Our approach involves identifying comparable control groups against which the changes in deforestation rates are assessed over time. The counterfactual approach in our analysis represents a methodological advantage relative to VERRA’s methods (Balmford et al., 2023).

Leakage is challenging to quantify as it requires the characterisation of enabling factors, such as labour and market conditions (Pfaff & Robalino 2017), with which to produce a forecast of potential displacement (or shielding) of deforestation activities. While acknowledging the limitations of our approach, by examining statistical differences in deforestation after projects became implemented, we did not observe strong evidence of systematic leakage effects into the buffer zones adjacent to REDD+ projects. However, leakage can occur across countries, through international market adjustments in response to local restrictions (Meyfroidt & Lambin 2009; Delzeit et al. 2018), but is very hard to quantify and was not accounted for in our study.

As our understanding of the enormous carbon stores in tropical forest ecosystems improves (Dargie et al. 2017), and as we gain further understanding of the feedbacks between tropical deforestation and climate change (Baccini et al. 2017; Wigneron et al. 2020), the case for tropical forests being central to climate change mitigation efforts grows stronger. Our analysis shows promising evidence that site-based REDD+ projects have helped to reduce deforestation, particularly in areas of higher deforestation threat. Yet, emissions reductions in the 40 REDD+ projects analysed represent a tiny fraction of global emissions: in total they amount to about 0.01% of 2018 emissions, or 0.13% of emissions from tropical deforestation in 2013. Considering that deforestation accounts for the second largest source of emissions after the energy sector, estimated at ~11% (Shukla et al. 2019), efforts to prevent forest-based emissions need to grow at scale to counteract current levels of forest-loss emission rates. Jurisdictional REDD+ programs, operating at regional or national scales following UNFCCC REDD+ framework of 2013, may address some of the major challenges faced by site-based REDD+ projects (Duchelle et al. 2019). Most importantly, larger-scale efforts may be better placed to address the fundamental challenge that key drivers of deforestation are embedded in global and domestic supply chains for commodities such as beef, palm oil and soya (Curtis et al. 2018; Pendrill et al. 2019; zu Ermgassen et al. 2020), so cannot be effectively tackled at site level (Delabre et al. 2020/ed). Encouragingly, there is evidence that jurisdictional programs can deliver results: Guyana's national-level program reduced tree cover loss by 35% between 2010 and 2015 (Roopsind et al. 2019). Applying the lessons from the last few decades to deliver effective and, crucially, equitable reductions in tropical forest degradation and deforestation will be critical if the Glasgow COP26 climate change objectives are to be met.

CHAPTER 3: SENSITIVITY OF AVOIDED DEFORESTATION ESTIMATES TO MATCHING SPECIFICATIONS

ABSTRACT

Reducing Emissions from Deforestation and Forest Degradation (REDD+) projects need to demonstrate a decrease in forest disturbances relative to a business-as-usual scenario (i.e. baseline) to be considered effective. Matching is becoming the preferred method for evaluating conservation interventions and making causal claims about their impact. Different a-priori valid matching specifications can result in different REDD+ baselines, which in turn affect estimates of REDD+ effectiveness. Here we examined 40 REDD+ projects spread across the tropics and explored the sensitivity of forest loss reductions to alternative matching specifications. We explored how variations in matching parameters such as algorithms, caliper size, allowing matching with replacement, and choices of covariate combinations impacted estimates of REDD+ effectiveness. To filter out matching options that delivered inadequate counterfactuals, we validated matched sets against four criteria: 1) post-matching covariate balance between treated and selected control samples, 2) proportion of units matched, 3) similarity of deforestation rates in the period prior to REDD+ (i.e. "parallel trends"), and 4) sensitivity of effect sizes to hidden confounders. Working with this filtered dataset, we found reductions in forest loss rates in 33 of the 40 sites, with an estimated mean of -0.08 % yr^{-1} (95% CI= -0.04 to -0.14 % yr^{-1}) when averaging across baseline scenarios. The median within-project standard deviation of baseline estimates was estimated at 0.043 % yr^{-1} (IQR= 0.02 to 0.084 % yr^{-1}), highlighting the variability in potential outcomes identified among robust matched sets selected as controls. Our results highlight the sensitivity of avoided deforestation estimates to variations in matching parameters and call attention to the importance of integrating structured robustness checks in counterfactual design.

INTRODUCTION

The need to robustly evaluate the effectiveness of conservation projects such as REDD+ is well established (Baylis et al. 2016). However robust evaluation is non-trivial because interventions are not designed as controlled experiments, so great care is needed to select meaningful, unbiased counterfactuals in often complex human-modified landscapes (Schleicher et al. 2017). Reducing Emissions from Deforestation and forest Degradation (REDD+) projects represent one of the most important initiatives for financing tropical forest conservation, with a growing adoption among participant countries (Wunder et al. 2020b). To access REDD+ benefits, projects need to demonstrate reductions in deforestation relative to a baseline scenario, e.g., a counterfactual scenario of the expected deforestation had the project not been implemented. Far from occurring at random, areas where REDD+ has been implemented meet conditions that make them more prone for project development, including exposure to local drivers of deforestation (Lin et al. 2012) and the pre-existence of social and institutional conditions that facilitate project implementation (Sunderlin & Sills 2012; Usmani et al. 2018). It is therefore crucial to account for such confounding factors when estimating the effectiveness of projects at reducing deforestation.

Matching is fast becoming the preferred tool for evaluating the impact of forest conservation interventions (Schleicher et al. 2019a; Börner et al. 2020). Matching refers to a suite of statistical methods that allow causal inferences from observational data, by pairing sets of treated and control groups that are as similar as possible with respect to confounders (Stuart 2010). The similarity between treatment and control groups is measured by the empirical overlap of confounding factors. Procuring a dataset with balanced confounder distributions is necessary to meet the “ignorability” assumption of observational data analysis: assignment to the treatment group could have taken place “as if random”, once conditioned on such confounders (Rosenbaum 2002; Angrist & Pischke 2008). Provided enough overlap exists between treated and control units on such covariates, the treatment effect is more likely to be correctly estimated and less prone to model dependence (Ho et al. 2007).

Matching in practice involves making choices on multiple parameters when defining the approach for selecting potential controls. Some choices are made based on the questions being addressed and the availability of data for each treatment regime (Stuart 2010). Others are guided by ongoing debates, for instance when determining which algorithms to use when assessing the similarity between treatment and control groups (i.e., Austin 2011; King & Nielsen 2019), or when selecting a caliper size to filter potential control observations based on their location in the covariate distributions (Rosenbaum 1985). In practice, these and other important parameters such as the refinement of covariate selection to characterise confounders are commonly chosen *ad-hoc*, for instance by iterative testing with the aim of maximising the quality of matched sets (as in Schleicher et al. 2019a). Each of these parameters can result in different sets of matched controls, each associated with different potential outcomes, which in turn can result different estimates of effect size. Although recommendations exist in the applied literature for determining the robustness of matched sets after matching is performed (Stuart 2010; Schleicher et al. 2019a), few studies have critically examined whether multiple robustly matched sets, produced with different matching parameters, result in meaningful variation of counterfactual (i.e. baseline) estimates, and therefore of treatment effects.

Biases can also arise from confounding variables omitted from the matching procedure (Cinelli & Hazlett 2020; Jones et al. 2022). In applications such as evaluating the effectiveness of REDD+, unobserved confounders can come about from variables that correlated with selection to treatment but are nonetheless hard to quantify. For instance, social and institutional factors that correlate with REDD+ adoption, such as land tenure and presence of local conservation initiatives (Lin et al. 2012; Sunderlin & Sills 2012; Wunder et al. 2020b; Usmani et al. 2022) are challenging to characterise across forested landscapes, and so are seldomly included in matching assessments. However, given the significant role of such factors in determining where REDD+ projects are established, and eventual project outcomes (e.g., Sunderlin & Sills 2012; Wunder et al. 2020b; Usmani et al. 2022), their omission could result in sizeable biases in effect size estimates. Despite the likelihood of unobserved confounding in conservation impact evaluations, the use of tests to assess the sensitivity to hidden bias remains limited (Jones et al. 2022).

Here we explored the sensitivity of REDD+ effectiveness to alternative matching specifications by performing matching runs with variations in the choice of algorithm, caliper size, allowing matching with replacement, and covariate combinations. We assessed the resulting matched sets against four robustness metrics: 1) post-matching covariate balance between treated and selected control samples, 2) proportion of units matched, 3) similarity of deforestation rates in the period prior to REDD+ (i.e. "parallel trends"), and 4) sensitivity of effect sizes to hidden confounders. We then filtered out poor quality matches and explored the variability of potential outcomes (i.e., baseline estimates) among quality-assured matched sets. We finally summarised the effect size for each REDD+ project combining all robustly matched sets.

Methods

We used a structured approach to build a database of forest cover and forest loss, perform matching, and evaluate the robustness of matching results. A schematic of the study design is provided in (Fig. 3.1) and developed in the following sections.

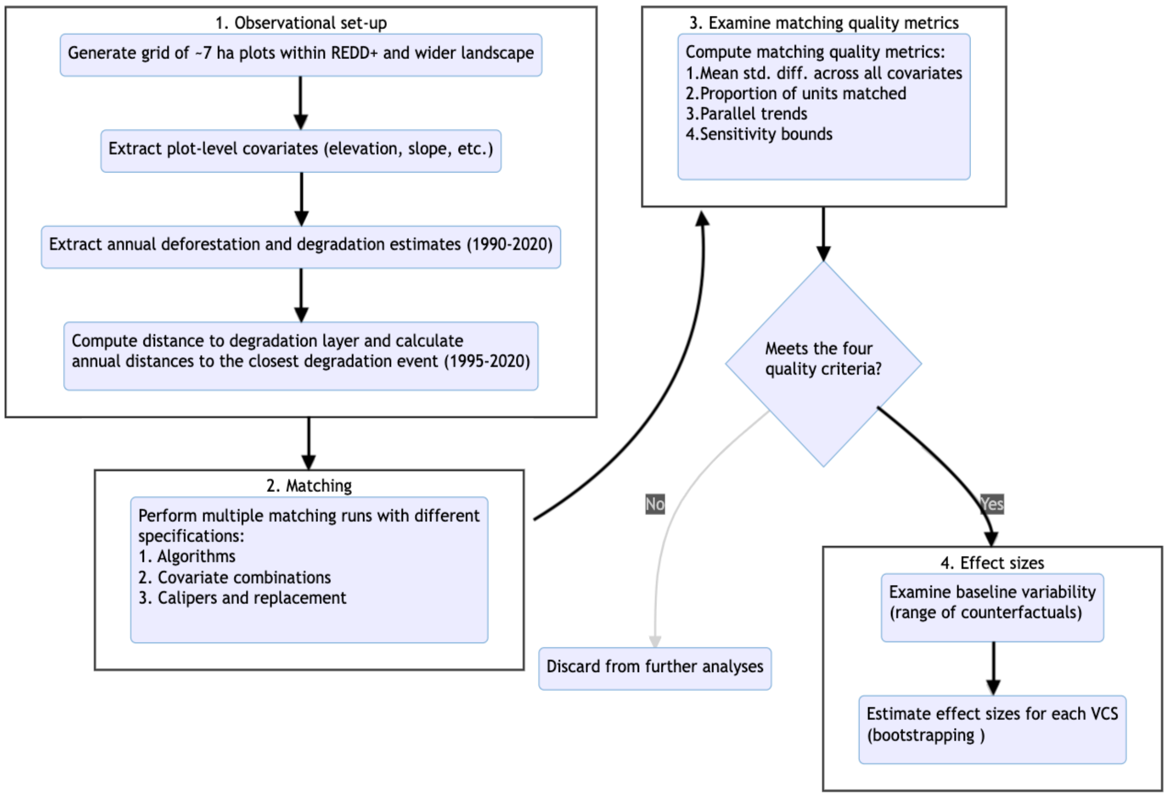

Figure 3.1: Schematic of pipeline used to evaluate the impact of matching parameters on estimates of REDD+ effectiveness. Blocks indicate components of the analysis and are described in more detail in the methods.

1. Observational set-up

We constructed a database of REDD+ projects (i.e. interventions) and potential controls observations following a similar approach to that of (Guizar-Coutiño et al. 2022). To characterise areas under REDD+ projects, we collected boundary data of interventions certified by the Verified Carbon Standards (VCS). To characterise outcomes, we constructed a time-series of annual deforestation rates across the moist tropics as characterised by Tropical Moist Forest (TMF) map (Vancutsem et al. 2021). We collected data on deforestation risks factors, including key bioclimatic and socio-demographic determinants of deforestation, across treated areas and forested areas from the wider landscape (outside REDD+ interventions) that overlapped with the TMF. This approach allowed us to produce sets of treated and control groups via matching with which to estimate the impact of site-based REDD+ projects at reducing deforestation.

Definition of treatment areas, comparison zones and sampling units. To assess the extent of reductions in deforestation due to REDD+ we compared changes in forest cover within treated areas with that taking place in similar forest patches within the wider landscape. We generated a regular grid of circular plots, each with 150 m radius (~7 ha) and spaced 1 km apart, within forested areas of the REDD+ boundaries. We considered the plot size adequate for characterising local deforestation dynamics and determinants of forest loss (Avelino et al. 2016). The extent of forest cover in 1990 forest cover in 1990 as delineated by the TMF (Vancutsem et al. 2021) was used as the reference area for sampling. In doing this we ensured that all samples contained at least some portion of moist forest cover when constructing the time-series of forest cover change (see *Characterising forest cover, forest disturbance and determinants of deforestation.*). We used a similar approach when collecting samples from the wider landscape, which were later used to select control groups. First, we generated a grid of 150 m radius plots, with centroids separated by 1 km, across tropical moist forests outside treatment areas, excluding a 15 km buffer zone that extended from the project area boundaries to avoid sampling form areas where local *leakage* effects could have taken place (e.g. deforestation displacements outside project area boundaries due to the intervention). Secondly, we selected random plots from the grid of samples computed for the whole landscape, keeping a ratio of up to 5:1 landscape samples for each plot inside REDD+ areas. This resulted in the selection of landscape-level plots across comparable tropical moist forests zones for matching (Fig. 3.2). Separation between samples reduced the spatial dependence of plots within treatment regimes (Robalino & Pfaff 2012). Moreover, by defining 7 ha plots through landscapes and REDD+ interventions we were able to collect data on the temporal dynamics of forest cover and forest disturbances, which was used at a later stage to assess the quality of the matches, by establishing the similarity of forest cover trajectories between REDD+ and selected landscapes samples.

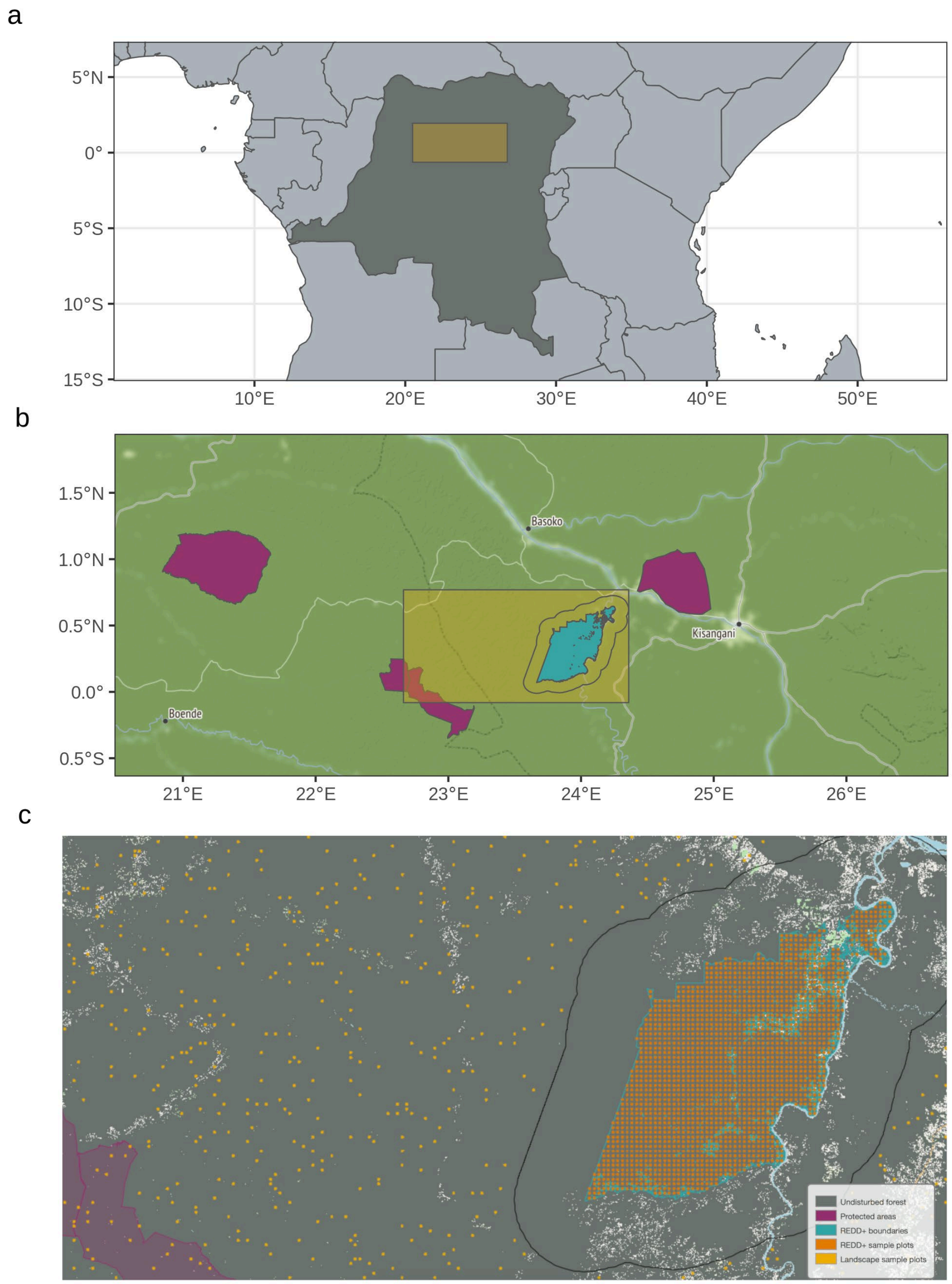

Figure 3.2: Schematic of the sampling design for a REDD+ project in Democratic Republic of the Congo. Yellow squares in *a* and *b* represent the boundaries for panels *b* and *c*, respectively. Features shown in the maps include protected areas (purple), REDD+ area (VCS 1359, cyan), treatment samples (orange), landscape samples (yellow). Leakage boundaries surrounding project area shown in black.

To conduct our analysis, we selected REDD+ sites with at least 80% forest cover, according to the TMF product, at the project start date. We also restricted our selection to projects that started operating between 2000 and 2014, to allow for at least 5 years of observations before and after project implementation, and because projects beginning before 2000 did this prior to the arrival of REDD+. We extracted boundary information for 71 projects from the VCS database, but 22 REDD+ projects were excluded because they did not contain high enough forest cover, and 5 more were excluded because they were not operating over the selected time period, bringing the number of projects available for examination to 44.

Characterising forest cover, forest disturbance and determinants of deforestation. Annual estimates of forest cover and deforestation were extracted from the Tropical Moist Forests (TMF) database (Vancutsem et al. 2021), which was derived from Landsat imagery with pixels of ~30 m^2 resolution. This database provides a long-term characterisation of forest disturbances and posterior land cover trajectories, on an annual basis spanning the period 1990 - 2019. Our analyses focused on examining temporal patterns of three main forest classes: The *undisturbed* class, which refer to closed evergreen or semi-evergreen forest areas that have not been degraded or deforested; the *degraded* class, which is characterised as a 'disturbance in tree canopy cover that is visible from space over a short time period (less than 2.5 years)', resulting from anthropogenic causes such as selective logging or from causes such as wind storms or fires; and the *deforested* class, representing long-term forest disturbances and complete removal of forest cover for all sampling plots (Vancutsem et al. 2021). For each year during the 1990 - 2019 period, we calculated the extent of undisturbed, and disturbed forests classes within sampled circular plots. From these metrics we derived our outcome of interest - the proportion of forest lost - defined as $r_t = \Delta p_t / p_d$, where Δp_t is the total area deforested and degraded in year t and p_d is the area of undisturbed forest at the start of the time series (1990).

We identified key covariates associated with deforestation (Busch & Ferretti-Gallon 2017), some of which were also likely to have been associated with assignment to the treatment (Lin et al. 2012; Sunderlin & Sills 2012). To characterise deforestation risks factors, we collected data on key and observable bioclimatic and socio-demographic determinants of deforestation (as in Busch & Ferretti-Gallon 2017). For each circular plot, we obtained mean estimates for elevation and slope (Jarvis et al. 2008), distance to the nearest urban centre (Weiss et al. 2018) and distance to forest edge (Ewers et al. 2011). To account for distance to forest edge, we produced annual time-series of the distance to the nearest deforested pixel, as characterised by the TMF map. First, for each pixel covered by the circular plots, we computed pixel-level estimates of the distance to the nearest pixel that had changed its status from undisturbed to deforested, or from degraded to deforested, during the observed year, for the period 1990-2019. We then produced plot-level annual rolling estimates of the mean distance to deforestation events in the previous five years, covering the period 1995-2019.

2. Matching

We used statistical matching to select control plots from the wider landscape around each REDD+ site, consisting of plots with similar initial forest cover density, exposed to a set of similar drivers of deforestation, which also exhibited similar trajectories of deforestation in the period prior to REDD+. These plots would allow us to evaluate the effectiveness of REDD+ in avoiding forest loss, i.e. what would have been the rates of deforestation inside REDD+ plots, had these not been protected? We identified key covariates associated with deforestation (Busch & Ferretti-Gallon 2017), some of which were also likely to have been associated with assignment to the treatment (Lin et al. 2012; Sunderlin & Sills 2012). We defined a key set of matching covariates (i.e., “base” combination) which included: 1) elevation, 2) slope (Jarvis et al. 2008), 3) distance to population centres (Weiss et al. 2018), 4) distance to degraded areas over the five years prior to project commencement and 5) country and 6) biome. By matching tropical moist forests plots from the same biomes and countries we ensured comparability of bioclimatic conditions for agricultural activities and of national governance.

To evaluate the sensitivity of effect sizes to matching specifications, we performed multiple matching runs iterating over a set of matching parameters across 3 components. We ran matching trials with alternative specifications on: 1) matching algorithms, where we determined the similarity of samples with the propensity score (PSM), Mahalanobis distance (MHN) (Legendre & Legendre 1988) and random forest (RFM) algorithms; 2) matching parameterisation, where we included runs of matching with and without replacement, and tested with multiple caliper sizes so control observations lied within 0.1, 0.3, 0.5, 0.7 and 0.9 SD from treatment observations in their covariate distributions; and 3) by adding up to two covariates to the key set of matching covariates (i.e., base), which included: a) extent of undisturbed forest in 1990 (forest area 1990), b) extent of undisturbed forest in the implementation year (forest area), and c) both additional covariates in the same matching run. This resulted in a total of 4 covariate combinations: base, base + forest area (1990), base + forest area, and base + forest area (1990) + forest area. The combination of all matching parameters resulted in 120 matching runs for the 44 selected REDD+ sites, giving a total of 5280 runs.

3. Metrics of matching quality

In the following section we describe the procedures for computing metrics of matching quality. Matching runs were assessed on four criteria: 1) post-matching covariate balance between treated and selected control samples; 2) proportion of treatment units matched and sampling bias; 3) similarity of deforestation rates in the period prior to REDD+ (i.e. “parallel trends”), and; 4) sensitivity of effect sizes to hidden confounders.

Covariate balance between treatment regimes. Controls units are required to be as similar as possible to the treated group to be considered credible counterfactuals. To assess the similarity of treated and control groups, we determined an absolute standardised mean difference of <0.25 between treated and control samples across all covariates included in the matching run, following (Stuart 2010).

Proportion of units matched. The proportion of matched plots in the treatment group defines the extent to which the estimations of effect sizes can be generalised across the target population (Stuart 2010). When sufficient treated plots are matched, computing effect sizes result in the estimation of the *average treatment effect on the treated* (ATT). For each matching run, we report the proportion of plots matched and specify a minimum of 80% of matched treated samples for a run to be considered valid.

Parallel trends. Causal inferences from time-series data require an assumption that the treatment and control groups followed similar outcome trajectories in the period prior to the intervention (e.g. parallel trends; Angrist & Pischke 2008). We examined the evidence of variation in forest loss rates between matched sets to establish the similarity of outcomes between treated and control groups prior to REDD+. We ran the following linear mixed-effects model to test for significant departures from parallel trends:

$$\text{PropAreaforest}_{\text{it}} = \beta_0 + \beta_1 \text{Year}_t + \beta_2 CI_i + \beta_3 Year \times CI_{\text{it}} + \varepsilon \qquad (1)$$

Where $\text{PropAreaforest}_{\text{it}}$ refers to the proportion of undisturbed forest area in the sample i at time t, Year_t refers to year t,and CI_i is the treatment group indicator (treated = 1 or control = 0). $\beta_3 Year \times CI_{\text{it}}$ is the parameter that allows us to test for significant differences in the trend lines of treated and control plots. We performed these tests using all available years before the project starting date. Departure from the parallel trend assumption was assessed by statistically testing whether the slope of the treatment and control groups were different, i.e. whether the interaction term β_3 was significant at $p< 0.05$. We provide a visual example of a run meeting the parallel trends assumption in the Appendix, Fig. B.4.

Robustness to hidden bias. In the absence of controlled experimental settings, the possibility of hidden omitted variables can never be ruled out. Tests of the sensitivity of effect sizes to omitted variables allow us to explore the risk of hidden confounding factors affecting our estimates (Cinelli & Hazlett 2020). We conducted a sensitivity analysis to obtain a residual variance threshold (RV) which refers to the amount of residual variation that an unmeasured confounder should explain in both the outcome and treatment term to neutralise the observed effects (Cinelli & Hazlett 2020). The RV value is useful in determining the strength that unobserved confounders need to have in order to eliminate our conclusions. After estimating the RV value, a further step involves assessing whether confounders with such strengths are plausible. This can be established by ‘bounding’ the strength of unobserved confounders relative to covariates that have been included in the analysis. We estimated the sensitivity bounds in relation to *distance to degradation*. This allowed us to determine the upper values of the residual variation that a(n) unobserved confounder(s) as strong as distance to degradation (X) could explain of the outcome ($R^2_{Y\sim Z|D,X}$) and the treatment ($R^2_{D\sim Z|X}$) (Cinelli & Hazlett 2020). If RV is larger than the residual bounds for the outcome and treatment ($R^2_{Y\sim Z|D,X}$ and $R^2_{D\sim Z|X}$ respectively) it suggests that the results are sensitive to an unmeasured confounder with the magnitude of distance to degradation.

To perform the sensitivity test, we first estimated effect sizes for each matched set with a difference-in-differences linear model, using the formula:

$$\text{PropForestLoss}_{\text{it}} = \beta_0 + \beta_1 BA_t + \beta_2 CI_i + \beta_3 BA\ \times CI_{\text{it}} + \varepsilon \qquad (2)$$

Where $\text{PropForestLoss}_{\text{it}}$ is the proportion of forest area lost in plot i at year t, and $\beta_1 BA_t$ and $\beta_2 CI_i$ are dummy indicators for period (before=0, after=1) and group (0=control, 1=REDD+). $\beta_3 BA\ \times CI_{\text{it}}$ is the difference-in-difference parameter that allows us to test for significant changes in the proportion of forest loss in the treated group after project implementation. To quantify the sensitivity to hidden confounding, we first estimated the residual variation value (RV) and computed the sensitivity bounds in relation to a hidden confounding as strong as distance to degradation. We then computed the differences between RV and the residual bounds for the outcome and treatment ($R^2_{Y\sim Z|D,X}$ and $R^2_{D\sim Z|X}$ respectively), and considered projects to be robust if the difference were above zero.

4. Selection of matching models

Only matching runs that met the minimum quality criteria were selected to examine the impact of REDD+ interventions on reducing deforestation. To pick models with which to evaluate the impact of REDD+ we specified the following criteria: 1) an absolute standardised mean difference of <0.25 between treated and control samples across all covariates as included in the matching run, 2) A minimum of 80% of treated samples matched, 3) evidence of no departure from parallel trends in pre-treatment deforestation rates, and 4) acceptable sensitivity to unobserved confounders, quantified as positive values for both the outcome and treatment boundaries. The filtered set of matching runs are henceforth described as the "robustly matched set".

5. Variations in potential outcomes (e.g. baseline levels) due to matching specifications

We examined variations in baseline levels by comparing the range of potential outcomes within robust matched sets. Baseline were extracted from the main effect model (Eq. 1) by combining the $\beta_0 + \beta_1 + \beta_2$ term estimates, as it pertains to the 'counterfactual' deforestation rates (e.g. deforestation in the REDD+ area, had it followed the same trajectory as that of the controls. We report the median and inter-quartile range (IQR) of baseline references for each site. We computed site-level standard deviations to quantify variations in baseline estimates.

6. Effect size estimation

To quantify the effect of REDD+ in reducing deforestation we pooled effect sizes for each REDD+ site using the model runs that met quality criteria. Average site-level effects were estimated by computing the mean of the treatment effect term ($\beta_3 BA \times CI_{\mathrm{it}}$, in Eq. 2) from the main effects models, and then estimated 95% confidence intervals by bootstrapping β_3 terms. To produce a global estimate of the impact of REDD+ in reducing deforestation, we computed the mean of the site-level averages and estimated 95% confidence intervals by bootstrapping on the site-level averages.

RESULTS

We ran 5280 marching runs on 44 sites, which included 34 in the Americas, 7 in Africa and 3 in Asia-Pacific.

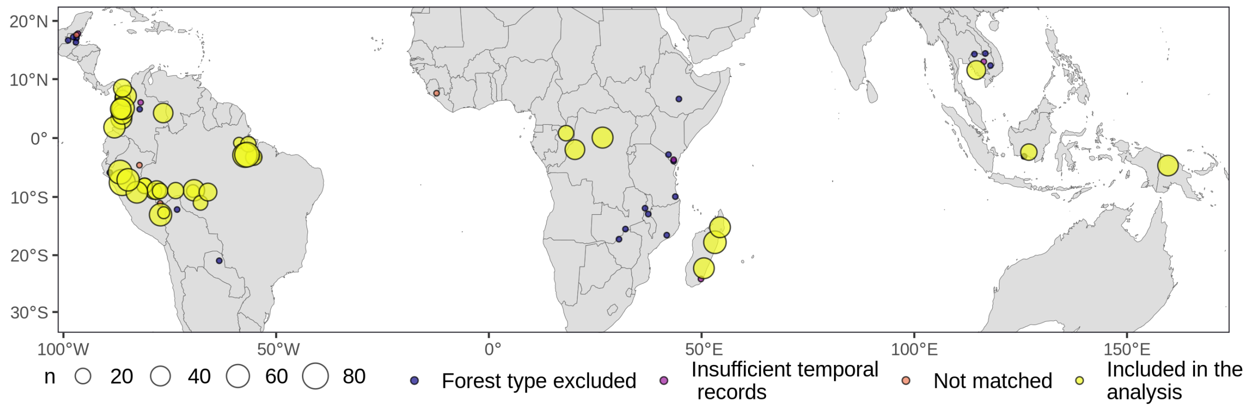

Figure 3.3: Location of the REDD+ projects included in the analysis. Of the 71 mapped sites, 22 had less than 80% evergreen forest cover at the start of projects (purple), 5 did not meet our time frame criteria, and 3 could not be matched with appropriate control pixels (orange), leaving 40 projects in the final dataset (yellow). Number of robustly matched sites shown the dot size.

Of the REDD+ sites that didn't produce a single valid matching run, sites 1326 and 1133 consistently didn't achieve the minimum proportion of matched observations required to be considered valid (of 80%), while sites 1201 and 844 did not produced models that met all the 4 inclusion criteria simultaneously, (e.g. models met some of the 4 criteria but were lacking in others when looking at a run by run basis).

Selection of valid models

Almost a third of the matching trials resulted in robust matched sets when assessed against the four quality criteria (n=1550) spread across 40 REDD+ sites. Selected sites were located in America (n=32), Africa (n=6) and Asia-Pacific (n=3) and had an average of 39 robust models (IQR= 24-49) with which to assess their effectiveness (Fig. 3.3). When examining factors that resulted in robust matches, runs that allowed matching with replacement and those using Mahalanobis distance as a metric of similarity led to the highest number of successful matches (Fig. 3.4). Taken together, the combination of factors that led to the highest number of successful matches (n=38 sites) was Mahalanobis matching with replacement, using the core matching variables (elevation, slope, accessibility and distance to degraded forests) as well as measures of forest area in 1990 and at the start of projects, with a 0.7 caliper. Further examination of the variations between matching factors and metrics of matching success suggests that the algorithm and covariate combination described above were significantly better than the alternatives in achieving robust matched runs (Appendix B).

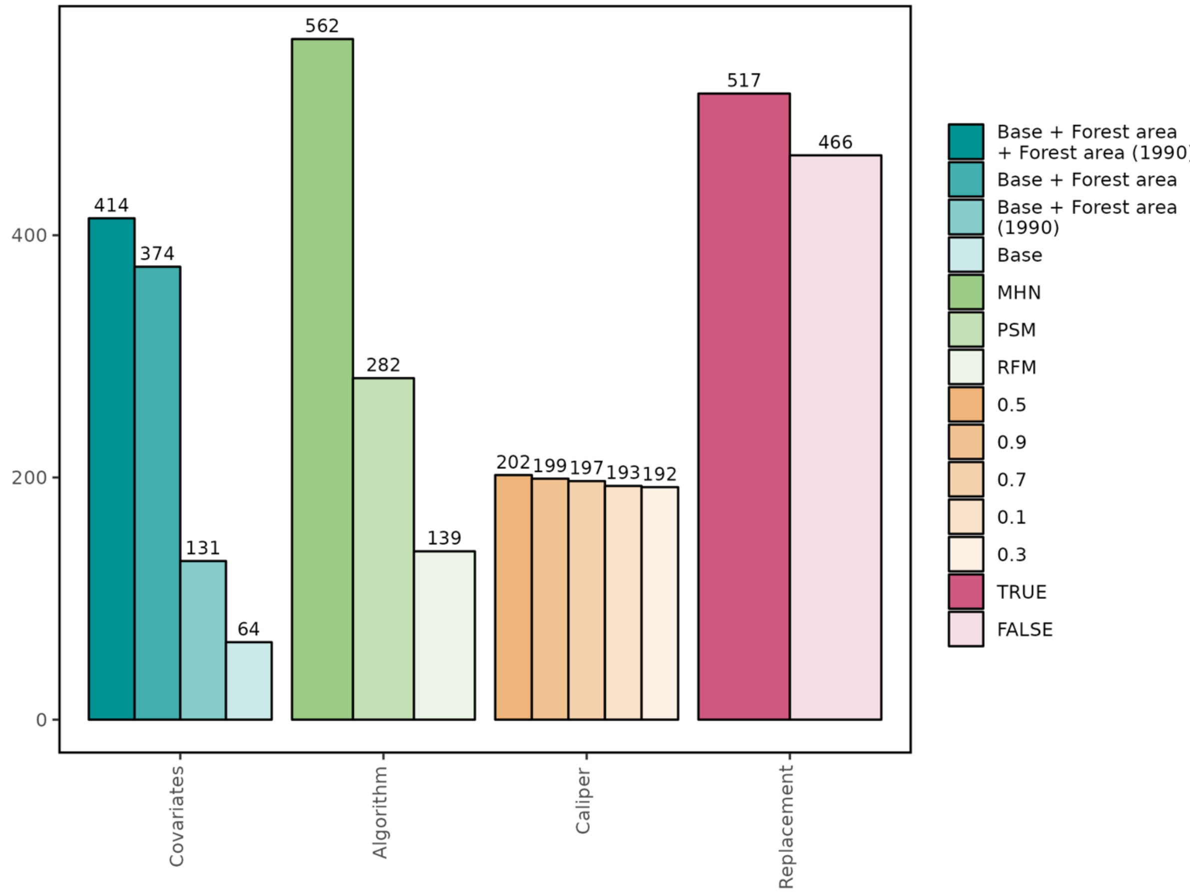

Figure 3.4: Count of successful matched sets by matching parameters. A total of 983 matching runs met all the quality criteria. Barcharts show counts of successful matched runs for each parameter (i.e. choice), in each of the examined factors: Covariates (aquamarines), algorithms (greens), calipers (oranges), and matching with replacement (pinks). The parameters tested included: a) 3 algorithms: Mahalanobis (MHN), propensity score (PSM), and random forest (RFM); b) 4 covariate combinations: the minimum set of covariates ('Base': elevation, slope, accessibility, distance to degradation, country and biome), and iterations that included forest area at the time of project implementation (forest area) and forest area at the beginning of the time series (forest area 1990); c) allowing to match with replacement, and; d) 5 caliper sizes: of 0.1, 0.3, 0.5, 0.7 and 0.9 SD.

Effect sizes

Overall, REDD+ projects led to significant reductions of forest loss over the period examined. Combining estimates of selected models with bootstrapping, we observed reductions in forest loss rates in 29 of the 40 examined sites (Fig. 3.5). Combining mean site-level estimates computed with quality-assured matched sets, we observe a mean reduction of 0.08 % yr^{-1} (95% CI= 0.04 to 0.14 % yr^{-1}) in the rates of forest loss compared to controls. Changes in the area of forest loss, although small in absolute terms, equates to a relative reduction of 33% (CI = 13-49%) over the examined period compared to the rates of forest loss observed in controls.

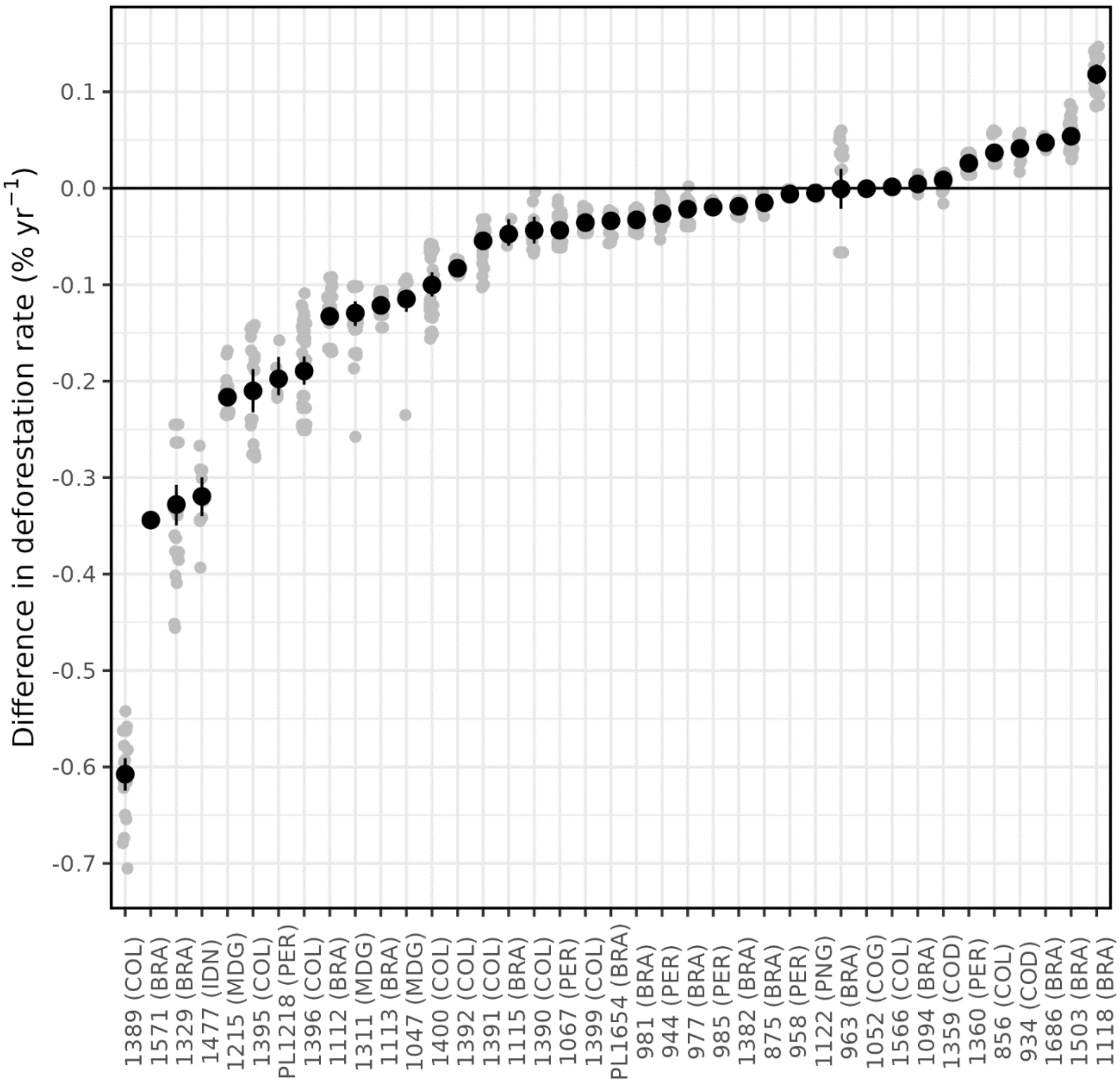

Figure 3.5: Change in forest loss rates due to REDD+ computed with robustly matched sets. Grey dots depict mean site-level difference-in-differences in loss rates (% yr^{-1}), i.e., the effect sizes, extracted from the term $\beta_3 BA \times CI_{\text{it}}$ of Eq. 2. Estimates mean and 95% confidence intervals estimated by bootstrapping (black). Values below zero represent a reduction in deforestation relative to controls.

Variability of potential outcomes (e.g. baseline estimates)

We observed important variability in the range of potential outcomes (e.g. baseline estimates) produced by the robust matched sets, against which the effect of REDD+ was estimated. For the 40 matched sites, we observed a median standard deviation of 0.041 % yr^{-1} (IQR=0.017-0.093 % yr^{-1}) of baseline estimates among quality-assured matched sets for the same REDD+ project (Fig. 3.6) This corresponds to nearly half of the magnitude in mean reductions in deforestation by REDD+, estimated at 0.08 % yr^{-1}.

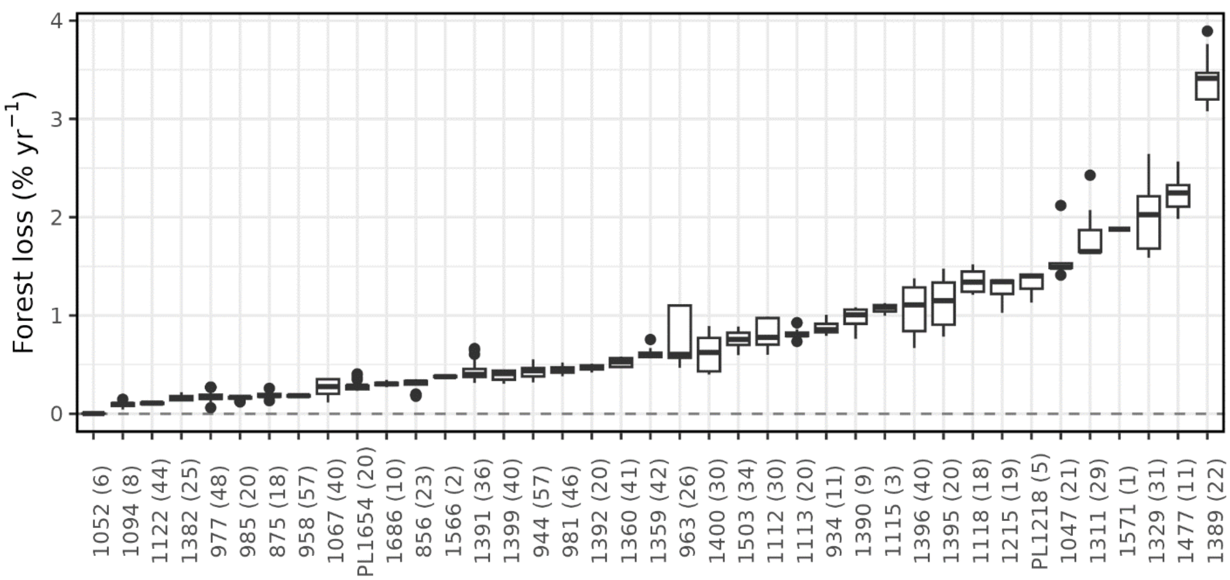

Figure 3.6: Range of baseline estimates for REDD+ sites. Boxplots show the distribution of site-level baseline estimates, defined as forest loss (% yr^{-1}) in control groups, in the period prior to project implementation. The number of robustly matched sets used to compute summaries shown in parenthesis.

Discussion

To our knowledge, this is the first attempt in trying to derive the impact of REDD+ using a multi-matching approach in which models were subjected to a battery of robustness tests, making explicit the potential range of baseline estimates that may come about from different matching choices. Matching was performed on 44 sites for which 1556 matching runs were selected to examine the effect of REDD+, involving 40 VCS sites. Our analysis shows significant but modest reductions in forest loss that are attributable to REDD+ when measured in area-based terms.

Our relative estimates of forest loss reductions, of 33%, aligns with the findings of a recent evaluation of a REDD+ national programme (Roopsind et al. 2019), and adds to the growing evidence base of the impacts of site-based REDD+ effectiveness, with studies ranging from positive (Simonet et al. 2019; Guizar-Coutiño et al. 2022) to little or no impact on reducing forest loss (West et al. 2020; Correa et al. 2020).

Our study highlights some of the challenges involved in counterfactual assessments and calls for the use of robust methodologies when developing REDD+ baselines. We attempted to reduce biases by applying multiple matching approaches and them by evaluating sets according to multiple validation criteria. Testing multiple matching combinations allowed us to explore the range of potential outcomes that may come about from valid matching set-ups, without making a-priori decisions favoring one approach versus another. While most of the sites showed effects of similar sizes once validated matches were selected, in a few sites we observed an important variability in the range of outcomes even when focusing on robust matched sets. For example, ten REDD+ sites in this study showed a positive impact when bootstrapping the effect sizes of the robustly matched sets, yet the standard deviations of their baseline estimates were of larger magnitudes than the bootstrapped effect sizes. For these sites, we could have arrived at a different conclusion about the impact of REDD+, had we settled for one of the robustly matched sets which suggested a null or negative impact. Moreover, we attempted to make explicit the presence of unmeasured confounders by incorporating sensitivity tests as part of the validation process, which despite their importance, are still rarely used in conservation impact evaluation (Jones et al. 2022). These tests are even more important in settings such as conservation, where a full characterisation of the factors influencing the selection of areas for project development is perhaps unfeasible.

We note an important difference between the aggregate estimates of REDD+ effectiveness between this chapter and that provided in Chapter 2. The examination of differences in effect sizes provided in Appendix B reveal implications for matching. In Chapter 2 we collected many more samples than in Chapter 3 since we used ~30m pixels as a unit of analysis. We matched with replacement, allowing control units to be paired with more than one treated unit, to ensure that as many treated samples as possible were matched. The analysis provided in

Appendix B point to potential caveats when using matching with replacement, particularly when this results in the selection of a few high-weight control units.

Importantly, our study highlights the non-negligible impact that the choice of matching parameters has on effect size estimation, where robustly matched sets for the same REDD+ area could lead to different effect size estimates depending on their parametrization. These findings contribute to ongoing discussions about making empirical research more transparent by examining the role of matching parameters in effect size estimation (Desbureaux 2021). We show that specific choices of algorithms and covariate combinations resulted in robustly matched set. Matching on relevant confounders is essential to isolate the causal effects of the treatment, it is therefore not surprising that the covariate combination that accounted for historical land-use resulted in more robustly matched sets than simpler covariate combinations. The fact that the Mahalanobis distance resulted in more matched robust sets aligns with recent studies looking at the impact of matching algorithms on matching quality (King & Nielsen 2019; Desbureaux 2021).

Our research advocates for the careful examination of evidence when deriving inferences from observational data analyses. In situations where the range of potential outcomes depicted by sets of validated control groups is of a larger magnitude than the observed effects, additional steps might be required to confidently derive inferences about interventions. We attempted to derive a ‘true’ treatment effect size by averaging and bootstrapping on robust matched sets within the same REDD+ projects, but other methods to combine effect sizes, such as fitting meta-analyses models, are available and their application for uncovering true effect sizes from multiple, equally-valid, estimates should be explored. Further studies examining how different matching parameters impact effect sizes would be necessary to advance methodological discussions and best practices when applying matching in conservation settings.

CHAPTER 4: PREVAILING SPILLOVER EFFECTS IN REDD+ PROJECTS: IMPLICATIONS FOR FORESTS AND CARBON

ABSTRACT

Conserving tropical forests has been identified as a cost-effective way of reducing GHG emissions while meeting climate and biodiversity targets. However, programmes that preserve tropical forests within a project area may displace deforestation activities beyond project boundaries, generating leakage effects that reduce overall effectiveness. While development protocols for REDD+ projects contemplate leakage, robust quantification of post-treatment leakage effects occurring in the vicinity of programmes have hardly ever been conducted. Here we combined estimates of forest area loss and carbon loss to examine the effectiveness of 38 REDD+ interventions, including a 15-km leakage zone, at reducing emissions. The projects reduced forest loss rates by 0.037% yr^{-1} (95% CI= 0.068 to 0.006% yr^{-1}) and reduced C loss by -0.03 MgC ha^{-1} yr^{-1} (95% CI = -0.06 to -0.01 MgC ha^{-1} yr^{-1}) relative to counterfactuals, within five years of project implementation. Although aggregate estimates of forest C declines reveal higher net levels of biomass loss in the combined REDD+ and leakage belt areas relative to controls, proportional changes suggest more biomass loss took place in control groups, that is, the amount of forest C lost relative to baseline stocks was larger in the combined control groups than that observed in the combined REDD+ and leakage belt areas. In the leakage belts, deforestation rates were greater in forests located farther away from REDD+ boundaries and was moderated by proximity to cities. The results provide evidence of potential positive and negative leakage effects in the vicinity of REDD+ after project implementation, with rates of deforestation increasing or decreasing significantly in the vicinity of projects after project implementation.

INTRODUCTION

Conservation scientists are increasingly adopting robust evaluation methods to understand the effectiveness of area-based interventions for conserving natural habitats, such as creating protected reserves or providing incentives to conserve for environmental services (Ferraro & Hanauer 2014; Baylis et al. 2016). Such evaluations are often based on comparing impacts in treated areas relative to a counterfactual (i.e.; untreated but otherwise similar areas), from which the difference in outcomes can be established (Ferraro 2009; Börner et al. 2020). Yet, these interventions may directly or indirectly cause impacts that extend beyond their boundaries, which are not captured by such approaches. Accounting for displacements is crucial in determining the net effect of interventions as these can enhance or undermine impacts achieved within project boundaries.

The effects of displacements are not usually incorporated in forest conservation assessments (Pfaff & Robalino 2017; Fuller et al. 2019), in part due to the complexity involved in determining attribution59 and characterising channels through which these effects take place (Pfaff & Robalino 2017; Meyfroidt et al. 2020). In forest conservation, the term "leakage" is commonly used to describe situations where the outcome of interest is negatively impacted in locations outside project boundaries (Lima et al. 2019). Conversely, "positive leakage" has been sometimes used for positive impacts beyond intervention zones, such as when a decline of forest loss rates outside intervention areas takes place (Herrera et al. 2019). The extent and scales at which leakage take place are determined by a host of interacting factors, ranging from project design, to policy and economic conditions that promote deforestation activities outside the intervention areas once restrictions become enforced (Atmadja & Verchot 2012; Pfaff & Robalino 2017; Meyfroidt et al. 2020).

Displacements of effects can occur at multiple scales. Deforestation displacements may take place to distant areas, even across countries, through adjustments in national and international supply chains (Meyfroidt & Lambin 2009; Delzeit et al. 2018). Displacements that occur in the vicinity of project boundaries, also referred as local leakage, has received much attention in the impact evaluation literature, including protected areas (Ewers & Rodrigues 2008; Robalino et al. 2017; Herrera et al. 2019; Ford et al. 2020), payment for ecosystem services interventions (Alix-Garcia et al. 2012; Velly et al. 2017), formalising land tenure initiatives (Wren-Lewis et al. 2020) and voluntary protection schemes (Nolte et al. 2019). These studies have found that local leakage effects vary by intervention type, with no consistent pattern of the expected magnitude, and is influenced by socio-economic factors such as population density, GDP and governance characteristics (Fuller et al. 2019). Moreover, local leakage patterns of opposing sign can occur heterogeneously across landscapes and are determined by socio-economical and governance factors (Robalino et al. 2017). For instance, in a payment for ecosystem services programme in Mexico, deforestation leakage varied by income and was moderated by access to credit instruments (Alix-Garcia et al. 2012). Similarly, accessibility by roads interacted with the locations of tourist centres in Costa Rican protected areas, describing patterns of positive and negative leakage effects, where forests closer to tourism hotspots that were accessible had reduced leakage rates (Robalino et al. 2017). Other examples of local leakage effects of opposing sign linked to protected areas have been described in Brazil, where federal protected areas located in high deforestation regions saw reduced deforestation rates outside the intervention zones, while indigenous lands in similar context saw increased deforestation in nearby areas (Herrera et al. 2019).

Accounting for leakage has been a sticking point in the development of a global Reducing Emissions from Deforestation and forest Degradation (REDD+) framework (Angelsen et al. 2012), and a challenging component in the implementation of site-level REDD+ interventions (Atmadja & Verchot 2012; Henders & Ostwald 2012). Countries developing their national REDD+ strategies are required to develop safeguards to tackle deforestation displacements within their national boundaries, for instance by strengthening public governance structures and enhancing monitoring, reporting and verification systems (Angelsen et al. 2012). Moreover, local REDD+ interventions are particular among site-based forest interventions in that considerations for leakage are included since the project development phase. This typically involve mapping existing drivers of deforestation from which a forecast of displacement risks is produced (Atmadja & Verchot 2012). Projects are required to address leakage by implementing mitigation strategies and monitor the identified areas where displacements might occur (Chagas et al. 2020). Methodologies such as that used by the Voluntary Carbon Standards (VCS) require a ‘leakage zone’ to be defined outside the intervention zone, which is used as a baseline to establish *ex-post* presence of leakage (Shoch et al. 2011). It is expected that leakage could undermine REDD+ benefits but understanding of the extent to which leakage occurs following REDD+ interventions remain limited.

In this chapter, we aim to understand the effectiveness of VCS REDD+ sites at avoiding forest loss and biomass loss within project areas, with a particular focus on local spillovers effects, and their implications for protecting forest carbon. We achieved this by examining post-project deforestation patterns in REDD+ project areas and within 15 km of their boundaries. We used a counterfactual approach that relied on matching forest plots in REDD+ and in the adjacent areas with forests that share similar characteristics and exposure to drivers of deforestation, which help to establish an estimate about the expected forest and biomass loss trajectories had the projects not been implemented. To examine local leakage patterns more closely, we also explored variations of deforestation pattens outside project boundaries given their distance to REDD+ and localities. Finally, we assessed the presence of key drivers of deforestation as mapped by Curtis *et al*. 2018, inside REDD+ and within 15 km of their boundaries, with which we assessed the feasibility of displacing deforestation activities to neighboring lands after the introduction of REDD+.

METHODS

RESEARCH DESIGN

Mapping of VCS REDD+ sites. Data on VCS REDD+ interventions were collected following the steps outlined in (Guizar-Coutiño et al. 2022) to compile the initial database. We focused on Verified Carbon Standard (VCS) projects given their prominence in the voluntary carbon markets (Donofrio et al. 2019). We gathered project design documents, validation reports and geospatial datasets depicting project area boundaries from the VCS registry (https://registry.verra.org/app/search/VCS). We focused on “Reducing Deforestation and Degradation” projects established in the tropics (Africa, South-East Asia, America and Oceania), of which 71 boundary files and documents were available. Individual layers were processed and assembled into a standardised database with normalised polygons (e.g.; removing overlaps), reprojecting features to Mollweide equal-area projection (Bingham et al. 2019). All data compilation and manipulation were performed in PostGIS (v 2.3).

Definition of leakage belts and sampling units. We devised a sampling strategy to collect information form areas within the REDD+ project boundary and neighbouring forests to assess REDD+ effectiveness and potential leakage effects. To define the zone within which to examine leakage effects, we computed a buffer ring that extended 15 km from the project boundaries (henceforth “leakage belt”) which were adjusted to exclude overlaps with protected areas, nearby REDD+ projects and any potential overlapping leakage belts. This approach is similar to VCS methodologies that require the definition of a surrounding belt which is monitored for increases in deforestation rates relative to defined baselines (although leakage zones are sometimes selected in non-neighbouring forests which might be impacted by the introduction of REDD+). We produced a grid of circular plots, each of 150 m radius (~7 ha) and 500 m spacing between the centroids, within the area encompassing REDD+ sites and leakage belts. To identify leakage samples from neighbouring REDD+ projects (e.g., where overlaps in the leakage belts existed, n=7) we assigned plots to the REDD+ project whose boundary was nearest. Sampling was performed so that the centroid of each circular plot overlapped a forest pixel (e.g. excluding areas with no forest cover) as defined by the Tropical Moist Forests (TMF) database, described in more detail in the next section (Vancutsem et al. 2021). We divided the sample plots into two sets of ‘treatment groups’ for each VCS site: one for REDD+ areas and another for observations in the leakage belt (Fig. 1.1). Finally, we collected a different set of samples from lands outside REDD+ and leakage belt areas that were used for potential controls. These consisted of circular plots of 150 m radius that were randomly dispersed across forested areas from the same biomes and countries as the REDD+ sites, with each of their centroids separated by at least 1 km. The separation between plots was determined to reduce spatial dependence between observations (Robalino & Pfaff 2012).

Annual deforestation maps. Annual estimates of forest cover and forest loss were taken from the TMF database (Vancutsem et al. 2021). This product characterises annual trajectories in forest disturbances and post-disturbances land cover types on a per-pixel basis (~30 m^2 resolution), from 1990 - 2019. Three major classes of forest stand are described in the database: the *undisturbed* forest class, representing closed evergreen or semi-evergreen forest areas); the *degraded* forest class, consisting of regrowing evergreen pixels after interruptions (which may be short-lived) in the undisturbed forest time-series; and the *deforested* class, representing a permanent removal of forest cover, from which other land-use classes ensue (Vancutsem et al. 2021). For each year during the 1990 - 2019 period, we estimated the extent of undisturbed and disturbed forest classes (e.g., deforestation and degradation) within each of the computed plots. We built a database of the changes in the proportion of forest cover within each plot, defined as: $r_t = \Delta p_t / p_d$, where Δp_t is the extent of deforested and degraded in year t and p_d is the area of undisturbed forest at the start of the time series (1990).

Characterising determinants of forest loss. We collected plot-level information on key determinants of forest loss to which we used in the matching procedure. We focused on determinants of deforestation (e.g. Busch & Ferretti-Gallon 2017) for which global layers were available or that could be computed with surrogate data. These included: elevation and slope (Jarvis et al. 2008), distance to the nearest urban centre (Weiss et al. 2018) and distance to forest edge (Ewers et al. 2011). Distance to forest edges was produced with the collected TMF time-series using a 2-step approach: 1) for each forest pixel within the circular plots, we computed the distance to the nearest pixel that had changed its status to deforested or degraded pixel during the observed year; and 2) we produced annual estimates of the mean distance to disturbance events at the plot level, considering pixel-level distance to disturbance events in the previous five years, spanning the period 2005-2019.

Accounting for protected areas. Additionality calculations are more credible if we account for the presence of other conservation efforts that may be active in the regions where REDD+ is operating (Schleicher et al. 2019a). We thus excluded all the areas of leakage belts and the wider landscape that overlapped with protected areas. To characterise protected areas, we downloaded the World Database of Protected Areas (WDPA) dated April 2019 (UNEP-WCMC & IUCN 2019). We cleaned this database by excluding areas defined as "Not designated" and "Inscribed", as well as UNESCO Biosphere Reserves (Bingham et al. 2019). We used this reduced version of the database to remove any intersections between the protected area geometries and REDD+ areas, their leakage belts, and any overlapping area from the wider landscape.

Baseline biomass estimates. We used the WCMC biomass layer (Soto-Navarro et al. 2020) to derive changes in forest carbon (C) after project implementation in REDD+ and leakage areas. Prior to deciding on this layer we examined the potential use of the latest gridded aboveground biomass density layer from the Global Ecosystem Dynamics Investigation (GEDI) project. We determined poor coverage over the REDD+ so it was deemed unfit for this study (Appendix C Table C.1). The WCMC biomass product was produced by combining estimates of above and below ground carbon stock in biomass and soils (Xia et al. 2014; Hengl et al. 2017; Bouvet et al. 2018), representing conditions in MgC Ha^{-1} for circa 2010 and provided at $300m^2$. Plot-level values of biomass were obtained by overlaying the matched samples plots with the biomass layer and estimating a mean biomass value for each plot (MgC ha^{-1}).

Annual biomass loss maps. The basis for assessing changes in biomass stocks due to REDD+ relied on combining annual rates of forest loss with plot-level estimates of biomass. Using the annual estimates of forest loss in each plot, we estimated biomass loss as $\text{biomassLoss}_{\text{it}} = (lossHa_{\text{it}} * meanBiomass_i)/7$, where the mean biomass value for a given plot (meanBiomass_i) was multiplied by the extent of forest cover lost in year t and divided by the size of the plot to provide a per-hectare estimate ($\text{lossH}a_{\text{it}}$; MgC Ha^{-1}). Importantly, these estimates are constrained to biomass loss, so secondary forest regrowth is not included in the calculations of annual biomass loss. This allowed us to compile a time-series of forest C loss for REDD+, leakage and control plots. This differs from VCS methodologies for producing baseline emissions scenarios, where published estimates of biomass stocks for the expected replacing land-cover (e.g. the land cover type that will likely follow after deforestation) is sufficient to derive an estimate of the expected emissions after conversion. To assess changes, we constructed a time series of biomass loss estimates for each matched plot.

ESTIMATING EFFECT SIZES

Matching. We used statistical matching to pair REDD+ and leakage forest plots with similar patches of forests outside REDD+ and leakage belt areas, establishing a counterfactual scenario of the expected trajectories in forest biomass loss had projects not been implemented. The matching approach we used followed the findings identified in Chapter 3, where a systematic approach to matching, involving trials were different algorithms, covariates and thresholds for selecting controls plots (i.e. calipers) were employed to match samples collected from REDD+ areas. From this exercise we identified a combination of parameters that resulted in the maximum number of robustly matched sets, which was used to construct counterfactual groups for this chapter. We used matching with replacement using the Mahalanobis distance metric (Legendre & Legendre 1988) to assess similarity between treated and control samples, assessed on the following covariates: 1) mean elevation, 2) mean slope, 3) mean distance to population centres, 4) mean distance to disturbances over the five years prior to project commencement, 5) deforestation rates in the five years prior to project commencement, 6) percent of undisturbed forest cover by the time of project commencement, and 8) country. Matching also included a caliper of 0.7 SD to enforce similarity of observations prior to performing the match.

To assess the quality of the matches we examined four quality parameters (discussed in detail in Chapter 3): 1) post-matching covariate balance between treated and selected controls, where matched sets showed an absolute standardised mean difference of < 0.25 across all covariates; 2) proportion of units matched, with at least 80% of samples matched; 3) similarity of deforestation rates between matched sets in the period prior to REDD+ (i.e."parallel trends"), examined with a linear mixed model to rule out significant differences (at the 5% level) in the temporal trends of deforestation between matched sets, prior to REDD+; and 4) sensitivity of effects to hidden confounders, where we tested matched sets for hidden confounding as strong as distance to disturbed forests in explaining the treatment effect.

A total of 38 REDD+ sites were successfully matched when assessed against the four-quality criterion. We proceeded to match the leakage belts of the 38 REDD+ sites referenced using the same matching parameterisation, which resulted in 27 out of the 38 leakage belts successfully matched: one leakage belt failed to compute a matched set, and 10 matches did not meet all the matching quality criteria established in specified in Chapter 3. There were multiple reasons as to why leakage belts were not successfully matched, including diverging trends in the rates of deforestation in the pretreatment period (n=7), imbalances in covariate distributions between leakage and control sets (n=2) and lack of enough leakage samples matched to generalise findings (n=1).

Site-level estimates of treatment effect for forest and biomass loss. We combined matching with regressions to estimate differences in deforestation and biomass loss in REDD+ and leakage belts after project implementation. The main estimates of deforestation rates differences were obtained with a linear fixed-effects model with a difference-in-differences (DD) specification to compute effect sizes across all the matched runs. Separate models were computed for REDD+ and leakage belt samples. To estimate changes in forest distance rates we fitted a model as:

$$\text{PropForestLoss}_{\text{it}} = \beta_0 + \beta_1 BA_t + \beta_2 CI_i + \beta_3 BA \times CI_{\text{it}} + \varepsilon \qquad (0.1)$$

Where $\text{PropForestLoss}_{\text{it}}$ is the proportional area of forest area loss in plot i at time t, and BA and CI are dummy indicators for period (before=0, after=1) and group (0=control, 1=treated). Since we had 2 treatment groups per VCS site (i.e.; the REDD+ project area and the leakage belt), we ran 2 separate analyses. A similar approach was used to estimate differences in biomass between treatment groups after REDD+, with a model specified as:

$$\text{BiomassLoss}_{\text{it}} = \beta_0 + \beta_1 BA_t + \beta_2 CI_i + \beta_3 BA \times CI_{\text{it}} + \varepsilon \qquad (0.2)$$

Where $\text{BiomassLoss}_{\text{it}}$ is the total biomass loss (MgG/Ha) i at time t, with the period and group dummy indicators. In both cases, the interacting term $\beta_3 BA \times CI_{\text{it}}$ provides the DD estimator. The analysis focused on measuring changes within five years of project implementation, compared to five years of pre-treatment observations. Model standard errors were clustered at the treatment-control pair level (Zeileis et al. 2020) to calculate 95% CIs.

Pooled effect size estimates we used a meta-analysis approach to derive a global effect size for the REDD+ and leakage observational groups, as well as a net effect combining REDD+ and leakage estimates. We fitted meta-analysis random effects models (Harrer et al. 2021) to pool effect sizes at three scales: a) net effect size for REDD+ interventions, b) net effect size for leakage belts, and c) a combined net effect of REDD+ and leakage belts. We used the standard errors from the treatment effects models to weight observations, such that effect sizes with higher precision (i.e. smaller standard errors) were given a greater weight. A Knapp-Hartung adjustments was used to calculate the confidence intervals.

Deforestation patterns within leakage belts. To understand whether deforestation patterns in the leakage belts could be described by determinants of deforestation and their distance to the REDD+ intervention, we fitted a linear mixed effects model using all matched sample plots in the leakage belts, defined as:

$$\text{LossRat}e_i = \beta_0 + \beta_1 Z_i + (1|s) + \varepsilon \qquad (0.3)$$

Including a random intercept for VCS site (1|s), where LossRate_i refers to the mean disturbance rate in the sample i in the five years period after REDD+ implementation, and $\beta_1 Z_i$ is a matrix of covariates that includes: distance to the closest REDD+ boundary, time travel to cities, mean distance to deforestation, elevation, and slope.

Analysis of drivers of deforestation. To assess the feasibility of deforestation displacing from REDD+ project areas to the buffer zones after project implementation, we examined the presence of key drivers of deforestation as characterised by Curtis *et al*. 2018. This dataset portrays likely causes of forest disturbance at a 10 km resolution since the year 2000, including causes such as commodity production, shifting agriculture, forestry activities, fires and urbanisation processes. We overlaid this dataset with REDD+ and buffer zones, and compared the proportional extent of each of these drivers in each of these zones (Fig, C.2). We visualised the proportional composition of drivers of forest loss in REDD+ and their leakage areas to assess the extent to which similar deforestation activities could take place across these landscapes. Presence of similar drivers of deforestation in the vicinity of REDD+ projects would allow us to infer the potential to displace deforesting activities based on the suitability to relocate activities in the neighboring areas.

RESULTS

Estimates of initial carbon (C) stocks.

We observed similar initial C stocks estimates across treated and control plots, for REDD+ and leakage belts plots respectively (Fig. 4.1), for the thirty-eight REDD+ sites that were successfully matched. Forests selected as controls for REDD+ had a median C density of 162 MgC ha^{-1} (inter-quartile range [IQR]= 138 - 189 MgC ha^{-1}), and where similar to C levels in REDD+ sites, estimated at 160 MgC ha^{-1} (IQR= 138 - 182 MgC ha^{-1}). Sample plots selected as controls for leakage belts had a median C concentration of 146 MgC ha^{-1} (IQR= 114 - 185 MgC ha^{-1}) and were similar to that observed form plots in leakage belts, estimated at 144 MgC ha^{-1} (IQR= 115 - 182 MgC ha^{-1}).

Total baseline carbon stocks, obtained by aggregating plot-level values, varied among observational groups due to differences in their total extent. We estimated total initial carbon stocks of 79697 GgC for the REDD+ area, 54289 GgC for the REDD+ control area, 99563 GgC for the leakage belt area and 78151 GgC in the leakage belt control area. Total area and C estimates for each observational group are provided in Appendix C Table C.2.

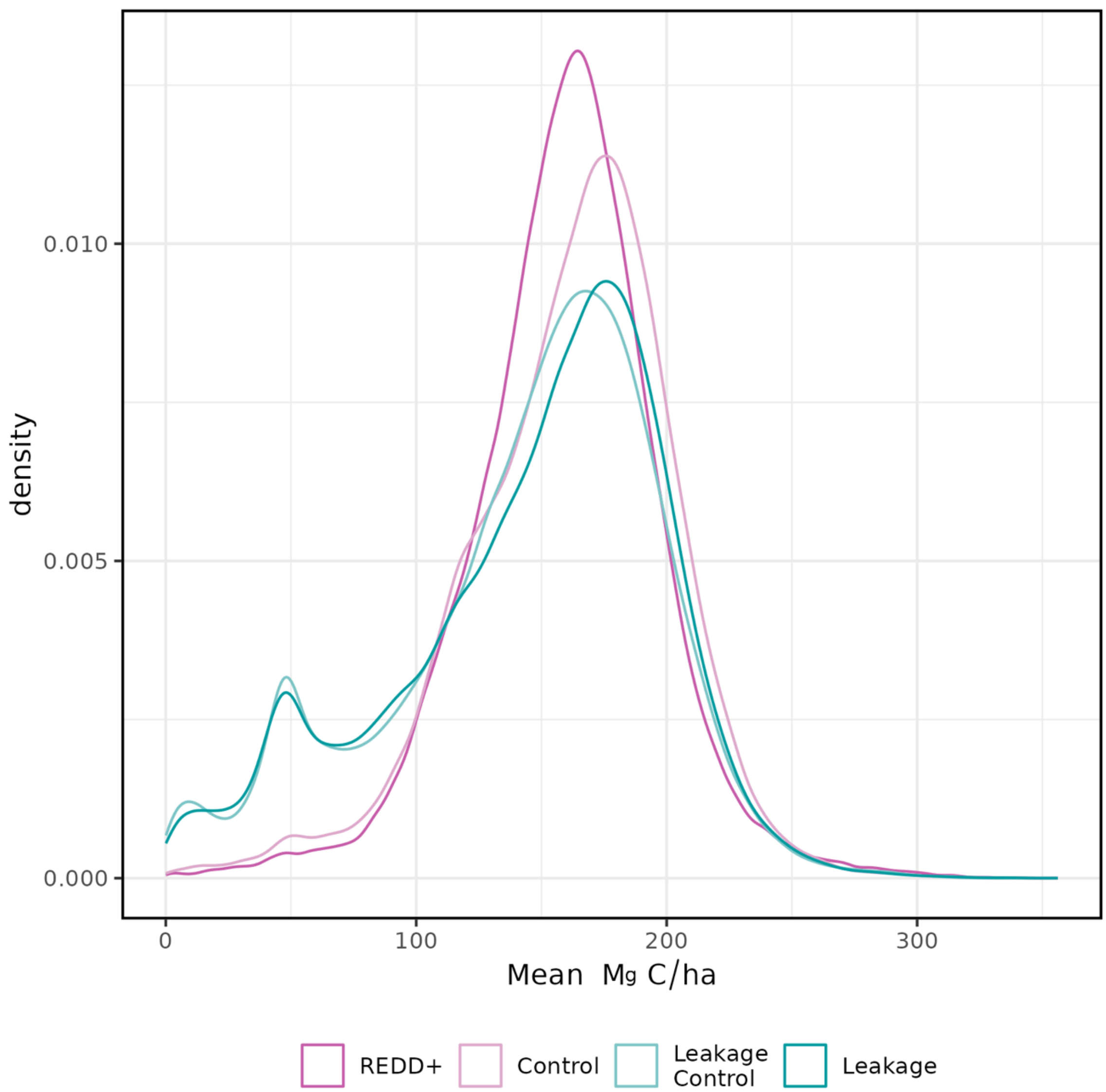

Figure 4.1: Probability distributions of plot-level mean biomass values (MgC ha^{-1}) for REDD+ and controls (pink), and leakage belts and their controls (turquoise).

Treatment effect on forest loss rates

Significant reductions in rates of forest loss were observed within REDD+ areas. The majority of REDD+ sites (n=22 of 38) reduced deforestation rates within their boundaries compared to controls in the 5 years period after project implementation, while remaining sites showed a null (n=12) or negative (n=4) effect in rates of forest loss during the same period. Combining effect sizes across projects we estimated a mean change in forest loss rates of -0.080% yr^{-1} (95% CI= -0.128 to -0.032% yr^{-1}). This effect size is different from the one estimated in Chapter 2, is it is based on the matching combination that yielded the largest number of valid matched sets (as opposed to combining effect sizes from matched sets).

Positive and negative leakage effects were observed in the 15-km buffer surrounding REDD+ sites. Almost half of leakage belts (n=17) showed reduced deforestation rates within their boundaries compared to controls after project implementation, while remaining leakage belts showed a null (n=6) or negative (n=14) impact on forest loss rates after projects were implemented. Changes in forest loss rates were close to zero when effect sizes were averaged using a meta-analysis model the average was -0.008% yr^{-1} which was not significantly different from zero (95% CI= -0.048 to 0.033% yr^{-1}).

Three distinctive patterns of effectiveness were apparent when effect sizes in REDD+ project areas were plotted against effect sizes in leakage belts (Fig. 4.2): a) overall positive effects (e.g. forest loss reductions in REDD+ and leakage belts, bottom-left quadrant), in 19 projects; b) positive effects in REDD+ but increased deforestation in the surrounding landscapes (e.g. the 'classical' leakage pattern, top-left quadrant), in 10 sites; and c) negative effects (e.g. increased deforestation in REDD+ and leakage belts, top-right quadrant) in 8 projects. Differences in the rates of deforestation between REDD+ sites and leakage belts were closely correlated among projects (Spearman's rho= 0.62, S=3192, p<0.0001). Combining effect sizes of treated and leakage belts, we estimated a small overall change in deforestation rates, of -0.037% yr^{-1} (95% CI= -0.068 to -0.006% yr^{-1}) by meta-analysis modelling. Individual effect sizes for REDD+, leakage belts and their combined effects is provided for reference in Appendix C Fig. 1.2.

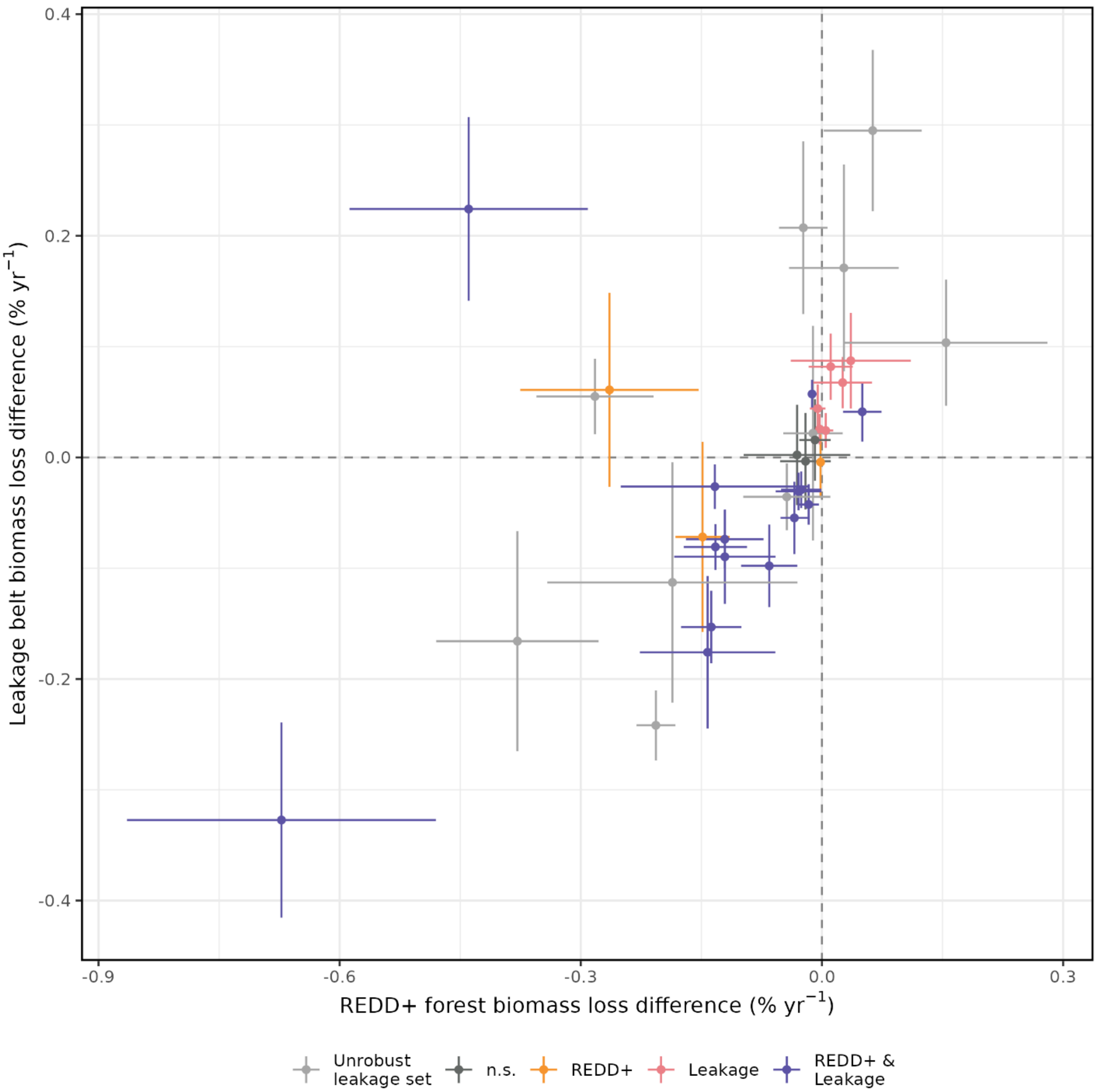

Figure 4.2: Differences in forest loss rates versus matched controls, in matched REDD+ project areas and leakage belts. Differences in annual rates of deforestation (% yr^{-1}) observed in 38 matched project sites (x-axis) and in their leakage belts (y-axis) five years after the implementation of REDD+. Quadrants resulting from intersecting the zero crossings on the x and y axis are shown to show the general patterns of effectiveness: **a**, an overall positive effect for REDD+ and leakage belts; **b**, positive effect in REDD+ but negative leakage effects, and; **c**, an overall negative effect in REDD+ and the surrounding landscapes. Note that no sites are located in the **d** quadrant. Colouring indicates statistical significance at the 5% threshold of comparing treatments (e.g. REDD+ or leakage belts) against their controls: REDD+ only (orange), leakage belt only (pink), REDD+ and leakage belts (purple) or none (grey). Results in light grey denote leakage matched sets that did not meet robust quality criteria. Error bars are for 95% CI where standard errors were clustered at the treatment-control pair level.

Treatment effects on forest biomass

Changes in the rates of forest loss due to REDD+ translated to similar patterns of avoided C losses in patches of forest under REDD+. The majority of REDD+ sites (n=20) reduced C losses in the 5 years period after project implementation, while remaining sites showed a null (n=13) or negative (n=4) trajectories in C loss (e.g. increased losses). Overall reductions in biomass loss within REDD+ sites was estimated at -0.058 MgC Ha^{-1} yr^{-1} (95% CI= -0.095 to -0.021 MgC Ha^{-1} yr^{-1}) when combining effect sizes with a meta-analysis model.

Changes in C loss after project implementation were both positive & negative across REDD+ leakage areas, converging at around zero when averaged. The majority of leakage belts (n=17) showed null changes in levels of C loss when compared to controls after project implementation. The remaining leakage belts showed significant reductions (n=11) or increases (n=9) in forest C loss during the same period. Combining estimates across with a meta-analysis model we observed non-significant changes in levels of C loss relative to controls, estimated at 0.001 MgC Ha^{-1} yr^{-1} (95% CI= -0.024 to 0.027 MgC Ha^{-1} yr^{-1}).

Similar patterns of effectiveness as observed in avoiding deforestation were revealed when comparing estimates of avoided biomass loss between REDD+ and leakage belts groups (Fig. 4.3): a) reductions in biomass loss was observed in 18 REDD+ and leakage belts; b) reductions in REDD+ but increased biomass loss in leakage belts was observed in 12 sites; c) increased biomass loss in REDD+ and leakage belts was found in 5 projects, and; d)increased biomass loss in REDD+ and reduced biomass loss in the leakage, in 2 sites, mostly driven by differences in baseline carbon levels, as forests in REDD+ areas for these two sites were more carbon dense than plots in leakage belts. As noted in the differences in deforestation rates, we a correlation in estimates of avoided C loss between REDD+ sites and leakage belts (Spearman's rho= 0.49, S=4312, p=0.002). Aggregating effect sizes across REDD+ sites and leakage belts suggest a small reduction in levels of forest C loss relative to controls in the first 5 years of REDD+ implementation, estimated at -0.03 MgC Ha^{-1} yr^{-1} (95% CI = -0.06 to -0.01 MgC Ha^{-1} yr^{-1}).

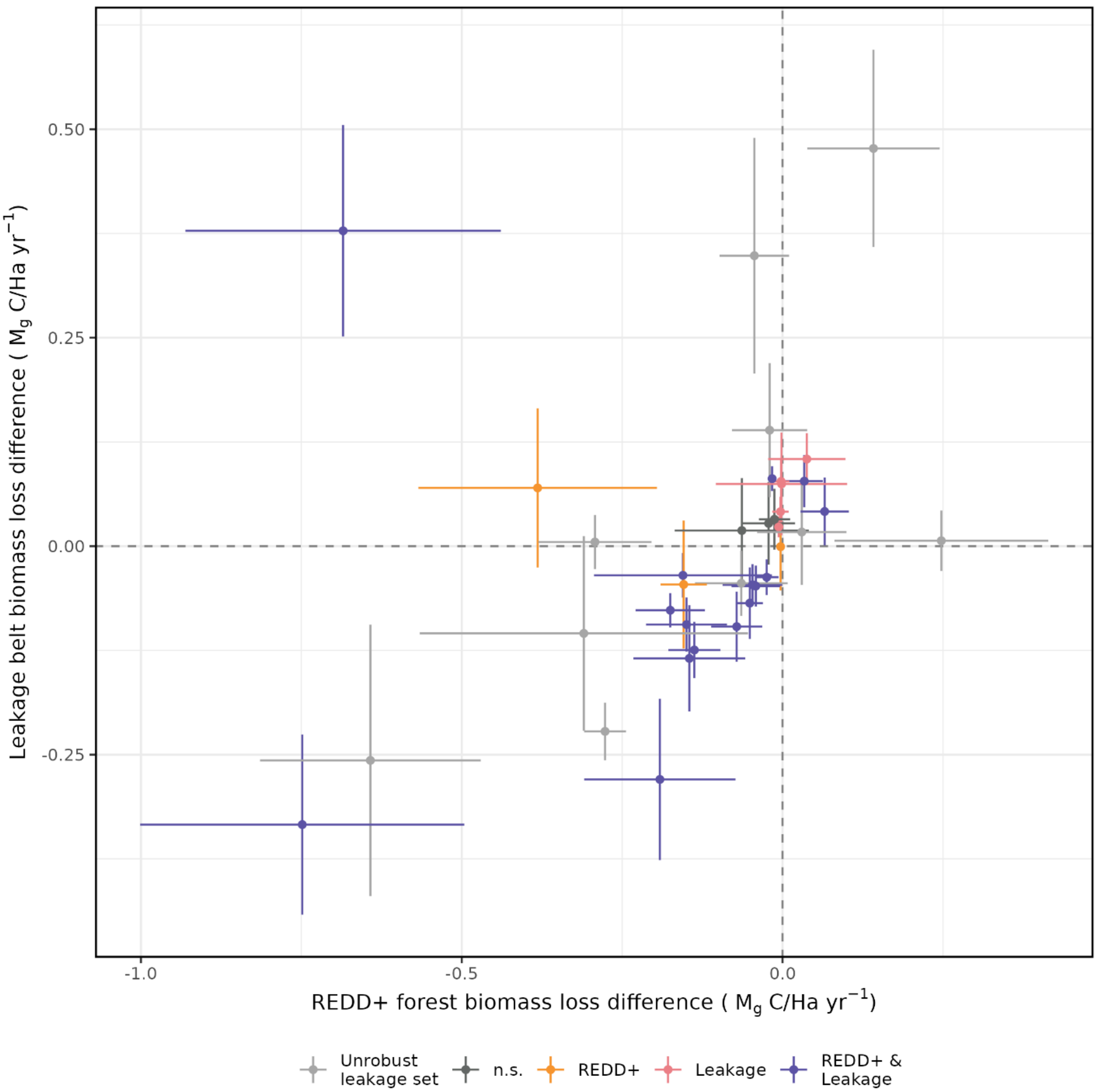

Figure 4.3: Differences in forest biomass loss versus matched controls, in matched REDD+ and leakage belts. Differences in annual C loss (MgC Ha^{-1} yr^{-1}) observed in 38 matched project sites (x-axis) and in their leakage belts (y-axis) five years after the implementation of REDD+. Quadrants resulting from intersecting the zero crossings on the x and y axis are shown to show the general patterns of effectiveness: **a**, an overall positive effect for REDD+ and leakage belts; **b**, positive effect in REDD+ but negative leakage effects; **c**, an overall negative effect in REDD+ and the surrounding landscapes, and; **d**, negative effects in REDD+ areas but positive in leakage belts. Colouring indicates group statistical significance at the 5% level: REDD+ only (orange), leakage belt only (pink), REDD+ and leakage belts (purple) or none (grey). Results in light grey denote leakage matched sets that did not met robust quality criterion. Error bars are for 95% CI where standard errors were clustered at the treatment-control pair level.

Estimates of C losses aggregated across REDD+ sites and leakage belts

Aggregating estimates of C loss across samples suggests net emission reductions in REDD+ sites and net emission increases in the 15 km spanning leakage belt areas. In the five years running up to REDD+ implementation, C losses in REDD+ sites were of a larger magnitude than C losses in controls, totaling 558 GgC and 356 GgC respectively, corresponding to losses of 0.70% and 0.66% of initial C stocks, estimated at 79697 GgC and 54289 GgC for REDD+ and controls respectively. Biomass loss in REDD+ sites increased at a smaller rate following project implementation than C losses observed in controls, totaling 1314 GgC and 1313 GgC, for REDD+ and controls respectively. This corresponded to a loss of 1.65% and 2.42% of initial C stocks, for REDD+ and controls respectively. Difference-in-differences changes in C loss amounted to a reduction of 201 GgC loss in REDD+ areas, or -0.82% of REDD+ C stocks losses compared to the percent of C stocks lost in controls (Fig. 4.4, **a**).

Leakage belts saw increased C losses relative to controls following REDD+ implementation. In the 5 years prior to REDD+ implementation, C losses totaled 2793 GgC and 2047 GgC, for leakage belts and controls respectively, corresponding to losses of 2.81% and 2.62% relative to initial C stocks, estimated at 99563 and 78151 GgC for leakage belts and controls respectively. Five years after project implementation, biomass losses were of a larger magnitude in leakage belts compared to controls, totaling 5367 and 3677 GgC, respectively, which corresponds to losses of 5.39% and 4.71% of initial C stocks, in leakage belts and controls respectively. Difference-in-differences changes in C losses amounted to an additional loss of 944.98 GgC in leakage belts following REDD+ implementation, or 0.50% more loss of C stocks in leakage belts compared to the percent of C stocks lost in controls (Fig. 4.4, **b**).

Aggregate estimates of forest C declines reveal higher net levels of C loss in the combined REDD+ and leakage belt areas, yet proportional losses of the initial C stocks were higher in controls. Total losses before and after project implementation, in the combined REDD+ and leakage belts areas, amounted to 3351 and 6682 GgC, respectively, while C losses in the combined control zones in the periods before and after project implementation totaled 2403 and 4991 GgC, respectively. Difference-in-differences changes in C loss amounted to an additional 743 GgC lost in the combined REDD+ and leakage belts areas after project implementation. However, expressing these changes as a percentage of the initial C stocks in each group, we see 0.10 % less emissions of C stocks in the combined REDD+ and leakage belts compared to the percent of C stocks lost in controls (Fig. 4.4, **c**).

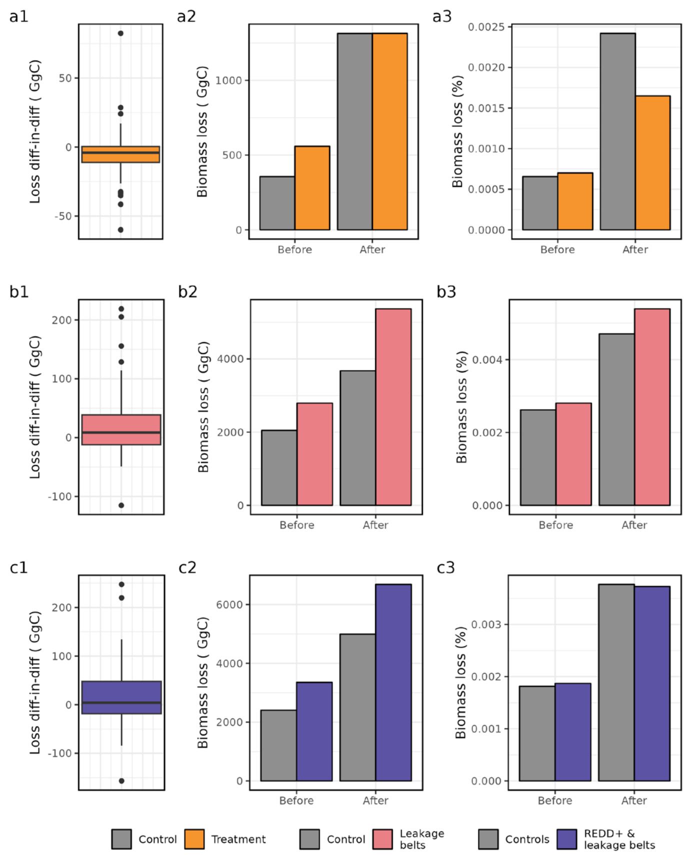

Figure 4.4: Total carbon loss, before and after REDD+ interventions, within 38 project sites. Rows show biomass loss estimates in REDD+ areas (orange), leakage belts (pink) and the sum of these zones (purple), contrasting these with controls (grey). Boxplots show the distribution of site-level difference-in-differences of avoided biomass loss, i.e.; the effect of REDD+ in avoiding biomass loss (left column); biomass loss estimates provided as the total (middle column) and as percentages of the initial C stocks in each group (%, right column).

Heterogeneity of deforestation patterns due to distance to cities and distance to REDD+ boundaries

Patterns of forest loss within leakage belts were determined by distance to population centres (mean= -0.003; t=-53; $p < 0.001$), where plots located further away from localities experienced less deforestation than average in the 5-years period after REDD+ implementation.

Deforestation patterns were also determined by the distance to REDD+ project boundaries (mean= 0.0002, t=2.9, $p < 0.01$; Appendix C Table C.2), suggesting that the rates of forest loss in the post-implementation period were higher for plots located further away from the REDD+ project (Fig. 4.5).

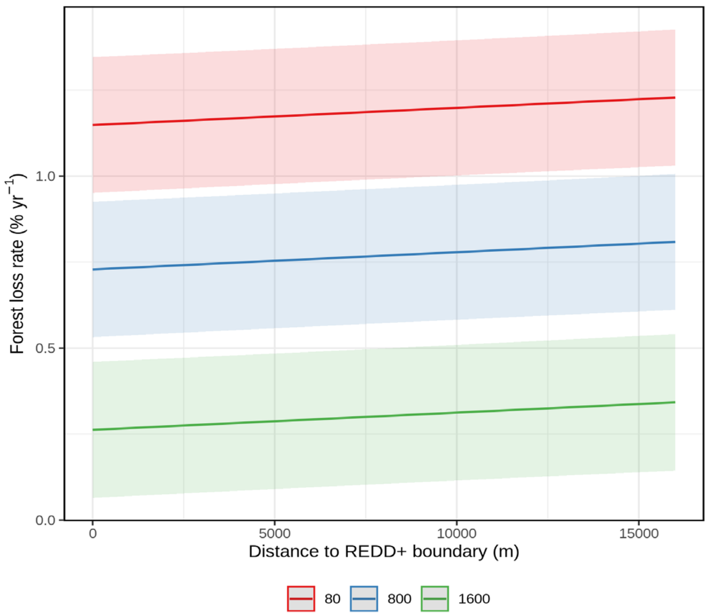

Figure 4.5: Impacts of distance from REDD+ project boundary and travel time to population centres on forest loss rates in leakage belts. Three travel times to the closest population centres are predicted: ~2 minutes (red), 15 minutes (blue) and 25 minutes (green). The median travel time was 10 minutes from the closest locality (IQR= 5 - 20 mins).

Analysis of drivers of deforestation

We observe a similar portfolio of drivers of deforestation in REDD+ and their surrounding areas. Shifting agriculture and commodity production were identified as predominant drivers of deforestation in some of the examined REDD+ projects and their surrounding areas, while other projects were predominantly free of drivers within and outside REDD+. For the remaining projects, we observe a combination of drivers occurring within and outside REDD+. Proportional occurrence of drivers of deforestation were closely correlated between REDD+ and buffer areas (Spearman's $\rho = 0.79$, $p < 0.0001$).

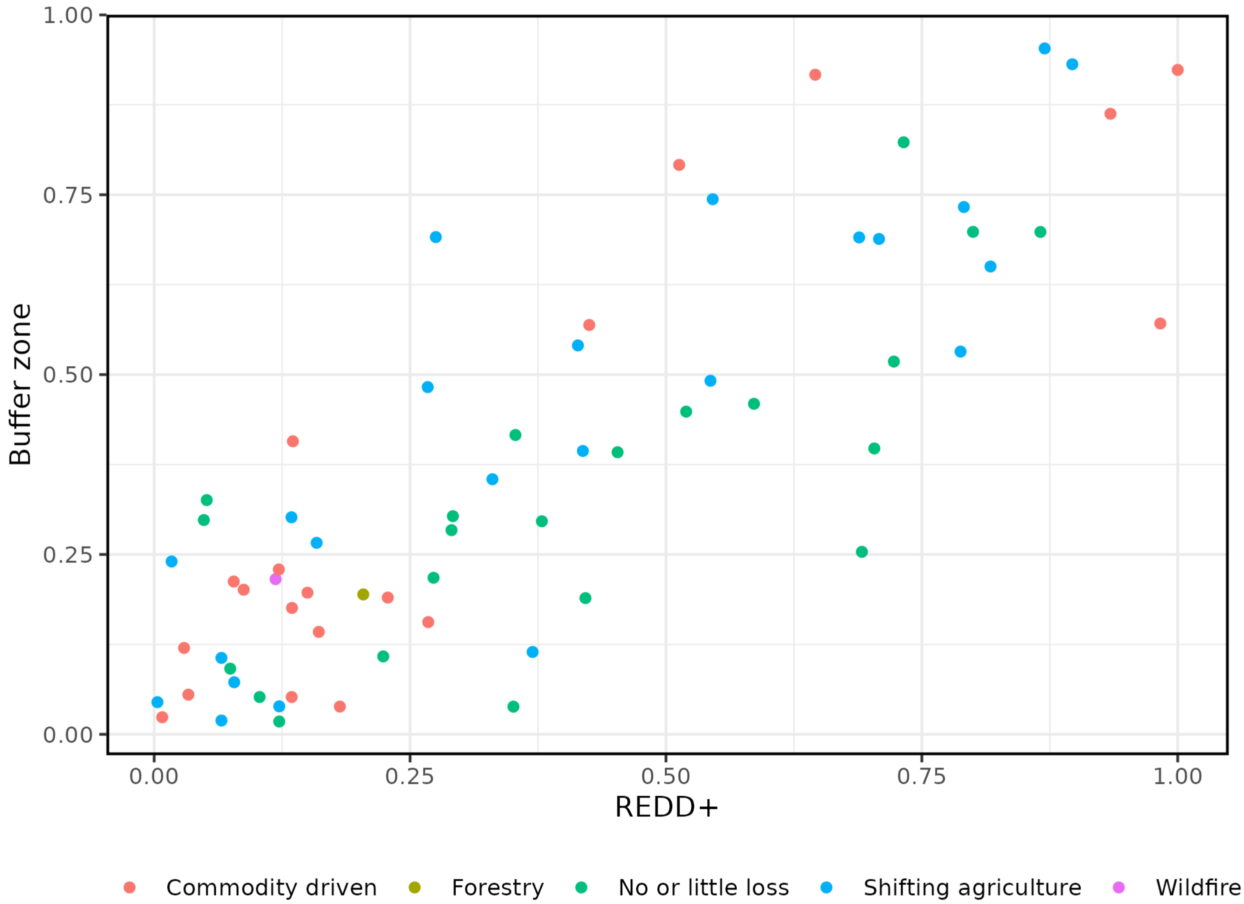

Figure 4.6. Drivers of forest loss inside REDD+ and buffer belts. Proportional coverage of up to 5 drivers of forest loss as characterised by Curtis et al. 2018, inside the examined REDD+ areas (x-axis) and their buffer belts (y-axis)

Discussion

To the best of our knowledge, this is the first global study of REDD+ effectiveness that explicitly integrates leakage into the total effect size calculations. Except for the few sites where the 'classical leakage' was evident (i.e.; deforestation decreased in REDD+ sites but increased in the surrounding area), analyses of post-treatment changes in REDD+ and proximate areas suggests a diffusion of the effects observed within REDD+ project areas into the surrounding landscapes: large forest loss reductions in REDD+ areas were accompanied by large forest loss reductions in leakage belts, while REDD+ sites with a null or negative effect in REDD+ corresponded with leakage belts where rates of deforestation remained the same or increased (Fig. 4.2). Analyses of deforestation patterns in the leakage belts add weight to this idea, showing that rates of forest loss increased with distance from REDD+ project boundaries (Fig. 4.5). These findings suggest that REDD+ activities shielded proximate forest patches outside intervention areas. However, failures in REDD+ design (Milne et al. 2019) and the inability of local-scale interventions to cope with market-driven pressures (Skutsch & Turnhout 2020) may have rendered REDD+ inadequate to prevent forest loss within the project areas and surround landscape in many sites, picked up in Fig. 4.5 as additional deforestation in both REDD+ sites and leakage belts.

Attributing leakage is always difficult to measure due to the complex, often unmeasurable, ways in which interventions impact forests elsewhere. In this study we determined the presence of leakage by identifying statistical differences in disturbance processes in areas exposed to the same drivers of deforestation, after project implementation. While the quantification of leakage would be better characterised by incorporating local-level information about the likely mechanisms that would trigger leakage effects (Pfaff & Robalino 2017), our results show strong positive and negative changes in forest loss in the vicinity of REDD+ areas. Much attention in the literature has been devoted to discussing potential negative leakage effects, however the potential of positive spillovers should be more looked at and examined.

Protecting forests can trigger market adjustments, displacing deforestation activities to areas outside the intervention, particularly if commercial activities are affected by restrictions. These market adjustments, also known as market leakage, can come about by a combination of factors, including an inelastic demand for commodities affected by the intervention (Meyfroidt et al. 2020), lack of technologies to enhance production levels to meet current demand (Meyfroidt et al. 2018), and the availability of mobile labour and capital to relocate production activities (Wunder 2008). Relocation of activities can also take place across countries through adjustments in the international supply chain (Meyfroidt & Lambin 2009; Delzeit et al. 2018). Methods for understanding pathways linking forest interventions and deforestation displacements are still emerging (Pfaff & Robalino 2017), in particular to ascertain long-distance displacements (Hertel et al. 2019; Meyfroidt et al. 2020); but despite these limitations, current research on distant effects of conservation policies and programmes suggests that market leakage can, in some instances, substantially undermine impacts achieved within forest reserves (Meyfroidt & Lambin 2009; Miranda et al. 2019) Moves to implement REDD+ as a jurisdictional approach could help in reducing country-level leakage through institutional coordination and enforcement of national REDD+ policies, but these efforts alone might not be sufficient to mitigate domestic deforestation displacements without addressing demand-side and market pressures (Streck 2021).

Important caveats need to be considered when examining these results. First, the leakage estimates for 10 sites come from matched sets that did not meet all the quality standards as the matched sets for the REDD+ projects criteria, and these should be interpreted with caution. Second, we used a cookie-cutter approach of leakage belt definition, computing 15 km around project interventions, however we cannot assume that leakage effects were prevalent for the entire leakage belt area, particularly in projects with relatively small size. This is supported by the result showing deforestation rates in leakage belts increased with distance from project area boundary. Third, anticipating regional or international market leakage effects due to restrictions such as those imposed by REDD+ is inherently difficult as this is dependent on socio-economic factors (e.g. environmental or development policies), demand-side responses to restrictions, and the capacity of the market to make adjustments along the supply chain (Meyfroidt et al. 2020). Regional or cross-boundary leakage could have taken place due to these interventions but are difficult to quantify. Nevertheless, while we are not able to ascertain the precise mechanism by which REDD+ may have produced leakage effects, the presence of a similar portfolio of drivers of deforestation in REDD+ and buffer zones (Fig. 4.6) can be indicative of the potential for the displacement of such activities to take place, particularly in areas were high prevalence of shifting agriculture and commodity production was identified and which experienced a surge in deforestation rates after REDD+ implementation.

Another set of caveats relate to how we measured biomass change, which are restricted to biomass losses, and does not consider fluctuations in emissions due to subsequent regrowth. Such variations can take place rapidly, particularly in the early re-greening phase (Rappaport et al. 2018). Furthermore, our analyses did not included changes in biomass due to degradation events, which can result in important releases of carbon even if canopy cover is not fully removed (Mitchard 2018). Moreover, quantification of biomass would be more precise with the use of new remote sensing products such as that produced by the Global Ecosystem Dynamics Investigation (GEDI) mission and Biomass ESA's forest mission. In the case of GEDI, the patchiness of the data prevented its use in the present study.

By incorporating counterfactual approaches not only into the estimation of effect sizes but also into the potential displacement of deforestation or threats that may occur in the vicinity of intervention areas, we were able to gain insights about the net effect of REDD+ within the targeted areas and surrounding landscapes. The consideration of potential displacements to proximate lands due to the presence of similar drivers of deforestation provides an avenue to anticipate how and where leakage might take place. By identifying potential areas of displacement, appropriate measures can be implemented to minimize or offset the negative impacts. With anthropogenic emissions from land use change estimated at +5.9±4.1 $GtCO_2eq/yr$ between 2010-2019, of which ~45% is attributable to deforestation (Shukla et al. 2019), emission reductions in the examined REDD+ projects represent only a tiny fraction of global emissions. This research suggest small progress is underway, but important challenges for project implementation and scale-up need to be tackled for REDD+ to realise its potential as a cost-effective initiative for mitigating climate change (Duchelle et al. 2018a; Milne et al. 2019; TSVCM 2021).

Chapter 5: Discussion

Overview

The fate of tropical forests in the coming decades will define whether we can successfully address socio-environmental challenges posed by the climate and biodiversity crises. Conserving tropical forests can be a cost-effective bridge towards meeting decarbonisation targets set out for this decade while delivering on social and environmental benefits (Griscom et al. 2017; Seddon et al. 2020). On the other hand, allowing deforestation to continue unabated will only exacerbate the current state of affairs, magnifying climate, biodiversity and sustainability challenges (Lewis et al. 2015; Baccini et al. 2017; Barlow et al. 2018; Symes et al. 2018).

The role of tropical forests in climate change mitigation is gaining momentum in multilateral fora and the re-emerging carbon markets. At the 26th Conference of the Parties of the United Nations Framework Convention on Climate Change (COP26), over a 100 countries committed to halting deforestation and forest degradation by 2030 through the Glasgow Leaders' Declaration on Forests and Land Use, which came attached with monetary pledges of almost USD 20 billions to support conservation activities (UNFCCC 2021). Furthermore, increased demand for forest-based carbon offsets is expected from the voluntary carbon markets, following a trend that has been observed in recent years (Donofrio et al. 2021), and is projected to grow exponentially by the end of this decade (TSVCM 2021).

Reducing Emissions from Deforestation and Forest Degradation (REDD+) encapsulate various approaches for supporting activities to reduce forest loss and enhance carbon stocks in tropical countries, with landscape REDD+ projects representing one of the earliest forms of on-the-ground interventions under the REDD+ branding. It has been almost 15 years since the initial wave of projects became implemented, and rigorous evaluations of these interventions are starting to yield insights into the multiple dimensions of REDD+ performance, most prominently on the livelihoods impacts (Duchelle et al. 2018; Sunderlin et al. 2018b, 2018a). Empirical impact evaluations looking at the ability of local-scale interventions to conserve forests are just starting to emerge. At the time when I commenced my PhD, in 2018, only one robust impact evaluation of forest carbon projects was available (Jayachandran et al. 2017). Since then, a few more empirical studies appeared in the literature, showing a mixed picture about the effectiveness of these initiatives to mitigate climate change by avoiding forest loss, from projects achieving reductions in deforestation (Simonet et al. 2019; Guizar-Coutiño et al. 2022), to null or negative results (West et al. 2020; Correa et al. 2020).

Summary of findings

In the preceding chapters of this thesis, I developed three pieces of research with the aim of exploring the impact of voluntary carbon projects on conserving tropical forests, exploring the contribution of these efforts in avoiding carbon emissions, as well as uncovering some of the challenges involved in quantifying impact from forest conservation interventions.

In **Chapter 2** we conducted a global analysis of REDD+ interventions, involving 40 voluntary REDD+ projects in 9 countries. To our knowledge this is the first study in which remotely sensed degradation and deforestation data were used to test voluntary REDD+ projects effectiveness across a sample of geographically dispersed projects that differ in deforestation drivers and social objectives. We showed average reductions in deforestation and forest degradation over the first 5 years of project operation, with reductions being particularly greater in regions where the threat of deforestation was greatest. We demonstrated modest reductions terms of the absolute area of spared forests, however in relative terms both rates were roughly halved by the projects we investigated. In this chapter we also found consistent reductions in deforestation and forest degradation when controlling for the presence of protected areas in the examined landscapes, suggesting that the observed gains in forest loss reductions were additional to that obtained by protected areas.

In **Chapter 3** we derived the impact of REDD+ using a multi-matching approach in which models were subjected to a battery of robustness tests, making explicit the potential impact that different matching parameters may have on effect size estimates. Matching was performed on 44 sites for which 1556 matching runs were selected to examine the effect of REDD+, involving 40 VCS sites. In line with the findings discussed in Chapter 2, our analysis showed significant reductions in forest loss due to REDD+, albeit of a lesser magnitude than the previous chapter. This difference could be partly explained by the different sampling techniques: Chapter 2 uses a pixel-based matching approach, with pixels of ~30 m in size, whereas Chapters 3 and 4 are based on area sampling, with plots of ~7 ha used to characterise landscapes. The examination of differences in effect sizes between Chapter 2 and Chapter 3 provided in Appendix B reveal implications for matching. In Chapter 2 we collected many more samples than in Chapter 3 since we used ~30m pixels as a unit of analysis. We matched with replacement, allowing control units to be paired with more than one treated unit, to ensure that as many treated samples as possible were matched. As is shown in the discussions in Appendix B, the use of matching with replacement, and in particular the selection of a few high-weight control units might have impacted our effect size estimations. Importantly, our study highlights the non-negligible impact that the choice of matching parameters has on effect size estimation, where robustly matched sets for the same REDD+ area could lead to different effect size estimates depending on their parametrisation. Future research on the impacts that different matching choices in addition to those examined in this chapter, such as the scale and size of units of analyses, would help advance best practices in the field.

In **Chapter 4** we built from the robust matching approaches developed in Chapter 3 to quantify forest carbon changes due to REDD+, in which local leakage was estimated using observational methods, and combined with effect sizes observed in REDD+ projects to derive a *net* total effect size. Except for the few sites where the ‘classical leakage’ was evident (i.e.; deforestation decreased in REDD+ sites but increased in the surrounding area), we observed a diffused effect of REDD+ into the surrounding landscapes: large forest loss reductions in REDD+ areas were accompanied by large forest loss reductions in leakage belts, while REDD+ sites with a null or negative effect in REDD+ corresponded with leakage belts where rates of deforestation remained the same or increased. Moreover, we took an empirical approach when assessing changes in forest carbon due to REDD+, by comparing plots of forests exposed to similar drivers of deforestation which contained similar biomass estimates. This approach differs from simpler baseline comparisons used in carbon offsetting projects (e.g. Shoch et al. 2011)Emission reductions in the examined REDD+ projects represent a tiny fraction of global emissions, and could be well below the estimates of avoided emissions as forecasted by carbon offsets projects such as those registered in the VCS. This suggests on the one hand that REDD+ on the ground is achieving carbon benefits, but there is a need to scale up activities to unlock the potential of forest carbon interventions for mitigating climate change. On the other hand, methodologies to estimate the carbon additionality for existing projects might have to be re-examined, considering recent methodological developments in observational designs.

Future directions

Challenges and future research in observational designs

Counterfactual analyses were a cornerstone in the development of this thesis, and the way in which these methodologies were used to estimate the impact of REDD+ illustrate some of the challenges and future research that is needed to enhance the robustness of observational designs for conservation impact evaluation.

The strengths and weaknesses of the different tools for counterfactual development need to be further explored. Different methods can be used to derive impacts of forest interventions when temporal observations are available for treatment and control groups, before and after project implementation, such as combining matching with difference-in-differences (Costedoat et al. 2015; Santika et al. 2021) or by using synthetic control methods to ensure similar pretreatment deforestation rates between treatment and control groups (West et al. 2020). We used two typical matching approaches in this thesis: a pixel-based approach in Chapter 2, in which pixels representing standing forests at the time of project implementation were selected to conduct the analyses (and in doing so, emulating VCS requirements for project implementation) (Shoch et al. 2011), followed by an after-intervention analyses of effect sizes; and a plot-based approach in Chapters 3 and 4, where the rates of deforestation in selected controls followed a similar trend than treated plots, followed by a difference-in-differences analysis to estimate effect sizes. A further approach not used in this thesis is the synthetic control method (Abadie et al. 2010), in which a ‘synthetic’ control group is constructed using a weighted combination of untreated geometric units in such a way that deforestation rates of the counterfactual group follows the same trajectory as the treated group before project implementation (e.g. West et al. 2020). All these different approaches will likely produce different effect size estimates. It is therefore necessary to explore the relative merits and drawbacks of these methods so they can be better tailored to the question at hand.

Another set of challenges relate to how parameters selected to perform matching may be influencing effect sizes. In Chapter 3, we attempted to reduce biases by applying multiple matching approaches and validated the resulting matched sets using multiple quality criteria. Testing multiple matching combinations allowed us to explore the range of potential outcomes that may come about from valid matching settings, or parameter choices, without making a-priori decisions favouring one choice of parameter versus another. We observed however an important variability in the range of counterfactual scenarios within robust matched sets. Further studies examining how the different choices of matching parameters impact effect sizes would be necessary to advance methodological discussions and best practices in counterfactual design.

The impact of hidden confounders is rarely explored in conservation impact evaluation, but the use of sensitivity tests is necessary to ascertain the robustness of effect size estimates. In these chapters we attempted to match REDD+ observations with potential controls based on appropriate drivers of deforestation, but there will inevitably be unobserved covariates. A key challenge in conservation settings remains on accounting for the uncertainty of how interventions are assigned in the first place (Jones et al. 2022). For example, local REDD+ projects have often been implemented in areas where NGOs have been operating for some time (Sunderlin & Sills 2012), and are likely associated with certain land tenure conditions (Wunder 2013). As these confounders remain unobserved because they are difficult to quantify, the need to use sensitivity analyses to assess the potential influence of hidden bias remains all the more pertinent in conservation evaluations (Jones et al. 2022).

The challenges above highlight the need to standardise counterfactual methodologies for impact evaluation, particularly those used to develop emissions baselines to evaluate the effectiveness of REDD+ (West et al. 2020; Guizar-Coutiño et al. 2022). For example, it is currently not possible to establish the aggregate impact of VCS REDD+ projects because the various methodologies used to forecast emissions reductions are incomparable and produce different baseline scenarios (Wilebore 2015). Furthermore, it remains difficult to reproduce baselines estimates for carbon offset projects given that the materials needed to replicate the methods are commonly not accessible. Making empirical research more transparent should include practices of code and data sharing to be able to replicate results, which will allow to examine uncertainty factors discussed above, such as the impact of subjective modelling choices on effect size estimations.

Leakage

Protecting forests can trigger market adjustments, displacing deforestation activities to areas outside the intervention, particularly if commercial activities are affected by restrictions. These market adjustments, also known as market leakage, can come about by a combination of factors, including an inelastic demand for commodities affected by the intervention (Meyfroidt et al. 2020), lack of technologies to enhance production levels to meet current demand (Meyfroidt et al. 2018), and the availability of mobile labour and capital to relocate production activities (Wunder 2008). Relocation of activities can also take place across countries through adjustments in the international supply chain (Meyfroidt & Lambin 2009; Delzeit et al. 2018). Methods for understanding pathways linking forest interventions and deforestation displacements are still emerging (Pfaff & Robalino 2017), in particular to ascertain long-distance displacements (Hertel et al. 2019; Meyfroidt et al. 2020); but despite these limitations, current research on distant effects of conservation policies and programmes suggests that market leakage can, in some instances, substantially undermine impacts achieved within forest reserves (Meyfroidt & Lambin 2009; Miranda et al. 2019). Moves to implement REDD+ as a jurisdictional approach could help in reducing country-level leakage through institutional coordination and enforcement of national REDD+ policies, but these efforts alone might not be sufficient to mitigate domestic deforestation displacements without addressing demand-side and market pressures (Streck 2021).

Local leakage can be a commonplace feature of REDD+ interventions. Attributing local leakage is always difficult due to the complex, often unmeasurable, ways in which interventions impact forests elsewhere. I attempted to measure local leakage in Chapters 2 and 4: in the latter case, I used an empirical approach by matching forested areas in the vicinity of REDD+ projects to similar but unprotected forests in the landscapes, and determined the presence of leakage by identifying statistical differences in deforestation rates compared to otherwise similar forestlands. While this analysis might not be able to uncover the mechanisms that triggered such effects in the vicinity of REDD+ projects, which are expected to vary depending on the local contexts (Pfaff & Robalino 2017), the results show strong positive and negative changes in forest loss in the vicinity of REDD+ areas. Much attention in the literature has been devoted to discussing potential negative leakage effects, where displacements undermine gains achieved within the intervention areas. Chapter 4 gives empirical evidence that local leakage can also take place as a diffused version of the effects observed within REDD+ sites (and perhaps that of other forest initiatives). By incorporating counterfactual approaches not only into the estimation of effect sizes but also into the potential displacement of deforestation or threats that may occur in the vicinity of intervention areas, we were able to gain insights about the net effect of REDD+ within the targeted areas and surrounding landscapes. The consideration of potential displacements to proximate lands due to the presence of similar drivers of deforestation provides an avenue to anticipate how and where leakage might take place. By identifying potential areas of displacement, appropriate measures can be implemented to minimize or offset the negative impacts. This is an area that should explored further to enhance our understanding of how forest-based conservation initiatives work.

MONITORING TROPICAL FOREST

Advances in remote sensing tools will help reduce uncertainties in the estimation of forest carbon. The Global Ecosystem Dynamics Investigation (GEDI) LiDAR instrument has been added to the constellation of active sensors in space and provides consistent measurements of forest structure across the global tropics. This sensor is helping to produce the first high resolution, globally consistent map of forest carbon stock for the tropics. Likewise, the European Space Agency (ESA) Sentinel (Berger et al. 2012) and BIOMASS missions (Le Toan et al. 2011) will support innovative ways to generate high resolution global biomass estimates. Developing improved methods for quantifying forest degradation based on open-access imagery will allow to generate better estimates of intervention performance for conserving carbon, increasing our ability to track forest degradation will help to guide monitoring and restoration efforts.

While remote sensing products play a huge role in upscaling project development, the role that local communities in the monitoring and quantification of forest biomass should not be underplayed. There is an important opportunity to increase REDD+ success by engaging communities with in-situ measurements, which can result in enhanced biomass estimations (Butt et al. 2015). Moreover, promoting community monitoring can help bring local perspectives into the planning and decision-making of the interventions (Danielsen et al. 2022), which could lead to a more equitable and long-lasting impact of projects (Butt et al. 2015; Duchelle et al. 2018).

Concluding remarks

Site-based REDD+ projects conform the first generation of REDD+ efforts and are representative of a broader set of nature-based solutions to climate change, from which valuable lessons can be harnessed (Duchelle et al. 2019). More than a decade after implementation reveals aspects that need to be carefully revisited: from exposing the inadequacy to face external threats (Milne et al. 2019; Skutsch & Turnhout 2020), to challenges on meeting its commitment to delivering social impacts, as well as challenges related to establishing credible baseline and demonstrable emission reductions.

Applying the lessons from the last few decades to deliver effective and, crucially, equitable reductions in tropical forest degradation and deforestation will be critical if the climate change objectives are to be met. It is essential to learn from the last decade of REDD+ implementation. The REDD+ projects have already provided invaluable lessons on the central role of land tenure in implementation (Angelsen et al. 2017), the necessity of ensuring that the rural poor do not bear the cost of forest conservation efforts (Duchelle et al. 2018; Poudyal et al. 2018; Skutsch & Turnhout 2020), and the need for effective benefit-sharing systems and appropriate participation in decision-making and governance (Luttrell et al. 2013; Milne et al. 2019).

Protecting and restoring natural forests could be a nature-based climate solution that is cost-effective and could have substantial impact if the many hurdles to implementation can be overcome (Duchelle et al. 2018; Milne et al. 2019; TSVCM 2021). Moreover, avoiding deforestation and forest degradation is paramount for safeguarding biodiversity and can play a role in safeguarding noncarbon ecosystem services, such as water regulation and soil productivity (Griscom et al. 2017). Our results provide some room for optimism. Despite the many challenges to just and economically sustainable implementation, the initial wave of REDD+ projects were effective at reducing forest loss.

REFERENCES

Abadie A, Diamond A, Hainmueller J. 2010. Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. Journal of the American Statistical Association **105**:493–505.

Adams WM, Aveling R, Brockington D, Dickson B, Elliott J, Hutton J, Roe D, Vira B, Wolmer W. 2004. Biodiversity conservation and the eradication of poverty. Science (New York, N.Y.) **306**:1146–1149.

Agrawal A, Nepstad D, Chhatre A. 2011. Reducing Emissions from Deforestation and Forest Degradation. Annual Review of Environment and Resources **36**:373–396. Annual Reviews.

Alix-Garcia JM, Shapiro EN, Sims KRE. 2012. Forest Conservation and Slippage: Evidence from Mexico's National Payments for Ecosystem Services Program. Land Economics **88**:613–638. University of Wisconsin Press.

Andam KS, Ferraro PJ, Pfaff A, Sanchez-Azofeifa GA, Robalino JA. 2008. Measuring the effectiveness of protected area networks in reducing deforestation. Proceedings of the National Academy of Sciences of the United States of America **105**:16089–16094.

Angelsen A. 2017. REDD+ as result-based aid: General lessons and bilateral agreements of Norway. Review of Development Economics **21**:237–264. Wiley Online Library.

Angelsen A, Brockhaus M, Sunderlin WD, Verchot LV. 2012. Analysing REDD+: Challenges and choices. CIFOR, Bogor, Indonesia.

Angelsen A, Kaimowitz D. 2001. Agricultural Technologies and Tropical Deforestation. CABI.

Angelsen A, Martius C, De Sy V, Duchelle AE, Larson AM, Thuy PT, editors. 2018. Transforming REDD+: Lessons and new directions. Center for International Forestry Research (CIFOR), Bogor, Indonesia.

Angrist JD, Pischke J-S. 2008. Mostly harmless econometrics: An empiricist's companion. Princeton University Press.

Arriagada RA, Ferraro PJ, Sills EO, Pattanayak SK, Cordero-Sancho S. 2012. Do Payments for Environmental Services Affect Forest Cover? A Farm-Level Evaluation from Costa Rica. Land Economics **88**:382–399. University of Wisconsin Press.

Asner GP et al. 2013. High-fidelity national carbon mapping for resource management and REDD+. Carbon Balance Manag. **8**:7.

Asner GP, Mascaro J. 2014. Mapping tropical forest carbon: Calibrating plot estimates to a simple LiDAR metric. Remote Sensing of Environment **140**:614–624.

Atmadja S, Verchot L. 2012. A review of the state of research, policies and strategies in addressing leakage from reducing emissions from deforestation and forest degradation (REDD+). Mitig Adapt Strateg Glob Change **17**:311–336. Available from https://doi.org/10.1007/s11027-011-9328-4 (accessed February 1, 2021).

Austin PC. 2011. Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. Pharmaceutical Statistics **10**:150–161. Available from https://onlinelibrary.wiley.com/doi/abs/10.1002/pst.433 (accessed August 20, 2019).

Avelino AFT, Baylis K, Honey-Rosés J. 2016. Goldilocks and the Raster Grid: Selecting Scale when Evaluating Conservation Programs. PLOS ONE **11**:e0167945. Public Library of Science. Available from https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0167945 (accessed December 10, 2020).

Avitabile V et al. 2016. An integrated pan-tropical biomass map using multiple reference datasets **22**:1406–1420.
Baccini A et al. 2012. Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps **2**:182–185. Nature Research.
Baccini A, Walker W, Carvalho L, Farina M, Sulla-Menashe D, Houghton RA. 2017. Tropical forests are a net carbon source based on aboveground measurements of gain and loss. Science (New York, N.Y.) **358**:230–234.
Badgley G, Freeman J, Hamman JJ, Haya B, Trugman AT, Anderegg WRL, Cullenward D. 2022. Systematic over-crediting in California's forest carbon offsets program. Global Change Biology **28**:1433–1445.
Balmford A et al. 2023. Credit credibility threatens forests. Science 380.6644 (2023): 466-467.
Barlow J et al. 2018. The future of hyperdiverse tropical ecosystems. Nature **559**:517–526.
Baylis K, Honey-Rosés J, Börner J, Corbera E, Ezzine-de-Blas D, Ferraro PJ, Lapeyre R, Persson UM, Pfaff A, Wunder S. 2016. Mainstreaming Impact Evaluation in Nature Conservation. Conservation Letters **9**:58–64.
Berger M, Moreno J, Johannessen JA, Levelt PF, Hanssen RF. 2012. ESA's sentinel missions in support of Earth system science. Remote Sensing of Environment **120**:84–90.
Bille Larsen P et al. 2021. Understanding and responding to the environmental human rights defenders crisis: The case for conservation action. Conservation Letters **14**:e12777.
Bingham HC et al. 2019. Sixty years of tracking conservation progress using the World Database on Protected Areas. Nature Ecology & Evolution **3**:737–743. Nature Publishing Group.
Börner J, Schulz D, Wunder S, Pfaff A. 2020. The effectiveness of forest conservation policies and programs **12**:45–64. Annual Reviews.
Bottazzi P, Wiik E, Crespo D, Jones JPG. 2018. Payment for Environmental "Self-Service": Exploring the Links Between Farmers' Motivation and Additionality in a Conservation Incentive Programme in the Bolivian Andes. Ecological Economics **150**:11–23.
Bouvet A, Mermoz S, Le Toan T, Villard L, Mathieu R, Naidoo L, Asner GP. 2018. An above-ground biomass map of African savannahs and woodlands at 25m resolution derived from ALOS PALSAR. Remote Sensing of Environment **206**:156–173. Available from https://www.sciencedirect.com/science/article/pii/S0034425717306053 (accessed December 30, 2022).
Brando PM, Paolucci L, Ummenhofer CC, Ordway EM, Hartmann H, Cattau ME, Rattis L, Medjibe V, Coe MT, Balch J. 2019. Droughts, Wildfires, and Forest Carbon Cycling: A Pantropical Synthesis. Annual Review of Earth and Planetary Sciences **47**:555–581.
Busch J, Ferretti-Gallon K. 2017. What Drives Deforestation and What Stops It? A Meta-Analysis. Review of environmental economics and policy **11**:3–23.
Butt N, Epps K, Overman H, Iwamura T, Fragoso JMV. 2015. Assessing carbon stocks using indigenous peoples' field measurements in Amazonian Guyana. Forest Ecology and Management **338**:191–199.
Carmenta R, Coomes DA, DeClerck FAJ, Hart AK, Harvey CA, Milder J, Reed J, Vira B, Estrada-Carmona N. 2020. Characterizing and Evaluating Integrated Landscape Initiatives. One earth (Cambridge, Mass.) **2**:174–187.
Carranza T, Balmford A, Kapos V, Manica A. 2013. Protected area effectiveness in reducing conversion in a rapidly vanishing ecosystem: The Brazilian Cerrado. Conservation Letters **7**:216–223. Wiley Online Library.
Chagas T, Galt H, Streck C. 2020. A close look at the quality of REDD+ carbon credits. Climate Focus.

Cinelli C, Hazlett C. 2020. Making sense of sensitivity: Extending omitted variable bias. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **82**:39–67. Available from https://onlinelibrary.wiley.com/doi/abs/10.1111/rssb.12348 (accessed June 24, 2022).

Corbera E, Estrada M, Brown K. 2010. Reducing greenhouse gas emissions from deforestation and forest degradation in developing countries: Revisiting the assumptions. Climatic Change **100**:355–388.

Correa J, Cisneros E, Börner J, Pfaff A, Costa M, Rajão R. 2020. Evaluating REDD+ at subnational level: Amazon fund impacts in Alta Floresta, Brazil. Forest Policy and Economics **116**:102178.

Costedoat S, Corbera E, Ezzine-de-Blas D, Honey-Rosés J, Baylis K, Castillo-Santiago MA. 2015. How Effective Are Biodiversity Conservation Payments in Mexico? PLOS ONE **10**:e0119881. Available from https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0119881 (accessed January 24, 2019).

Curtis PG, Slay CM, Harris NL, Tyukavina A, Hansen MC. 2018. Classifying drivers of global forest loss. Science (New York, N.Y.) **361**:1108–1111.

Danielsen F, Eicken H, Funder M, Johnson N, Lee O, Theilade I, Argyriou D, Burgess ND. 2022. Community monitoring of natural resource systems and the environment. Annual Review of Environment and Resources **47**:637–670. Annual Reviews.

Dargie GC, Lewis SL, Lawson IT, Mitchard ETA, Page SE, Bocko YE, Ifo SA. 2017. Age, extent and carbon storage of the central Congo Basin peatland complex. Nature **542**:86–90. Nature Publishing Group. Available from https://www.nature.com/articles/nature21048 (accessed February 1, 2021).

Dawson N et al. 2021. The role of Indigenous peoples and local communities in effective and equitable conservation. Ecology and Society **26**. The Resilience Alliance. Available from https://www.ecologyandsociety.org/vol26/iss3/art19/ (accessed December 19, 2022).

De Sy V, Herold M, Achard F, Avitabile V, Baccini A, Carter S, Clevers JG, Lindquist E, Pereira M, Verchot L. 2019. Tropical deforestation drivers and associated carbon emission factors derived from remote sensing data. Environmental Research Letters **14**:094022. IOP Publishing.

DeFries R, Hansen A, Newton AC, Hansen MC. 2005. Increasing isolation of protected areas in tropical forests over the past twenty years. Ecological Applications **15**:19–26.

Delabre I, Boyd E, Brockhaus M, Carton W, Krause T, Newell P, Wong GY, Zelli F. 2020/ed. Unearthing the myths of global sustainable forest governance. Global Sustainability **3**:e16. Cambridge University Press. Available from https://www.cambridge.org/core/journals/global-sustainability/article/unearthing-the-myths-of-global-sustainable-forest-governance/661FE54EF21F34BD75CD874BB28B6B6F (accessed February 5, 2021).

Delzeit R, Klepper G, Zabel F, Mauser W. 2018. Global Economic–Biophysical assessment of midterm scenarios for agricultural Markets—Biofuel policies, dietary patterns, cropland expansion, and productivity growth. Environmental Research Letters **13**:025003. IOP Publishing. Available from https://doi.org/10.1088/1748-9326/aa9da2 (accessed January 6, 2022).

Desbureaux S. 2021. Subjective modeling choices and the robustness of impact evaluations in conservation science. Conservation Biology **35**:1615–1626. Available from https://onlinelibrary.wiley.com/doi/abs/10.1111/cobi.13728 (accessed January 6, 2022).

Dinerstein E et al. 2017. An Ecoregion-Based Approach to Protecting Half the Terrestrial Realm. Bioscience **67**:534–545. Available from https://academic.oup.com/bioscience/article-lookup/doi/10.1093/biosci/bix014.

Donofrio S, Maguire P, Merry W, Zwick S. 2019. Financing Emissions Reductions for the Future: State of the Voluntary Carbon Markets 2019. Forest Trends' Ecosystem Marketplace, Washington, DC.

Donofrio S, Maguire P, Meyers K, Daley C, Lin K. 2021. Markets in motion: State of the voluntary carbon markets 2021. Ecosystem Marketplace, Washington, DC.

Duchelle AE, de Sassi C, Sills E, Wunder S. 2018a. People and communities: Well-being impacts of REDD+ on the ground. Pages 131–141 Transforming REDD+: Lessons and new directions. CIFOR.

Duchelle AE, Seymour F, Brockhaus M, Angelsen A, Larson A, Moira M, Wong GY, Pham TT, Martius C, others. 2019. Forest-based climate mitigation: Lessons from REDD+ implementation. World Resources Institute. Available from https://www.wri.org/research/forest-based-climate-mitigation-lessons-redd-implementation.

Duchelle AE, Simonet G, Sunderlin WD, Wunder S. 2018b. What is REDD+ achieving on the ground? Current Opinion in Environmental Sustainability **32**:134–140.

Duncanson L et al. 2021. Aboveground woody biomass product validation good practices protocol. Page Good Practices for Satellite Derived Land Product Validation. Land Product Validation Subgroup (WGCV/CEOS). Available from https://pure.iiasa.ac.at/id/eprint/17135/1/CEOS_WGCV_LPV_Biomass_Protocol_2021_V1.0.pdf .

Eklund J, Blanchet FG, Nyman J, Rocha R, Virtanen T, Cabeza M. 2016. Contrasting spatial and temporal trends of protected area effectiveness in mitigating deforestation in Madagascar. Biological Conservation **203**:290–297.

Eklund J, Coad L, Geldmann J, Cabeza M. 2019. What constitutes a useful measure of protected area effectiveness? A case study of management inputs and protected area impacts in Madagascar. Conservation Science and Practice **1**:e107.

Ewers RM et al. 2011. A large-scale forest fragmentation experiment: The Stability of Altered Forest Ecosystems Project. Philosophical Transactions of the Royal Society B: Biological Sciences **366**:3292–3302. Royal Society. Available from https://royalsocietypublishing.org/doi/10.1098/rstb.2011.0049 (accessed December 22, 2021).

Ewers RM, Rodrigues ASL. 2008. Estimates of reserve effectiveness are confounded by leakage. Trends in Ecology & Evolution **23**:113–116. Available from http://www.sciencedirect.com/science/article/pii/S0169534708000402 (accessed April 1, 2019).

Ferraro PJ. 2009. Counterfactual thinking and impact evaluation in environmental policy. New Directions for Evaluation **2009**:75–84.

Ferraro PJ. 2017. Are payments for ecosystem services benefiting ecosystems and people? Page Effective Conservation Science: Data Not Dogma. Oxford University Press. Available from https://oxford.universitypressscholarship.com/view/10.1093/oso/9780198808978.001.0001/oso-9780198808978-chapter-25.

Ferraro PJ, Hanauer MM. 2014. Advances in Measuring the Environmental and Social Impacts of Environmental Programs. Annual Review of Environment and Resources **39**:495–517.

Ferraro PJ, Pattanayak SK. 2006. Money for nothing? A call for empirical evaluation of biodiversity conservation investments. Plos Biology **4**:e105.

Ferraro PJ, Simorangkir R. 2020. Conditional cash transfers to alleviate poverty also reduced deforestation in Indonesia. Science Advances **6**:eaaz1298. American Association for the Advancement of Science.

Ford SA, Jepsen MR, Kingston N, Lewis E, Brooks TM, MacSharry B, Mertz O. 2020. Deforestation leakage undermines conservation value of tropical and subtropical forest protected areas. Global Ecology and Biogeography **n/a**. Available from https://onlinelibrary.wiley.com/doi/abs/10.1111/geb.13172 (accessed August 20, 2020).

Fuller C, Ondei S, Brook BW, Buettel JC. 2019a. First, do no harm: A systematic review of deforestation spillovers from protected areas. Global Ecology and Conservation **18**:e00591. Available from http://www.sciencedirect.com/science/article/pii/S2351989419301143 (accessed June 28, 2019).

Fuller C, Ondei S, Brook BW, Buettel JC. 2019b. First, do no harm: A systematic review of deforestation spillovers from protected areas. Global Ecology and Conservation **18**:e00591.
Gatti LV et al. 2021. Amazonia as a carbon source linked to deforestation and climate change. Nature **595**:388–393. Nature Publishing Group.
Geist HJ, Lambin EF. 2002. Proximate Causes and Underlying Driving Forces of Tropical Deforestation: Tropical forests are disappearing as the result of many pressures, both local and regional, acting in various combinations in different geographical locations. BioScience **52**:143–150. Available from https://doi.org/10.1641/0006-3568(2002)052[0143:PCAUDF]2.0.CO;2 (accessed May 24, 2021).
Geldmann J, Barnes M, Coad L, Craigie ID, Hockings M, Burgess ND. 2013. Effectiveness of terrestrial protected areas in reducing habitat loss and population declines. Biological Conservation **161**:230–238.
Geldmann J, Manica A, Burgess ND, Coad L, Balmford A. 2019. A global-level assessment of the effectiveness of protected areas at resisting anthropogenic pressures. Proceedings of the National Academy of Sciences of the United States of America **116**:23209–23215.
Gibson L et al. 2011. Primary forests are irreplaceable for sustaining tropical biodiversity. Nature **478**:378–381. Nature Publishing Group.
Goetz SJ, Hansen M, Houghton RA, Walker W, Laporte N, Busch J. 2015. Measurement and monitoring needs, capabilities and potential for addressing reduced emissions from deforestation and forest degradation under REDD+. Environmental Research Letters **10**:123001. IOP Publishing.
Griscom BW et al. 2017. Natural climate solutions. Proceedings of the National Academy of Sciences **114**:11645–11650.
Guizar-Coutiño A, Jones JPG, Balmford A, Carmenta R, Coomes DA. 2022. A global evaluation of the effectiveness of voluntary REDD+ projects at reducing deforestation and degradation in the moist tropics. Conservation Biology **n/a**:e13970.
Hansen AJ et al. 2020a. A policy-driven framework for conserving the best of Earth's remaining moist tropical forests. Nature Ecology & Evolution **4**:1377–1384. Nature Publishing Group.
Hansen MC et al. 2013. High-resolution global maps of 21st-Century forest cover change. Science (New York, N.Y.) **342**:850–853.
Hansen MC, Krylov A, Tyukavina A, Potapov PV, Turubanova S, Zutta B, Ifo S, Margono B, Stolle F, Moore R. 2016. Humid tropical forest disturbance alerts using Landsat data. Environmental Research Letters **11**:034008. IOP Publishing.
Hansen MC, Wang L, Song X-P, Tyukavina A, Turubanova S, Potapov PV, Stehman SV. 2020b. The fate of tropical forest fragments. Science Advances **6**:eaax8574. American Association for the Advancement of Science.
Harrer M, Cuijpers P, A FT, Ebert DD. 2021. Doing meta-analysis with R: A hands-on guide, 1st edition. Chapman & Hall/CRC Press, Boca Raton, FL and London.
Harris NL, Brown S, Hagen SC, Saatchi SS, Petrova S, Salas W, Hansen MC, Potapov PV, Lotsch A. 2012. Baseline map of carbon emissions from deforestation in tropical regions. Science (New York, N.Y.) **336**:1573–1576.
Henders S, Ostwald M. 2012. Forest Carbon Leakage Quantification Methods and Their Suitability for Assessing Leakage in REDD. Forests **3**:33–58. Molecular Diversity Preservation International. Available from https://www.mdpi.com/1999-4907/3/1/33 (accessed February 1, 2022).
Hengl T et al. 2017. SoilGrids250m: Global gridded soil information based on machine learning. PLOS ONE **12**:e0169748. Public Library of Science. Available from https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0169748 (accessed December 30, 2022).

Herrera D, Pfaff A, Robalino J. 2019. Impacts of protected areas vary with the level of government: Comparing avoided deforestation across agencies in the Brazilian Amazon. Proceedings of the National Academy of Sciences of the United States of America **116**:14916–14925. National Academy of Sciences. Available from https://www.pnas.org/content/116/30/14916 (accessed October 9, 2020).

Hertel TW, West TAP, Börner J, Villoria NB. 2019. A review of global-local-global linkages in economic land-use/cover change models. Environmental Research Letters **14**:053003. IOP Publishing. Available from https://doi.org/10.1088/1748-9326/ab0d33 (accessed December 18, 2021).

Ho D, Imai K, King G, Stuart EA. 2011. MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. Journal of Statistical Software **42**:1–28. Available from https://www.jstatsoft.org/index.php/jss/article/view/v042i08 (accessed March 4, 2019).

Ho DE, Imai K, King G, Stuart EA. 2007. Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. Political Analysis **15**:199–236. Oxford University Press. Available from https://academic.oup.com/pan/article-abstract/15/3/199/1484592.

Holloway V, Giandomenico E. 2009. The History of REDD Policy. Carbon Planet.

Houghton RA, Byers B, Nassikas AA. 2015. A role for tropical forests in stabilizing atmospheric CO2 **5**:1022. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved. Available from http://dx.doi.org/10.1038/nclimate2869.

Hubau W et al. 2020. Asynchronous carbon sink saturation in African and Amazonian tropical forests. Nature **579**:80–87. Nature Publishing Group.

Iacus SM, King G, Porro G. 2019. A Theory of Statistical Inference for Matching Methods in Causal Research. Political Analysis **27**:46–68. Cambridge University Press.

Jack BK, Jayachandran S. 2018. Self-selection into payments for ecosystem services programs. Proceedings of the National Academy of Sciences of the United States of AmericaDOI: 10.1073/pnas.1802868115. National Acad Sciences. Available from http://dx.doi.org/10.1073/pnas.1802868115.

Jarvis A, Reuter HI, Nelson A, Guevara E. 2008. Hole-filled SRTM for the Globe. V4. CGIAR / CGIAR. Available from http://srtm.csi.cgiar.org.

Jayachandran S, de Laat J, Lambin EF, Stanton CY, Audy R, Thomas NE. 2017. Cash for carbon: A randomized trial of payments for ecosystem services to reduce deforestation. Science (New York, N.Y.) **357**:267–273.

Jones JPG, Barnes M, Eklund J, Ferraro PJ, Geldmann J, Oldekop JA, Schleicher J. 2022. Quantifying uncertainty about how interventions are assigned would improve impact evaluation in conservation: Reply to Rasolofoson 2022. Conservation Biology **n/a**:e14007.

Joppa L, Pfaff A. 2010. Reassessing the forest impacts of protection: The challenge of nonrandom location and a corrective method. Annals of the New York Academy of Sciences **1185**:135–149.

Joppa LN, Pfaff A. 2009. High and far: Biases in the location of protected areas. PLoS One **4**:e8273.

Joppa LN, Pfaff A. 2011. Global protected area impacts **278**:1633–1638.

Joshi N, Mitchard ETA, Woo N, Torres J, Moll-Rocek J, Ehammer A, Collins M, Jepsen MR, Fensholt R. 2015. Mapping dynamics of deforestation and forest degradation in tropical forests using radar satellite data. Environmental Research Letters **10**:034014. IOP Publishing.

Kariuki J, Birner R. 2016. Are Market-Based Conservation Schemes Gender-Blind? A Qualitative Study of Three Cases From Kenya. Society & Natural Resources **29**:432–447. Routledge.

King G, Nielsen R. 2019. Why Propensity Scores Should Not Be Used for Matching. Political Analysis.

Laurance WF et al. 2011. The fate of Amazonian forest fragments: A 32-Year investigation. Biological Conservation **144**:56–67. Available from https://www.sciencedirect.com/science/article/pii/S0006320710004209 (accessed December 22, 2021).

Laurance WF, Sayer J, Cassman KG. 2014. Agricultural expansion and its impacts on tropical nature. Trends in Ecology & Evolution **29**:107–116. Available from http://dx.doi.org/10.1016/j.tree.2013.12.001.

Lausch A, Erasmi S, King DJ, Magdon P, Heurich M. 2017. Understanding Forest Health with Remote Sensing-Part II—A Review of Approaches and Data Models. Remote Sensing **9**:129. Multidisciplinary Digital Publishing Institute.

Le Toan T et al. 2011. The BIOMASS mission: Mapping global forest biomass to better understand the terrestrial carbon cycle. Remote Sensing of Environment **115**:2850–2860.

Leblois A, Damette O, Wolfersberger J. 2017. What has Driven Deforestation in Developing Countries Since the 2000s? Evidence from New Remote-Sensing Data. World Development **92**:82–102.

Legendre P, Legendre L. 1988. Numerical ecology (2nd Edition), 2nd edition. Elsevier.

Lewis SL, Edwards DP, Galbraith D. 2015. Increasing human dominance of tropical forests. Science (New York, N.Y.) **349**:827–832.

Lewis SL, Wheeler CE, Mitchard ETA, Koch A. 2019. Restoring natural forests is the best way to remove atmospheric carbon. Nature **568**:25–28. Nature Publishing Group. Available from https://www.nature.com/articles/d41586-019-01026-8 (accessed February 16, 2021).

Lima MGB, Persson UM, Meyfroidt P. 2019. Leakage and boosting effects in environmental governance: A framework for analysis. Environmental Research Letters **14**:105006. IOP Publishing. Available from https://doi.org/10.1088/1748-9326/ab4551 (accessed January 26, 2022).

Lin L, Pattanayak SK, Sills EO, Sunderlin WD. 2012. Site selection for forest carbon projects. Pages 209–230 Analysing REDD+: Challenges and choices. CIFOR, Bogor, Indonesia.

Lui GV, Coomes DA. 2015. A Comparison of Novel Optical Remote Sensing-Based Technologies for Forest-Cover/Change Monitoring. Remote Sensing **7**:2781–2807. Multidisciplinary Digital Publishing Institute.

Melo JB, Ziv G, Baker TR, Carreiras JMB, Pearson TRH, Vasconcelos MJ. 2018. Striking divergences in Earth Observation products may limit their use for REDD+. Environmental Research Letters **13**:104020. IOP Publishing.

Meyfroidt P et al. 2018. Middle-range theories of land system change. Global Environmental Change **53**:52–67. Available from http://www.sciencedirect.com/science/article/pii/S0959378018302280.

Meyfroidt P, Börner J, Garrett R, Gardner T, Godar J, Kis-Katos K, Soares-Filho BS, Wunder S. 2020. Focus on leakage and spillovers: Informing land-use governance in a tele-coupled world. Environmental Research Letters **15**:090202. IOP Publishing. Available from https://doi.org/10.1088/1748-9326/ab7397 (accessed January 26, 2022).

Meyfroidt P, Lambin EF. 2009. Forest transition in Vietnam and displacement of deforestation abroad. Proceedings of the National Academy of Sciences of the United States of America **106**:16139–16144. National Academy of Sciences. Available from https://www.pnas.org/content/106/38/16139 (accessed January 6, 2022).

Milne S, Mahanty S, To P, Dressler W, Kanowski P, Thavat M. 2019. Learning From `Actually Existing' REDD+: A Synthesis of Ethnographic Findings. Conservation & Society **17**:84–95. [Ashoka Trust for Research in Ecology and the Environment, Wolters Kluwer India Pvt. Ltd.].

Miranda J, Börner J, Kalkuhl M, Soares-Filho B. 2019. Land speculation and conservation policy leakage in Brazil. Environmental Research Letters **14**:045006. IOP Publishing. Available from https://doi.org/10.1088/1748-9326/ab003a (accessed February 1, 2022).

Mitchard ETA. 2018. The tropical forest carbon cycle and climate change. Nature **559**:527. Nature Publishing Group.

Mitchard ETA, Meir P, Ryan CM, Woollen ES, Williams M, Goodman LE, Mucavele JA, Watts P, Woodhouse IH, Saatchi SS. 2013. A novel application of satellite radar data: Measuring carbon

sequestration and detecting degradation in a community forestry project in Mozambique. Plant Ecology & Diversity **6**:159–170.
Miteva DA, Pattanayak SK, Ferraro PJ. 2012. Evaluation of biodiversity policy instruments: What works and what doesn't? Oxford review of economic policy **28**:69–92. Oxford University Press.
Nolte C, Leverington F, Kettner A, Marr M, Nielsen G, Bomhard B, Stolton S, Stoll-Kleemann S, Hockings M. 2010. Protected Area Management Effectiveness Assessments in Europe: A review of application, methods and results. Bundesamt für Naturschutz.
Nolte C, Meyer SR, Sims KRE, Thompson JR. 2019. Voluntary, permanent land protection reduces forest loss and development in a rural-urban landscape. Conservation Letters **12**:e12649. Available from https://conbio.onlinelibrary.wiley.com/doi/abs/10.1111/conl.12649 (accessed April 16, 2021).
Oldekop JA et al. 2020. Forest-linked livelihoods in a globalized world. Nature Plants **6**:1400–1407. Nature Publishing Group.
Oldekop JA, Sims KRE, Karna BK, Whittingham MJ, Agrawal A. 2019. Reductions in deforestation and poverty from decentralized forest management in Nepal. Nature sustainability **2**:421–428.
Pagiola S, Honey-Rosés J, Freire-González J. 2016. Evaluation of the Permanence of Land Use Change Induced by Payments for Environmental Services in Quindío, Colombia. PLOS ONE **11**:e0147829. Public Library of Science.
Pan Y et al. 2011. A large and persistent carbon sink in the world's forests. Science (New York, N.Y.) **333**:988–993.
Partners NA. 2020. Goal 1 assessment: Striving to end natural forest loss. Climate Focus.
Pascual U et al. 2017. Valuing nature's contributions to people: The IPBES approach. Current Opinion in Environmental Sustainability **26–27**:7–16.
Pearson TRH, Brown S, Casarim FM. 2014. Carbon emissions from tropical forest degradation caused by logging. Environmental Research Letters **9**:034017. IOP Publishing.
Pendrill F et al. 2022. Disentangling the numbers behind agriculture-driven tropical deforestation. Science (New York, N.Y.) **377**:eabm9267. American Association for the Advancement of Science.
Pendrill F, Persson UM, Godar J, Kastner T. 2019. Deforestation displaced: Trade in forest-risk commodities and the prospects for a global forest transition. Environmental Research Letters **14**:055003. IOP Publishing. Available from https://doi.org/10.1088/1748-9326/ab0d41 (accessed January 30, 2021).
Pettorelli N, Laurance WF, O'Brien TG, Wegmann M, Nagendra H, Turner W. 2014. Satellite remote sensing for applied ecologists: Opportunities and challenges. Journal of Applied Ecology **51**:839–848.
Pfaff A, Robalino J. 2017. Spillovers from Conservation Programs. Annual Review of Resource Economics **9**:299–315. Available from https://doi.org/10.1146/annurev-resource-100516-053543 (accessed June 28, 2019).
Pynegar EL, Gibbons JM, Asquith NM, Jones JPG. 2019. What role should randomized control trials play in providing the evidence base for conservation? Oryx : the journal of the Fauna Preservation Society:1–10. Cambridge University Press.
Rappaport DI, Morton DC, Longo M, Keller M, Dubayah R, dos-Santos MN. 2018. Quantifying long-term changes in carbon stocks and forest structure from Amazon forest degradation. Environmental Research Letters **13**:065013. IOP Publishing.
Rasolofoson RA. 2022. Statistical matching for conservation science revisited: Response to Schleicher et al. 2020. Conservation Biology **n/a**:e14006.
Rasolofoson RA, Ferraro PJ, Jenkins CN, Jones JPG. 2015. Effectiveness of Community Forest Management at reducing deforestation in Madagascar. Biological Conservation **184**:271–277.

Reed J et al. 2020. The extent and distribution of joint conservation-development funding in the tropics. One earth (Cambridge, Mass.) **3**:753–762.
Robalino J, Pfaff A, Villalobos L. 2017. Heterogeneous Local Spillovers from Protected Areas in Costa Rica. Journal of the Association of Environmental and Resource Economists **4**:795–820. The University of Chicago Press. Available from https://www.journals.uchicago.edu/doi/full/10.1086/692089 (accessed November 15, 2021).
Robalino JA, Pfaff A. 2012. Contagious development: Neighbor interactions in deforestation. Journal of Development Economics **97**:427–436. Available from https://www.sciencedirect.com/science/article/pii/S0304387811000605 (accessed November 22, 2021).
Roopsind A, Sohngen B, Brandt J. 2019. Evidence that a national REDD+ program reduces tree cover loss and carbon emissions in a high forest cover, low deforestation country. Proceedings of the National Academy of Sciences of the United States of America **116**:24492–24499. Available from https://www.pnas.org/content/early/2019/11/12/1904027116 (accessed November 24, 2019).
Rosenbaum PR. 2002. Observational studies / Paul R. Rosenbaum. Page (Rosenbaum PR, editor)2nd ed. New York : Springer, c2002., New York. Available from https://idiscover.lib.cam.ac.uk/primo-explore/fulldisplay?docid=44CAM_ALMA21524262980003606&context=L&vid=44CAM_PROD&search_scope=SCOP_CAM_ALL&tab=cam_lib_coll&lang=en_US.
Samii C, Lisiecki M, Kulkarni P, Paler L, Chavis L, Snilstveit B, Vojtkova M, Gallagher E. 2014. Effects of Payment for Environmental Services (PES) on Deforestation and Poverty in Low and Middle Income Countries: A Systematic Review. Campbell Systematic Reviews **10**:1–95.
Santika T, Wilson KA, Law EA, St John FAV, Carlson KM, Gibbs H, Morgans CL, Ancrenaz M, Meijaard E, Struebig MJ. 2021. Impact of palm oil sustainability certification on village well-being and poverty in Indonesia. Nature sustainability **4**:109–119. Nature Publishing Group. Available from https://www.nature.com/articles/s41893-020-00630-1 (accessed June 25, 2021).
Schleicher J, Eklund J, Barnes MD, Geldmann J, Oldekop JA, Jones JPG. 2019a. Statistical matching for conservation science. Conservation Biology **34**:538–549.
Schleicher J, Peres CA, Amano T, Llactayo W, Leader-Williams N. 2017. Conservation performance of different conservation governance regimes in the Peruvian Amazon. Scientific Reports **7**:11318.
Schleicher J, Peres CA, Leader-Williams N. 2019b. Conservation performance of tropical protected areas: How important is management? Conservation Letters **0**:e12650.
Seddon N, Chausson A, Berry P, Girardin CAJ, Smith A, Turner B. 2020. Understanding the value and limits of nature-based solutions to climate change and other global challenges. Philos. Trans. R. Soc. B Biol. Sci. **375**:20190120. Royal Society.
Sexton JO et al. 2013. Global, 30-m resolution continuous fields of tree cover: Landsat-based rescaling of MODIS vegetation continuous fields with lidar-based estimates of error. International Journal of Digital Earth **6**:427–448.
Seymour F. 2020. 4 Reasons Why a Jurisdictional Approach for REDD+ Crediting Is Superior to a Project-Based Approach. Available from https://www.wri.org/insights/insider-4-reasons-why-jurisdictional-approach-redd-crediting-superior-project-based (accessed January 26, 2022).
Seymour F, Busch J. 2016. Why forests? Why now?: The science, economics, and politics of tropical forests and climate change. Brookings Institution Press.
Seymour F, Duchelle AE, Brockhaus M, Angelsen A, Larson AM, Moeliono M, Wong GY, Pham TT, Martius C. 2018. REDD+: Lessons from National and Subnational Implementation. World Resources Institute. Available from https://www.cifor.org/knowledge/publication/6934/ (accessed August 19, 2021).
Seymour F, Harris NL. 2019. Reducing tropical deforestation. Science (New York, N.Y.) **365**:756–757. American Association for the Advancement of Science.

Shimada M, Itoh T, Motooka T, Watanabe M, Shiraishi T, Thapa R, Lucas R. 2014. New global Forest/Non-Forest maps from ALOS PALSAR data (2007–2010). Remote Sensing of Environment **155**:13–31.

Shoch D, Eaton J, Settelmyer S. 2011. Project developer's guidebook to VCS REDD methodologies. Version 2.0. Conservation International.

Shukla PR et al. 2019. Climate Change and Land: An IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems.

Sills EO, Atmadja SS, de Sassi C, Duchelle AE, Kweka DL, Resosudarmo IAP, Sunderlin WD. 2014. REDD+ on the ground: A case book of subnational initiatives across the globe. CIFOR. Available from https://market.android.com/details?id=book-co4UBgAAQBAJ.

Simonet G et al. 2020. ID-RECCO, International Database on REDD+ projects and programs: Linking Economics, Carbon and Communities. V4. CIFOR. Available from http://www.reddprojectsdatabase.org.

Simonet G, Subervie J, Ezzine-de-Blas D, Cromberg M, Duchelle AE. 2019. Effectiveness of a REDD+ Project in Reducing Deforestation in the Brazilian Amazon. American journal of agricultural economics **101**:211–229.

Simula M. 2009. Towards defining forest degradation: Comparative analysis of existing definitions. Forest Resources Assessment Working Paper **154**. FAO. Available from http://www.ardot.fi/Documents_2/Degradationdefintions.pdf.

Skutsch M, Turnhout E. 2020. REDD+: If communities are the solution, what is the problem? World Development **130**:104942. Available from http://www.sciencedirect.com/science/article/pii/S0305750X20300681 (accessed July 14, 2020).

Soto-Navarro C et al. 2020. Mapping co-benefits for carbon storage and biodiversity to inform conservation policy and action. Philosophical Transactions of the Royal Society B: Biological Sciences **375**:20190128. Royal Society. Available from https://royalsocietypublishing.org/doi/full/10.1098/rstb.2019.0128 (accessed October 8, 2020).

Stern NH. 2007. The Economics of Climate Change: The Stern Review. Cambridge University press.

Stickler C, Duchelle AE, Nepstad D, Ardila JP. 2018. Subnational jurisdictional approaches. Page Transforming REDD+: Lessons and new directions. CIFOR.

Streck C. 2021. REDD+ and leakage: Debunking myths and promoting integrated solutions. Climate Policy. Taylor & Francis. Available from https://www.tandfonline.com/doi/abs/10.1080/14693062.2021.1920363 (accessed January 7, 2022).

Stuart EA. 2010. Matching Methods for Causal Inference: A Review and a Look Forward. Statistical Science **25**:1–21. Available from https://projecteuclid.org/euclid.ss/1280841730 (accessed January 11, 2019).

Sunderlin WD, de Sassi C, Sills EO, Duchelle AE, Larson AM, Resosudarmo IAP, Awono A, Kweka DL, Huynh TB. 2018a. Creating an appropriate tenure foundation for REDD+: The record to date and prospects for the future. World Development **106**:376–392.

Sunderlin WD, Larson AM, Sarmiento Barletti J. 2018b. Land and carbon tenure: Some–but Insufficient–Progress. Page Transforming REDD+: Lessons and new directions. CIFOR.

Sunderlin WD, Sills EO. 2012. REDD+ projects as a hybrid of old and new forest conservation approaches. Pages 177–191 Analysing REDD+: Challenges and choices. CIFOR, Bogor, Indonesia.

Symes WS, Edwards DP, Miettinen J, Rheindt FE, Carrasco LR. 2018. Combined impacts of deforestation and wildlife trade on tropical biodiversity are severely underestimated. Nature communications **9**:4052. Nature Publishing Group.

Sze JS, Carrasco LR, Childs D, Edwards DP. 2022. Reduced deforestation and degradation in Indigenous Lands pan-tropically. Nature sustainability **5**:123–130. Nature Publishing Group.

Tagesson T, Schurgers G, Horion S, Ciais P, Tian F, Brandt M, Ahlström A, Wigneron J-P, Ardö J, Olin S. 2020. Recent divergence in the contributions of tropical and boreal forests to the terrestrial carbon sink. Nature Ecology & Evolution **4**:202–209. Nature Publishing Group.

TSVCM. 2021. Taskforce on scaling voluntary carbon markets. Institute of International Finance. Available from www.iif.com/Portals/1/Files/TSVCM_Report.pdf.

Turubanova S, Potapov PV, Tyukavina A, Hansen MC. 2018. Ongoing primary forest loss in Brazil, Democratic Republic of the Congo, and Indonesia. Environmental Research Letters **13**:074028. IOP Publishing.

Tyukavina A, Baccini A, Hansen MC, Potapov PV, Stehman SV, Houghton RA, Krylov AM, Turubanova S, Goetz SJ. 2015. Aboveground carbon loss in natural and managed tropical forests from 2000 to 2012. Environmental Research Letters **10**:074002. IOP Publishing.

UNEP-WCMC, IUCN. 2019. Protected Planet: The World Database on Protected Areas (WDPA). Available from www.protectedplanet.net.

UNFCCC. 2021. Glasgow Leaders' Declaration on Forests and Land Use. Available from https://ukcop26.org/glasgow-leaders-declaration-on-forests-and-land-use/ (accessed January 11, 2022).

UN-REDD. 2022. Pricing Forest Carbon.

Usmani F, Jeuland M, Pattanayak SK. 2018. NGOs and the effectiveness of interventions. UNU-WIDER, Helsinki, Finland.

Usmani F, Jeuland M, Pattanayak SK. 2022. NGOs and the Effectiveness of Interventions. The Review of Economics and Statistics:1–45.

Vancutsem C, Achard F, Pekel J-F, Vieilledent G, Carboni S, Simonetti D, Gallego J, Aragão LEOC, Nasi R. 2021. Long-term (1990–2019) monitoring of forest cover changes in the humid tropics. Science Advances **7**:eabe1603. American Association for the Advancement of Science. Available from https://advances.sciencemag.org/content/7/10/eabe1603 (accessed March 12, 2021).

Vancutsem C, Achard F, Pekel J-F, Vieilledent G, Carboni S, Simonetti D, Gallego J, Marelli A. 2020. Long-term monitoring of tropical moist forest extent (from 1990 to 2019) Description of the dataset. JRC European Comission. Available from https://ec.europa.eu/jrc/en/publication/long-term-monitoring-tropical-moist-forest-extent-1990-2019 (accessed March 12, 2021).

Velly GL, Sauquet A, Cortina-Villar S. 2017. PES Impact and Leakages over Several Cohorts: The Case of the PSA-H in Yucatan, Mexico. Land Economics **93**:230–257. University of Wisconsin Press. Available from http://le.uwpress.org/content/93/2/230 (accessed December 18, 2020).

Watson JEM et al. 2018. The exceptional value of intact forest ecosystems. Nature ecology & evolution **2**:599–610. Nature Publishing Group.

Watson JEM, Dudley N, Segan DB, Hockings M. 2014. The performance and potential of protected areas. Nature **515**:67–73.

Weiss DJ et al. 2018. A global map of travel time to cities to assess inequalities in accessibility in 2015. Nature **553**:333–336. Available from http://dx.doi.org/10.1038/nature25181.

Weisse M, Goldman ED. 2020. We lost a football pitch of primary rainforest every 6 seconds in 2019. WRI.

West TAP, Börner J, Sills EO, Kontoleon A. 2020. Overstated carbon emission reductions from voluntary REDD+ projects in the Brazilian Amazon. Proceedings of the National Academy of Sciences **117**. National Academy of Sciences. Available from https://www.pnas.org/content/early/2020/09/08/2004334117 (accessed September 15, 2020).

West TAP, Caviglia-Harris JL, Martins FSRV, Silva DE, Börner J. 2022. Potential conservation gains from improved protected area management in the Brazilian Amazon. Biological Conservation **269**:109526.

Wheeler D, Guzder-Williams B, Petersen R, Thau D. 2018. Rapid MODIS-based detection of tree cover loss. International Journal of Applied Earth Observation and Geoinformation **69**:78–87.

Wigneron J-P, Fan L, Ciais P, Bastos A, Brandt M, Chave J, Saatchi S, Baccini A, Fensholt R. 2020. Tropical forests did not recover from the strong 2015–2016 El Niño event. Science Advances **6**:eaay4603. American Association for the Advancement of Science. Available from https://advances.sciencemag.org/content/6/6/eaay4603 (accessed February 1, 2021).

Wiik E, d' Annunzio R, Pynegar E, Crespo D, Asquith N, Jones JPG. 2019. Experimental evaluation of the impact of a payment for environmental services program on deforestation. Conservation Science and Practice **0**:e8.

Wilebore B, Voors M, Bulte EH, Coomes D, Kontoleon A. 2019. Unconditional Transfers and Tropical Forest Conservation: Evidence from a Randomized Control Trial in Sierra Leone. American journal of agricultural economics **101**:894–918. Oxford Academic. Available from https://academic.oup.com/ajae/article/101/3/894/5364749 (accessed October 25, 2020).

Wilebore RAL. 2015. Valuing forests in tropical landscapes in the context of REDD+. University of Cambridge.

Withey K et al. 2018. Quantifying immediate carbon emissions from El Niño-mediated wildfires in humid tropical forests. Philosophical Transactions of the Royal Society B: Biological Sciences **373**:20170312. Royal Society.

Wren-Lewis L, Becerra-Valbuena L, Houngbedji K. 2020. Formalizing land rights can reduce forest loss: Experimental evidence from Benin. Science Advances **6**:eabb6914. American Association for the Advancement of Science. Available from https://advances.sciencemag.org/content/6/26/eabb6914 (accessed July 15, 2020).

Wunder S. 2008. How do we deal with leakage. Pages 65–75 Moving ahead with REDD: Issues, options and implications.

Wunder S. 2013. When payments for environmental services will work for conservation. Conservation Letters **6**:230–237. Available from https://onlinelibrary.wiley.com/doi/abs/10.1111/conl.12034 (accessed November 5, 2021).

Wunder S, Börner J, Ezzine-de-Blas D, Feder S, Pagiola S. 2020a. Payments for environmental services: Past performance and pending potentials. Annual Review of Resource Economics **12**:209–234. Annual Reviews.

Wunder S, Brouwer R, Engel S, Ezzine-de-Blas D, Muradian R, Pascual U, Pinto R. 2018. From principles to practice in paying for nature's services. Nature Sustainability **1**:145.

Wunder S, Duchelle AE, de Sassi C, Sills EO, Simonet G, Sunderlin WD. 2020b. REDD+ in Theory and Practice: How Lessons From Local Projects Can Inform Jurisdictional Approaches **3**. Frontiers. Available from https://www.frontiersin.org/articles/10.3389/ffgc.2020.00011/full (accessed September 7, 2020).

Xia J, Liu S, Liang S, Chen Y, Xu W, Yuan W. 2014. Spatio-Temporal Patterns and Climate Variables Controlling of Biomass Carbon Stock of Global Grassland Ecosystems from 1982 to 2006. Remote Sensing **6**:1783–1802. Multidisciplinary Digital Publishing Institute. Available from https://www.mdpi.com/2072-4292/6/3/1783 (accessed December 30, 2022).

Zeileis A, Köll S, Graham N. 2020. Various Versatile Variances: An Object-Oriented Implementation of Clustered Covariances in R. Journal of Statistical Software **95**:1–36. Available from https://www.jstatsoft.org/index.php/jss/article/view/v095i01 (accessed January 9, 2021).

zu Ermgassen E, Godar J, Lathuillière MJ, Löfgren P, Gardner T, Vasconcelos A, Meyfroidt P. 2020. The origin, supply chain, and deforestation risk of Brazil's beef exports. Proceedings of the National

Academy of Sciences **117**:31770–31779. National Academy of Sciences. Available from https://www.pnas.org/content/early/2020/11/30/2003270117 (accessed December 2, 2020).

APPENDIX A

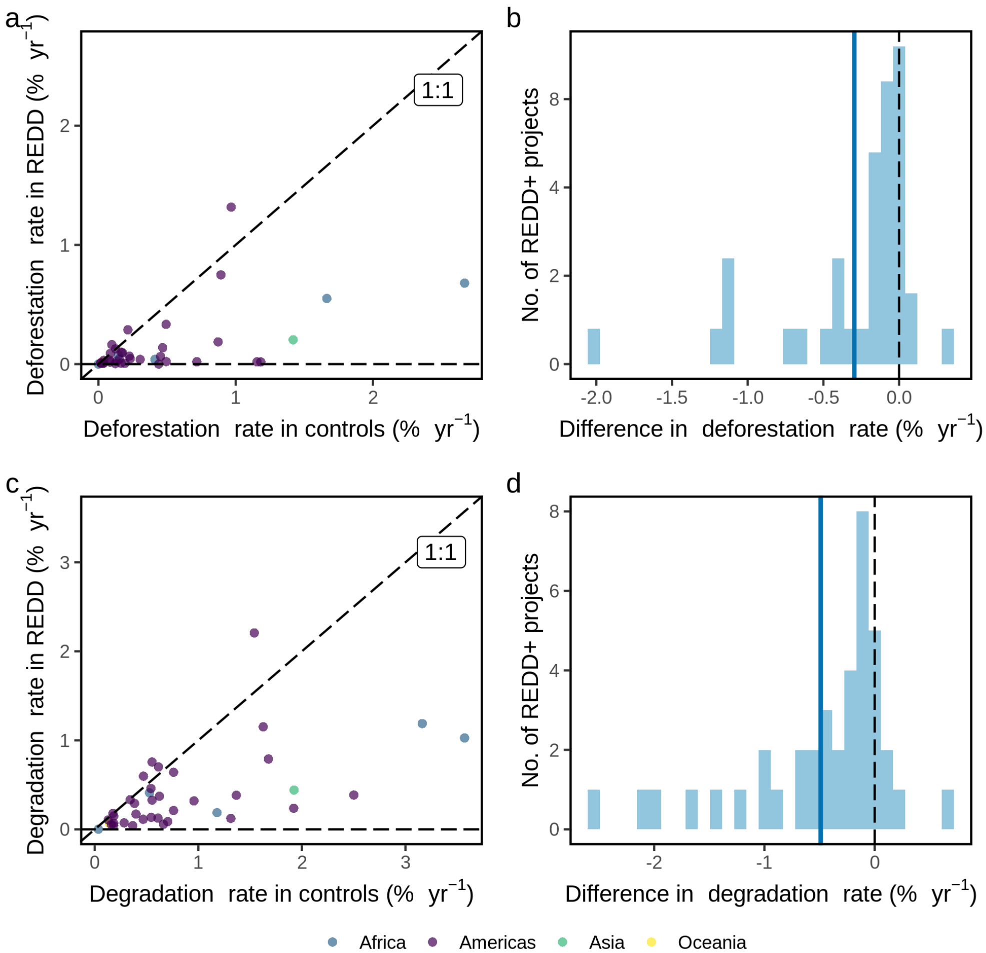

Figure A.1: Reductions in deforestation and degradation rates resulting from REDD+ projects over their first five years of operation, excluding protected area portions. **a** and **b** scatterplots of deforestation and degradation rates in REDD+ projects versus matched control pixels; the reduction in deforestation resulting from a project is given by the vertical distance between the datapoint and the 1:1 line (with one example shown by a blue arrow); **c** and **d** histograms of the reduction in deforestation and degradation rates, with the mean shown as a blue line.

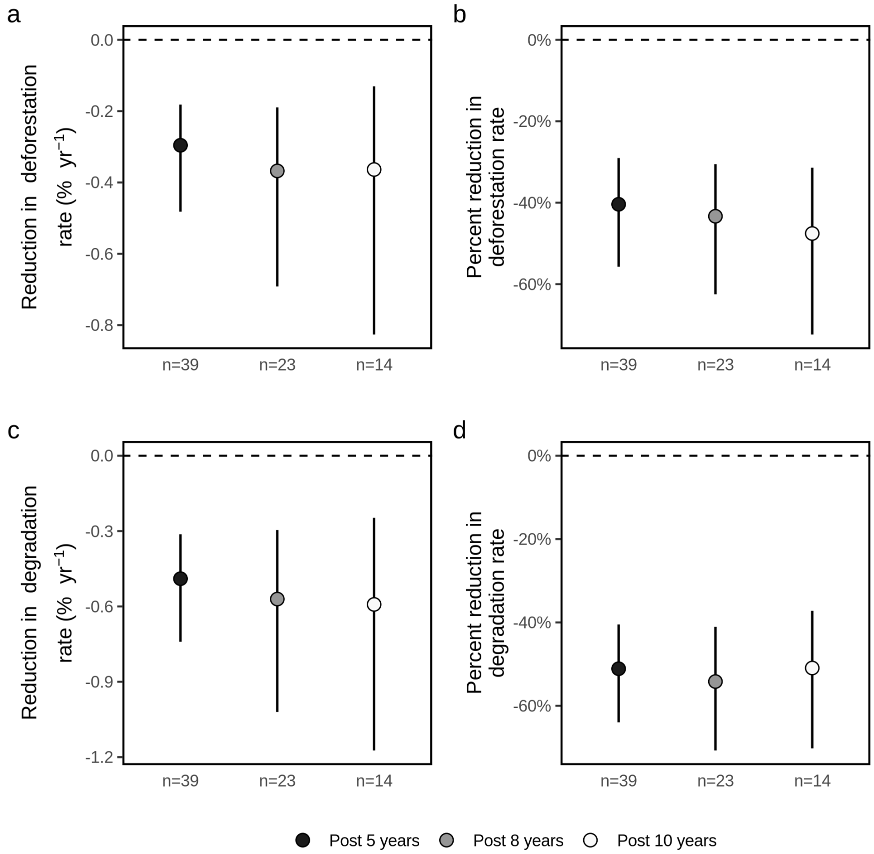

Figure A.2: Avoided deforestation and degradation associated with REDD+ projects considering all forested sites, excluding protected area portions. **a** and **b** mean reductions in annual deforestation and degradation rates. **c** and **d** mean percent reductions in deforestation and degradation rates (relative to controls). Whisker lines show 95% confidence intervals estimated by nonparametric bootstrapping**.** We analysed the temporal variation in effect sizes by examining subsets of projects operating for at least 8 years(n=23) and 10 years (n=14).

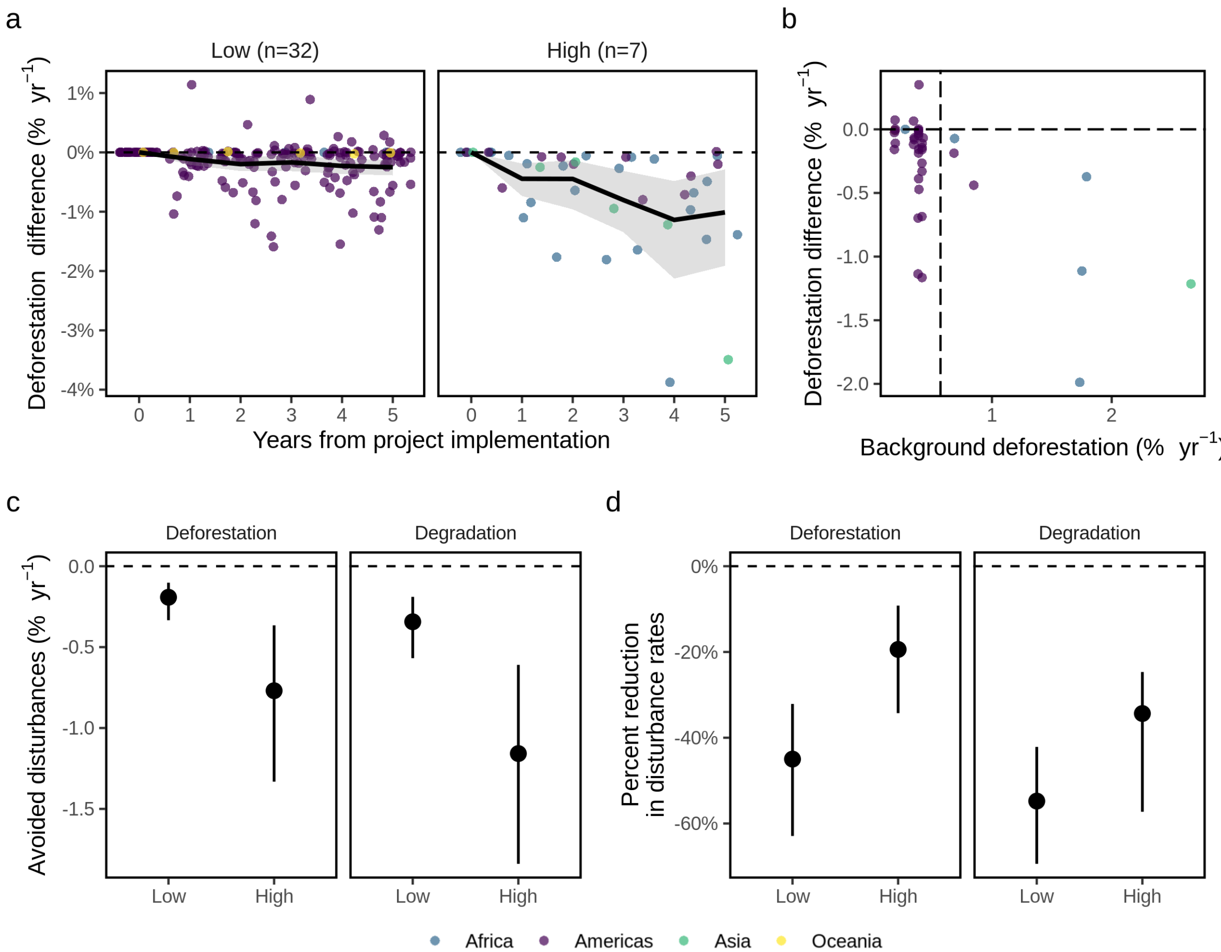

Figure A.3: REDD+ project effectiveness by background rates of deforestation, excluding protected area portions. **a** annual differences in deforestation (with jitter; % yr^{-1}) between project areas and matched controls over five years after project implementation, with a black line showing the mean annual differences and 95% CI shaded in grey; **b** mean differences in deforestation rates (% yr^{-1}) against country-level background deforestation rates (calculated for the project implementation period) with a vertical line showing the pantropical average rate of deforestation (0.57% yr^{-1}); **c** mean differences in deforestation and degradation rates; **d** mean percent reductions in deforestation and degradation rates relative to controls. The 95% CIs displayed at **a, c** and **d** were estimated using non-parametric bootstrapping.

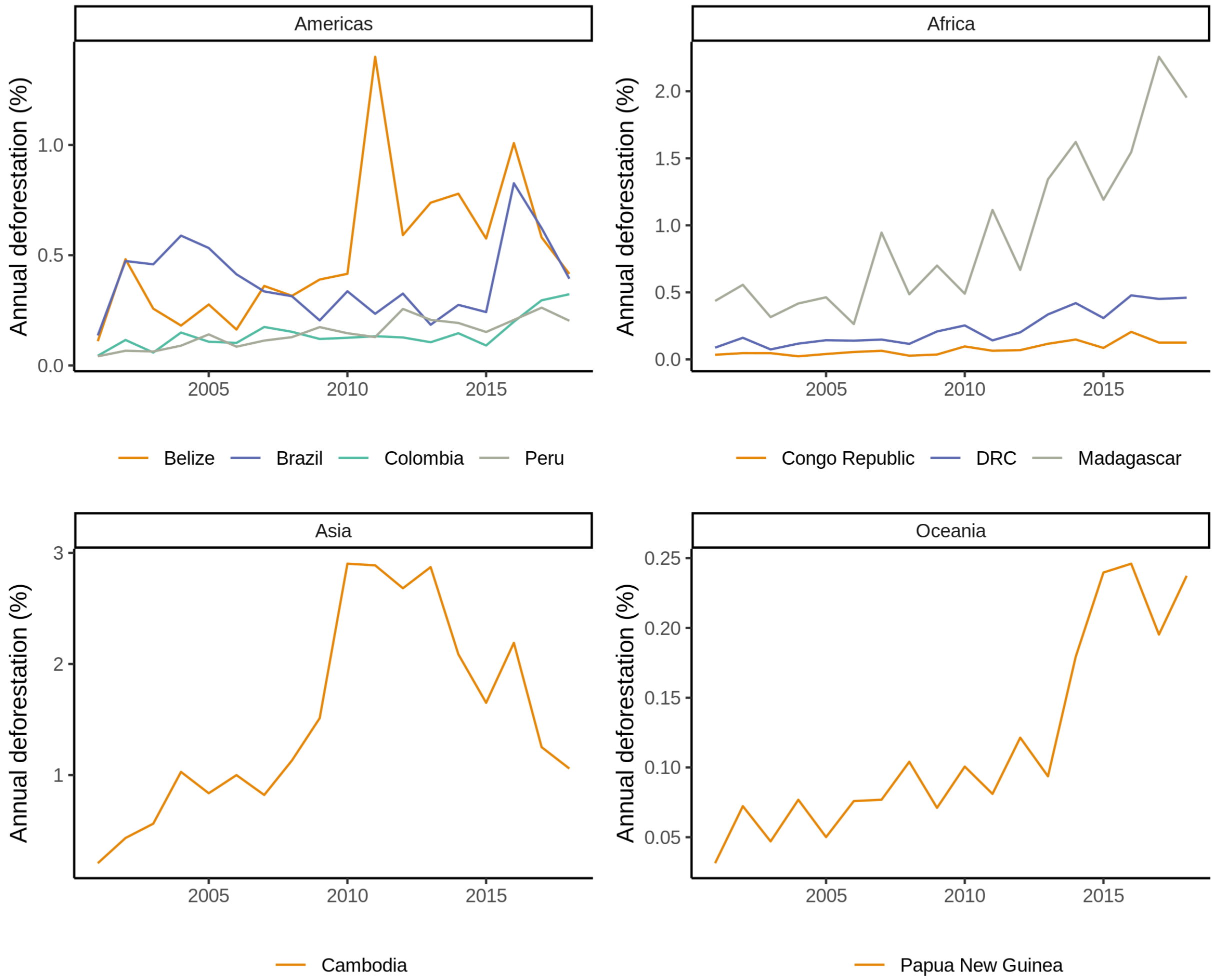

Figure A.4: Country-wide estimates of forest loss in the humid tropics for countries selected in the analysis. The time series were produced by subtracting deforestation and conversion to other land classes from annual estimates undisturbed and degraded Tropical Moist across all countries spanned by our sample of 40 REDD+ projects.

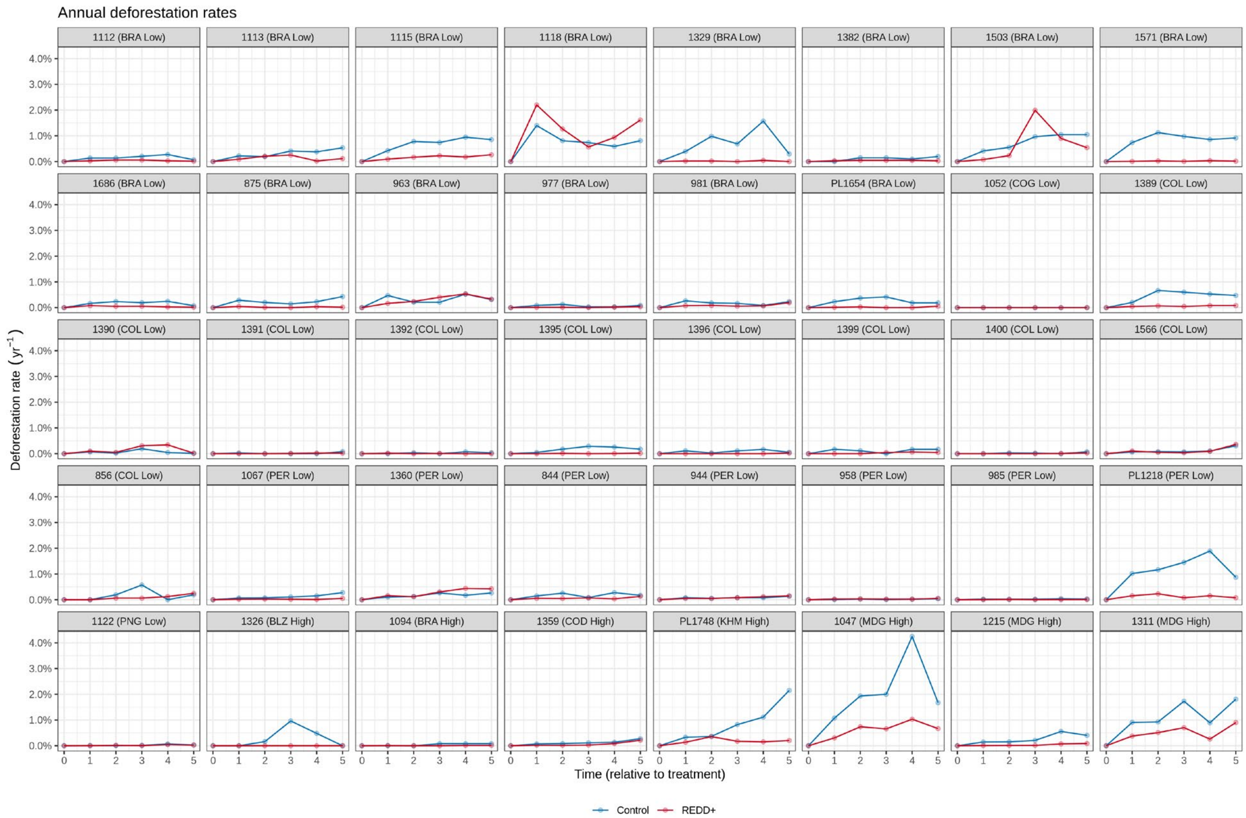

Figure A.5: Annual deforestation rates. Time series of the annual rates of deforestation (% yr^{-1}) observed in the 40 matched project sites and their controls, five years after the implementation of REDD+. Banner indicates the project ID (see Table A.4), followed by the ISO-3 country code and the classification according to the project background rates of deforestation.

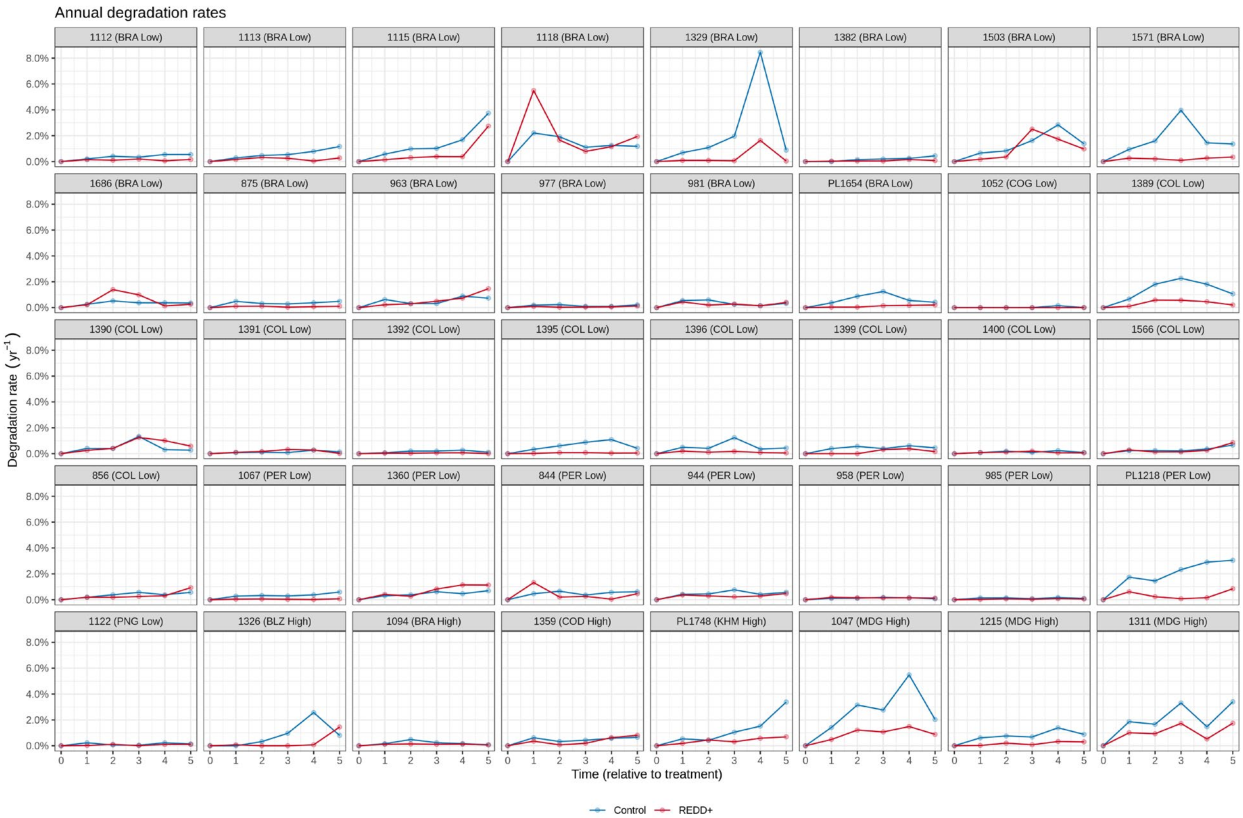

Figure A.6: Annual degradation rates. Time series of the annual rates of forest degradation (% yr^{-1}) observed in the 40 matched project sites and their controls, five years after the implementation of REDD+. Banner indicates the project ID (see Table A.4), followed by the ISO-3 country code and the classification according to the project background rates of deforestation.

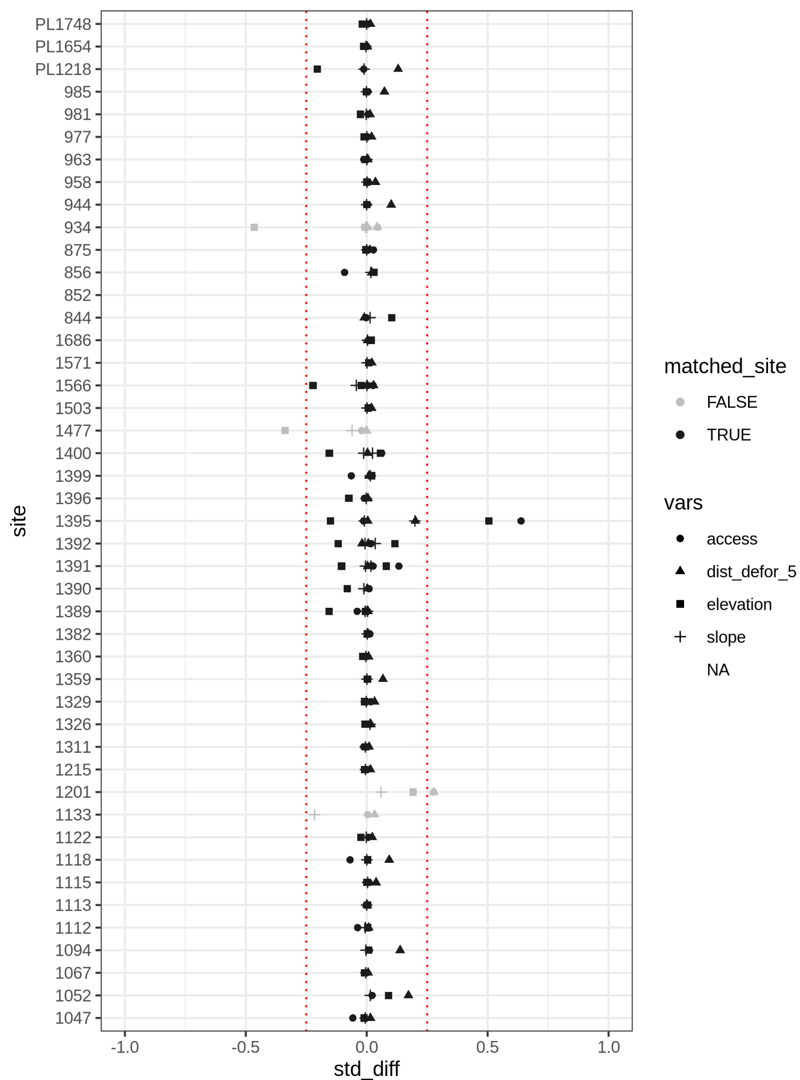

Figure A.7: Post-matching standardized mean differences across covariates. Sites were matched on accessibility (access), mean distance to deforestation 5 years prior to project commencement (dist_defor_5), elevation and slope; with pixels selected from the same biome and country. As 10 sites are intersected by two biomes, we constructed separate matched sets for each project/biome split (shown above as rows with 2 sets of dots). Sites were considered as successfully matched if at least 90% of the sampled pixels scored an absolute standardized mean difference of <0.25 (red lines) across all covariates. Project 1395, intersects two biomes. The pixels in one biome were not adequately matched but represent less than 10% of the total pixel count so this project is retained.

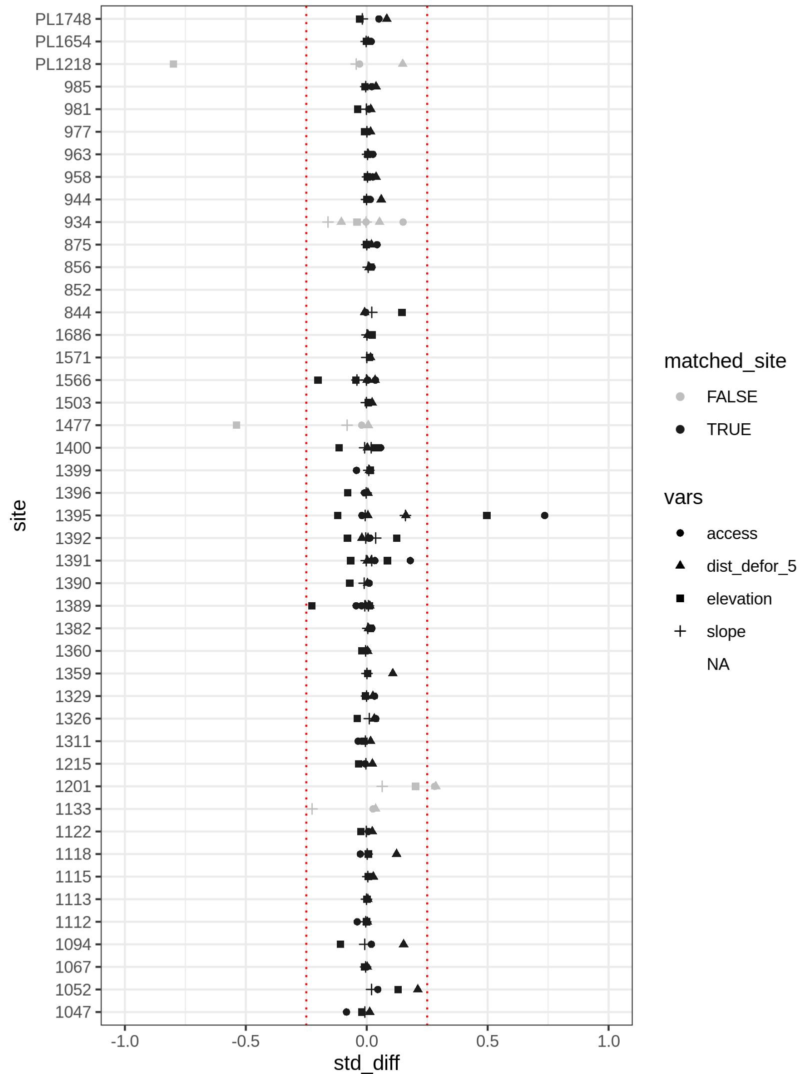

Figure A.8: Post-matching standardized mean differences across covariates excluding protected area portions. Sites were matched on accessibility (access), mean distance to deforestation 5 years prior to project commencement (dist_defor_5), elevation and slope; with pixels selected from the same biome and country. As 10 sites are intersected by two biomes, we constructed separate matched sets for each project/biome split (shown above as rows with 2 sets of dots). Sites were considered as successfully matched if at least 90% of the sampled pixels scored an absolute standardized mean difference of <0.25 (red lines) across all covariates. Project 1395, intersects two biomes. The pixels in one biome were not adequately matched but represent less than 10% of the total pixel count so this project is retained.

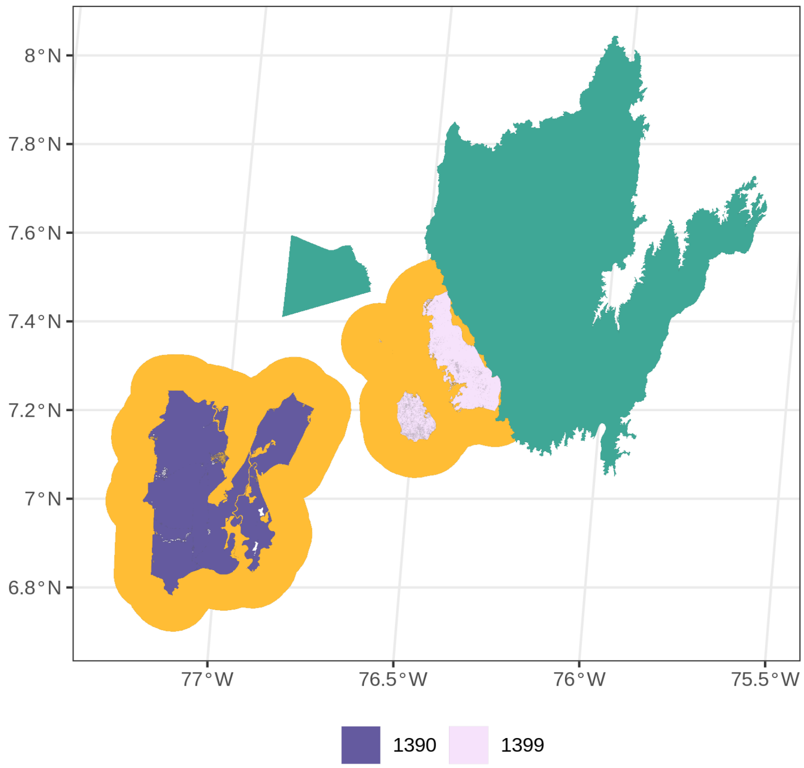

Figure A.9: Representation of buffer areas (in yellow) examined for leakage patterns (VCS projects 1390 and 1399). Overlaps with protected areas (green) were excluded from the examination

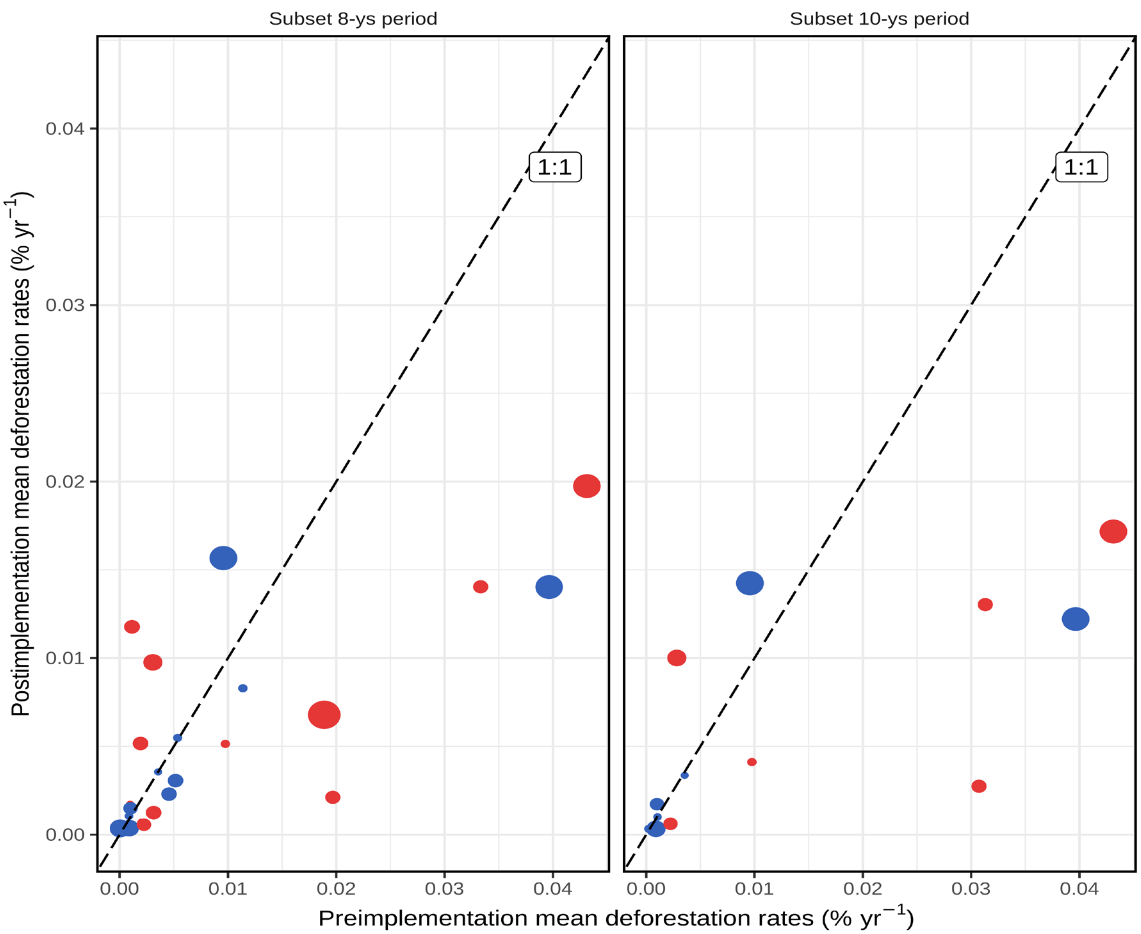

Figure A.10: Evidence of deforestation leakage over eight (a) and ten (b) years after project implementation. Dots depict the mean rates of deforestation in the buffer areas before (x-axis), and after (y-axis) the commencement of projects. Red colouring indicates significant differences in annual deforestation rates before and after project implementation (bootstrapped t-tests, $p < 0.05$) across the 10 km boundary area. A diagonal dotted line was added to depict a theoretical 1:1 relationship between axes. Dot sizes were scaled to reflect the background deforestation rates observed at the host country within the first 5 years of project implementation

Figure A.11: Distribution of land cover classes at the project starting date. The banner indicates the project ID, followed by the total count of pixels within the boundaries and the host country, for the 71 sites for which we were able to collect REDD+ boundary information. Projects with at least 80% of evergreen forest cover at the starting date were included in the analyses.

TABLE A.1: COVARIATES SELECTED FOR MATCHING. WE EXACT MATCHED ON BIOME AND COUNTRY (SHOWN IN BOLD).

Category	Rationale for inclusion	Source	Variables	Format and resolution
Physical characteristics	Elevation and slope influence land use practices, with higher elevation and steeper slopes consistently associated with a lower risk of deforestation (Joppa & Pfaff 2009).	SRTM Digital Elevation Data Version 4 (Jarvis et al. 2008)	Elevation (meters) Slope (degrees)	Raster (90m2)
Accessibility	Accessibility to forest lands from urban centres is associated with a higher risk of deforestation (Chomitz & Thomas 2003; Rudel et al. 2009).	Global accessibility map (Weiss et al. 2018)	Time travel to population centres (seconds)	Raster (300m2) Temporal res: 2015
Bioclimatic	The biophysical characteristics of biomes, such as temperature, precipitation and primary productivity; can influence the probability of	Terrestrial Ecoregions	Biome	Vector

	land clearance due to agricultural suitability (Busch & Ferretti-Gallon 2017).	(Dinerstein et al. 2017).		
Governance / political economy	Deforestation is influenced by domestic policies, governance, economic processes and markets (Umemiya et al. 2010; Ceddia et al. 2014)	Large Scale International Boundary (LSIB) dataset	Country	Vector Temporal res: 2013
Proximity to deforestation	Proximity to cleared lands is consistently associated with higher deforestation risk (Busch & Ferretti-Gallon 2017). We computed the average distance to the closest deforested pixel in the 5 years prior project implementation, for each pixel included in the matching assessment.	Own analysis using the JRC Tropical Moist Forest database (Vancutsem et al. 2021)	Euclidean distance to deforested areas (meters)	Raster (30m2)

Table A.2: Logistic regression (binary response variable, 0 = VCS REDD+ sites excluded from the analysis N=41; 1 = sites included in the analysis N=30) to assess the probability of a project being selected due to covariate characteristics. Statistically significant effects (P<0.05) indicate the probability of a selection bias in our sample (shown in bold). The p-value were adjusted through a post-hoc Bonferroni correction.

	Estimate	Std. Error	z value	Pr(>)
(Intercept)	-7.79	2.83	-2.76	0.04
Time travel to population centres (seconds) in 2015	0.00	0.00	2.39	0.02
Elevation (m)	0.00	0.00	-0.47	1
Slope (º)	0.00	0.08	0.03	1
Human Development Index (2010-2018)	9.22	3.71	2.49	0.06
Project area (km)	0.00	0.00	0.14	1
Background deforestation rate (yr^{-1})	0.00	0.00	-0.08	1

Table A.3: Characteristics of the 40 VCS projects included in the analysis.

ID	Name	Country and ISO 3 code	Start year	Area (ha)	PA overlap (%)	Mean deforestation REDD+ (y^{-1})	Mean deforestation control (y^{-1})	Mean degradation REDD+ (y^{-1})	Mean degradation control (y^{-1})	Deforestation at host country	Threat group	Undisturbed forest cover (%) at starting date
1326	Laguna Seca Forest Carbon Project	Belize (BLZ)	2011	8723	0.2	0	0.32	0.32	0.93	0.8	High	94.9
1094	Ecomapua Amazon REDD Project	Brazil (BRA)	2002	99149	99.8	0.01	0.05	0.11	0.22	0.7	High	97.8

981	ADPML Portel-Para REDD Project	Brazil (BRA)	2008	150204	0	0.09	0.18	0.29	0.37	0.4	Low	94.6
977	RMDLT Portel-Para REDD Project	Brazil (BRA)	2008	211131	0	0.02	0.07	0.07	0.16	0.4	Low	98
1118	Suruí Forest Carbon Project	Brazil (BRA)	2009	33626	100	1.32	0.87	2.21	1.53	0.4	Low	89.3
875	Florestal Santa Maria Project	Brazil (BRA)	2009	72120	0	0.02	0.26	0.09	0.39	0.4	Low	98
1113	The Valparaiso Project	Brazil (BRA)	2011	29195	0	0.14	0.35	0.21	0.65	0.4	Low	94.2
963	The Purus Project	Brazil (BRA)	2011	35746	7.1	0.33	0.35	0.64	0.58	0.4	Low	96.8
1112	The Russas Project	Brazil (BRA)	2011	42788	0	0.04	0.16	0.13	0.41	0.4	Low	97.8
1115	Jari/Amapá REDD+ Project	Brazil (BRA)	2011	78232	1.4	0.19	0.75	0.79	1.6	0.4	Low	85.1
1329	Maísa REDD+ Project	Brazil (BRA)	2012	30242	0	0.02	0.79	0.39	2.61	0.4	Low	97.1
1382	The Envira Amazonia Project - A Tropical Forest Conservation Project in Acre	Brazil (BRA)	2012	39749	0	0.04	0.12	0.07	0.21	0.4	Low	99.2
1503	Resex Rio Preto-Jacundá REDD+ Project	Brazil (BRA)	2012	101592	99.1	0.75	0.81	1.15	1.47	0.4	Low	98
PL1654	Fortaleza Ituxi	Brazil (BRA)	2013	47327	0	0.02	0.28	0.12	0.69	0.4	Low	97.3

	REDD Project											
1571	Manoa REDD+ Project	Brazil (BRA)	2013	73620	0.1	0.02	0.92	0.24	1.87	0.4	Low	98.5
1686	Agrocortex REDD Project	Brazil (BRA)	2014	187740	0	0.05	0.18	0.6	0.38	0.4	Low	98.9
PL1748	Southern Cardamoms REDD+ Project	Cambodia (KHM)	2010	500452	29.6	0.2	0.96	0.44	1.38	2.7	High	84.3
856	The Chocó-Darién Conservation Corridor REDD Project	Colombia (COL)	2010	10687	95.7	0.1	0.19	0.37	0.42	0.4	Low	95.4
1399	Mutatá REDD+ Project	Colombia (COL)	2013	40471	0.7	0.03	0.12	0.17	0.48	0.4	Low	99.1
1391	SUPP REDD+ Project	Colombia (COL)	2013	54210	49.7	0.01	0.02	0.18	0.14	0.4	Low	98.5
1389	Acapa - Bajo Mira Y Frontera REDD+ Project	Colombia (COL)	2013	55448	11.5	0.06	0.49	0.38	1.52	0.4	Low	91.8
1400	Concosta REDD+ Project	Colombia (COL)	2013	64997	49.7	0.01	0.02	0.11	0.14	0.4	Low	98.2
1392	Cajambre REDD+ Project	Colombia (COL)	2013	65845	1.4	0.01	0.03	0.04	0.17	0.4	Low	98.5
1395	Bajo Calima y Bahía Málaga (BCBM) REDD+ Project	Colombia (COL)	2013	91971	27.3	0.01	0.19	0.06	0.67	0.4	Low	97.7
1566	REDD+ Project Resguardo Indigena Unificado Selva de	Colombia (COL)	2013	1557190	0	0.13	0.12	0.33	0.34	0.4	Low	83.5

	Mataven											
1396	Rio Pepe y ACABA REDD+ Project	Colombia (COL)	2014	61132	0	0	0.09	0.13	0.59	0.3	Low	98.3
1390	Carmen del Darién REDD+ Project	Colombia (COL)	2014	134243	0	0.16	0.07	0.7	0.54	0.3	Low	91.6
1052	North Pikounda REDD+	Congo (COG)	2012	92735	2.8	0	0	0	0.03	0.3	Low	100
1359	Isangi REDD+ Project	Congo DRC (COD)	2009	199421	0	0.07	0.14	0.41	0.52	0.7	High	98.1
1215	The Makira Forest Protected Area in Madagascar	Madagascar (MDG)	2005	374708	0	0.04	0.29	0.19	0.86	1.8	High	90
1047	Carbon Emissions Reduction Project in the Forest Corridor Ambositra-Vondrozo (COFAV)	Madagascar (MDG)	2007	144336	0.6	0.68	2.19	1.03	2.97	1.7	High	90.8
1311	Carbon Emissions Reduction Project in the Corridor Ankeniheny-Zahamena (CAZ) Protected Area	Madagascar (MDG)	2008	392643	0.8	0.55	1.25	1.19	2.34	1.8	High	94.5
1122	April Salumei	Papua New	2009	653933	35	0.02	0.03	0.07	0.14	0.2	Low	96.6

	REDD Project	Guinea (PNG)										
844	Madre de Dios Amazon REDD Project	Peru (PER)	2006	98539	0	0.07	0.19	0.46	0.53	0.2	Low	93.1
985	Cordillera Azul National Park REDD project	Peru (PER)	2006	1362722	100	0.01	0.02	0.05	0.13	0.2	Low	97.5
944	Alto Mayo Conservation Initiative	Peru (PER)	2007	178704	98	0.09	0.08	0.33	0.52	0.2	Low	89.5
1067	Reduction of deforestation and degradation in Tambopata National Reserve and Bahuaja-Sonene National Park within the area of Madre de Dios region	Peru (PER)	2008	565849	99.6	0.02	0.13	0.04	0.37	0.2	Low	95.7
1360	Forest Management to reduce deforestation and degradation in Shipibo Conibo and Cacataibo Indigenous commu	Peru (PER)	2010	129260	0	0.29	0.18	0.76	0.49	0.2	Low	97.6

	nities of Ucayali region											
958	Biocorredor Martín Sagrado REDD+ project	Peru (PER)	2010	300275	0	0.03	0.02	0.15	0.12	0.2	Low	95
PL1218	Evio Kuiñaji Ese´Eja Cuana, To Mitigate Climate Change, Madre de Dios	Peru (PER)	2011	8724	18.4	0.14	1.28	0.38	2.3	0.2	Low	93.7

Predictors of deforestation

We fitted models using a sample of treated and control pixels to explore the importance of the matching covariates in predicting deforestation outcomes and selection to treatment. We used a stratified random sampling design to select humid forest pixels within each treated area and their surrounding landscapes within the same biome and country. We sampled 20k observations distributed in proportion to the area size of REDD+ interventions, with a balanced composition of treated and control observations (e.g. equal number of treated and untreated pixels). For each pixel, we collected information on elevation, slope, mean distance to the cities and mean distance to the closest deforested pixels, and distance to agricultural fields. To account for differing socio-economic factors present in the examined countries, we included estimates of the Human Development Index (HDI), Gross Domestic Product (GDP) and the World Bank (WB) governance indicators over the period 2010-2018. Before fitting the models, we investigated potential collinearity among selected covariates and identified high collinearity for HDI, GDP and the WB indicators. We fitted univariate models for collinear variables to predict deforestation and treatment assignment, and selected the one with the best model fit, assessed with the AIC. This led to the inclusion of the world bank indicators in further analyses.

We fitted generalised linear models (GLM) with increasing complexity to explore variability in deforestation outcomes and its major determinants. Our minimal specification model with a

Bernoulli distribution included the following covariates: elevation, slope, distance to cities, and mean distance to the closest deforested pixel prior to deforestation. The response variable is a binary variable indicating a deforested pixel (1) or a pixel that retained forest cover (0) by the end of the examined period (encompassing the years 1990-2019). We scaled all covariates excepting the WB indicator before modelling. We removed extreme residuals (n=33) after an initial exploration and determined minor underdispersion (0.89) of the minimal model. We explored more complex models that included WB indicators and distance to crops and assessed their performance using Bayes Factor scores.

The final parsimonious model included the WB indicators in addition to the initial model covariates. This suggest that distance to recent forest clearings and distance to population centres are strong predictors of deforestation. Quality checks suggests that the model explains a moderate proportion of the observed variation in deforestation across the examined landscapes (Nagelkerke's r^2 0.47) although with some departure from assumptions.

```
Call:
glm(formula = is_defor ~ elevation + slope + access + dist_defor_mean_5 +
wb, family = binomial, data = data)

Deviance Residuals:
Min    1Q  Median    3Q   Max
-1.2587 -0.2130 -0.0021  0.0000  3.6939

Coefficients:
Estimate Std. Error z value Pr(>|z|)
(Intercept)    -1.451e+02 5.669e+00 -25.596 < 2e-16 ***
elevation     -2.111e-01 4.360e-02 -4.842 1.29e-06 ***
slope       -1.284e-01 4.396e-02 -2.922 0.003481 **
access      -6.405e-01 7.050e-02 -9.084 < 2e-16 ***
dist_defor_mean_5 -1.230e+03 4.857e+01 -25.324 < 2e-16 ***
wb         5.679e-02 1.576e-02  3.603 0.000315 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 9899.2 on 20026 degrees of freedom
Residual deviance: 5877.1 on 20021 degrees of freedom
AIC: 5889.1

Number of Fisher Scoring iterations: 15

APPENDIX B

Examining the differences of Chapter 2 and Chapter 3 effect sizes.

Given the differences observed on the estimates of reduced deforestation between Chapter 2 and Chapter 3, we took a closer look at comparing the results of these two different approaches to matching and their results. A comparison of the matching setups and the resulting outcomes are provided below.

Parameter	Chapter 2	Chapter 3
Units of analysis	~30m^2 pixel	~7 ha plots
Sampling	250 mts. distance between treated samples, random sampling in the landscape area	1 km distance between treated samples, random sampling in landscapes with at least 1 km distance
Matching Covariates	Elevation, Slope, Dist. to deforestation, Accessibility, Biome	Elevation, Slope, Dist. Degradation, Accessibility, Biome, Forest extent at the time of project implementation, Forest extent in 1990
Algorithm	Mahalanobis, matching with replacement. No calipers	Mahalanobis, Random Forest and Propensity score matching. Matching with and without replacement. 5 calipers used (0.1, 0.3, 0.5, 0.7, 0.9).

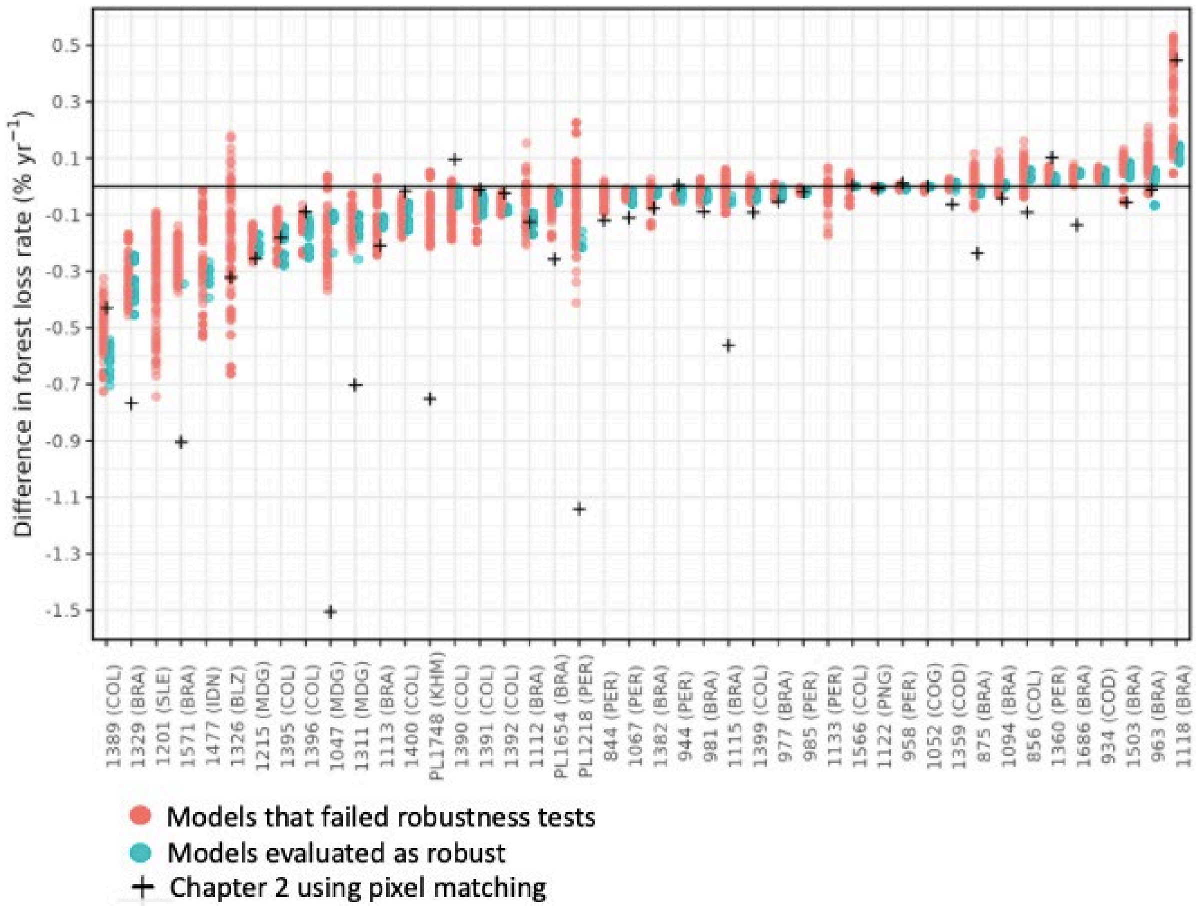

Figure B.1: Avoided deforestation estimates using all matched models following Chapter 3 methods. Dots indicate the models' mean estimates (95% CIs omitted), coloured by matching quality scoring (red=rejected, green=accepted). Where data is available, Chapter 2 estimates are provided with a '+' symbol.

A post-matching examination of matching quality metrics revealed no major differences between Chapter 2 and Chapter 3 sets, as shown in the covariate similarity metrics obtained in both studies, particularly when looking at the matched sets in Chapter 3 that were constructed using a Mahalanobis algorithm and a similar covariate combination than that in Chapter 2. In other words, the differences in the rates of deforestation observed in the control groups of Chapter 2 and Chapter 3 do not seem to be driven by differences in covariate characteristics, as these tended to be similar across studies.

We hypothesised that the differences in sampling density and the decision to match with replacement in Chapter 2 could explain the observed differences in effect sizes. In Chapter 2 we collected many more samples than in Chapter 3 since we used ~30m pixels as a unit of analysis. We matched with replacement, allowing control units to be paired with more than one treated unit, to ensure that as many treated samples as possible were matched. As will be shown in the following section, this led to some control units taking a disproportionate large share of the treatment observations, which in turn might have impacted our effect size estimations.

Uncovering potential issues of matching with replacement.

Matching with replacement led to control observations being selected repeatedly to ensure each treated pixel was matched with a control. In some cases, this led to some controls units being paired with a high number of treated units. While it is expected for control units selected with replacement to have a larger weight in the observational pool due to the number of treatment units attached to them, 'high influence' units, i.e. those holding the largest share of the treatment pool, can have a disproportionate impact in the estimation of outcomes for the control group, which in turn impacts effect size estimations.

To exemplify potential issues with high influence units I examined the control pool selected for the COFAV site in Chapter 2, established in the Ambositra-Vondrozo corridor in Madagascar (VCS site 1047), as it is here where the largest discrepancies are shown between Chapter 2 and Chapter 3, of **-**1.2%**/yr** and -0.2%/yr, respectively, when computing estimates of REDD+ effectiveness. High influence units that changed their outcome status, from forest to deforested, would signal a high deforestation rate in the control group, even though such deforestation took place on a few individual pixels.

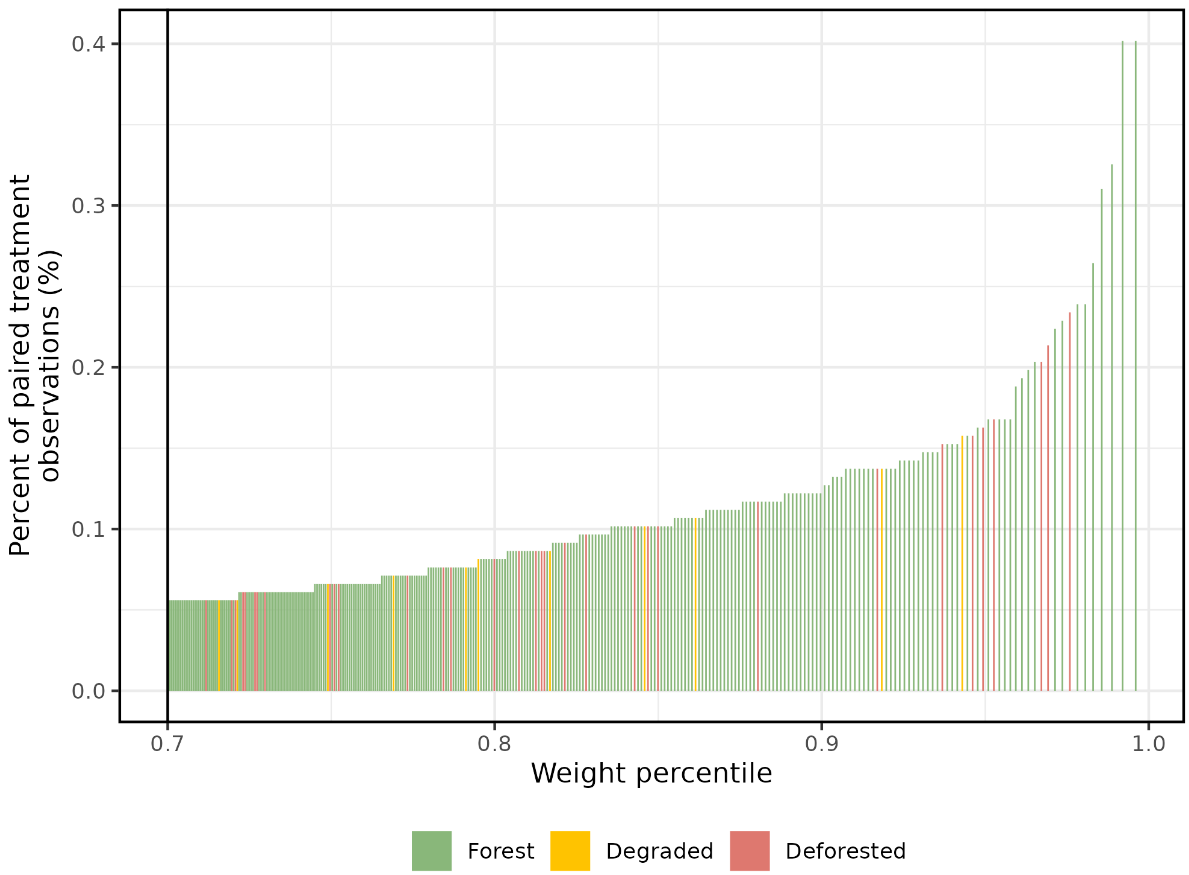

Figure B.2: Forest outcomes in the top 30 weight percentile of the control group, for VCS 1047. Observations arranged by weight percentile (i.e., number of treated units, x-axis) and the percent of paired treated observations allocated to each of them (y-axis). Colour indicates the outcomes status after 5 years of project implementation. The 32 deforested control pixels in this category (red) were paired with 1400 treatment pixels.

When examining the impact of high influence observations across all sites, we observed a correlation between the deforestation outcomes in the top 30 weight percentile and the deforestation rate in the control group as a whole (Spearman's rho= 0.8, p<0.0001, panel a below). This highlights the sensitivity of reported outcomes in the control groups to the outcomes observed in high influence observations, which in our sample amounted to up to 50 units in the observed sites. For VCS 1047, we identified 32 deforested pixels that were paired with 1400 treatment pixels.

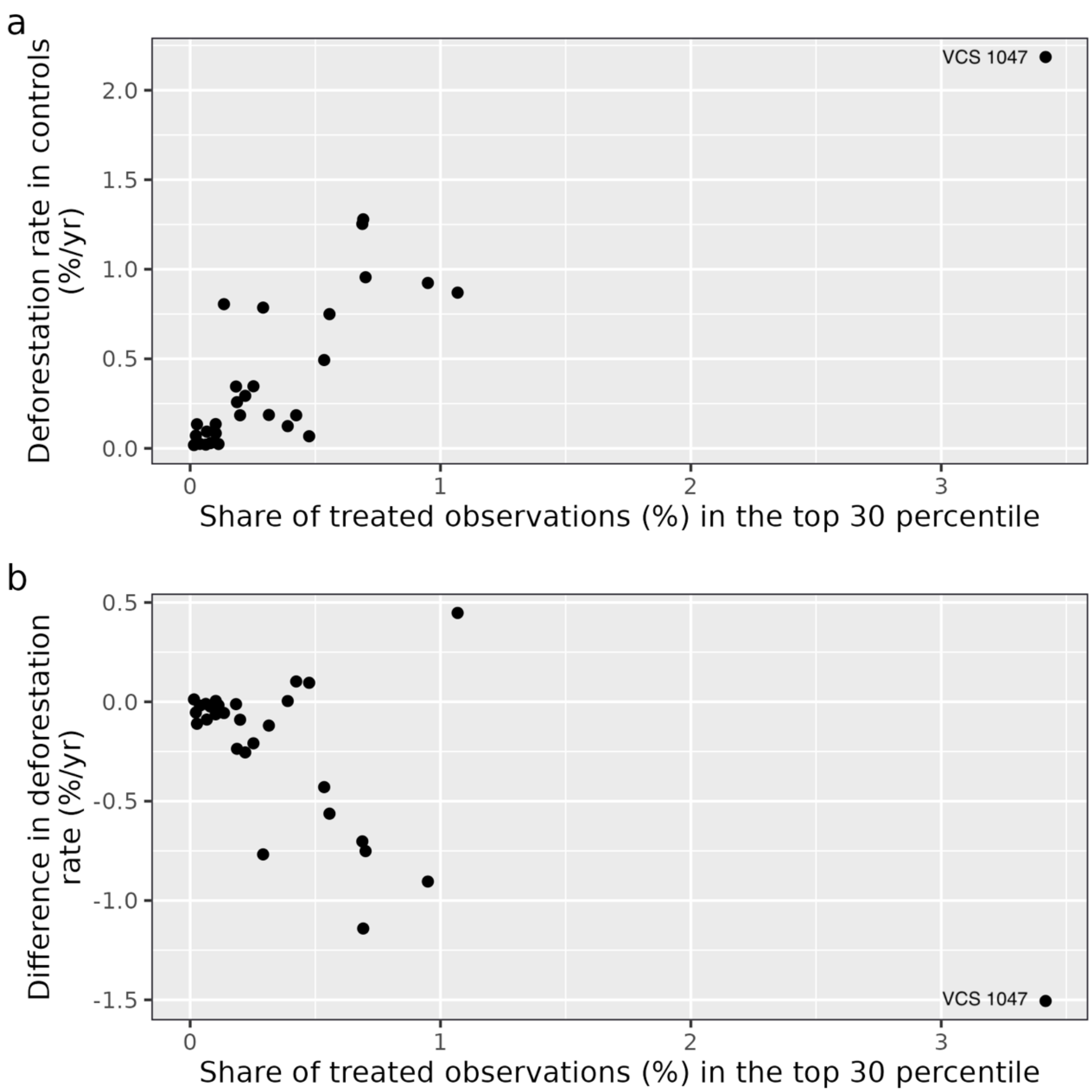

Figure B.3: Deforestation outcomes in the top 30 weight percentile of the control group. Deforested pixels in the top 30 weight percentile, with the percent of total treated units paired to these shown in the x-axis. Deforestation observed in controls after 5 years of project implementation (a), and effect sizes (b) shown int the y-axis. High influence deforested units amounted to relatively few observations, with up to 50 observations in the examined sites. For VCS 1047 (highlighted in the plot), 32 deforested pixels were paired with 1400 treatment pixels.

Concluding remarks.

Matching with replacement is a commonly used form of matching, particularly in situations when finding suitable controls is constrained by the availability of potential candidates (Stuart et al. 2010), however, the implications of having units with a large share of treated observations has not been put in context.

This analysis suggest caution when using matching with replacement, and highlight the needs to introduce quality checks the verify that the outcomes in the control groups is not driven by a few high weight units.

Supplementary figures

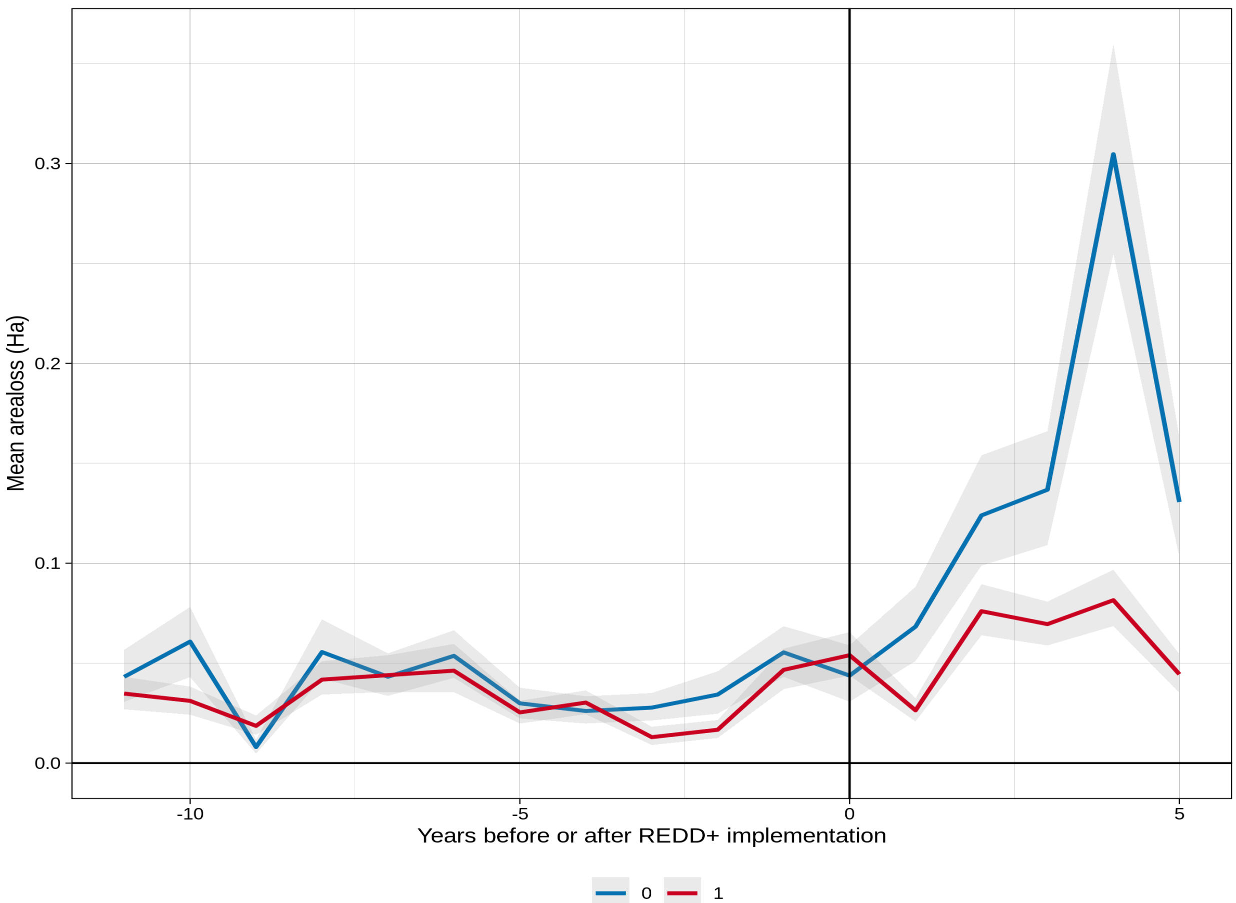

Figure B.4: Patterns of deforestation, before and after project implementation, for VCS site 1047. Plot-level mean estimates of forest loss (Ha) in treated (blue) and control (red) samples. 95% confidence intervals (grey) generated by bootstrapping. This particular time-series comes from a matched set that meets all quality criteria.

Supplementary Tables

TABLE B.1: POOLED TREATMENT EFFECT ESTIMATES OF SELECTED MODELS.

VCS ID (ISO 3)	Area (km2)	No. of matched sets	Bootstrapped effect size [95% CI] (%/yr)	Median baseline (SD) (%/yr)
1118 (BRA)	336	17	0.12 [0.11 0.14]	0.75 (0.1)
1329 (BRA)	302	24	-0.39 [-0.4 -0.37]	0.21 (0.14)
1112 (BRA)	428	20	-0.12 [-0.13 -0.11]	0.08 (0)
PL1654 (BRA)	473	31	-0.03 [-0.04 -0.03]	0.14 (0.04)
1113 (BRA)	292	20	-0.12 [-0.12 -0.11]	0.39 (0.02)
963 (BRA)	357	20	0.04 [0.04 0.05]	0.22 (0.02)
1382 (BRA)	397	35	-0.02 [-0.02 -0.02]	0.04 (0)
875 (BRA)	721	29	0 [0 0.01]	0.2 (0.12)
1571 (BRA)	736	11	-0.35 [-0.36 -0.32]	0.12 (0.02)
1115 (BRA)	782	7	-0.03 [-0.04 -0.03]	0.56 (0.1)
1094 (BRA)	991	20	0.03 [0.02 0.04]	0.19 (0.16)
1503 (BRA)	1016	49	0.07 [0.06 0.08]	0.17 (0.08)
981 (BRA)	1502	70	-0.03 [-0.04 -0.03]	0.24 (0.04)
1686 (BRA)	1877	24	0.05 [0.05 0.05]	0.02 (0.02)
977 (BRA)	2111	73	-0.02 [-0.03 -0.02]	0.07 (0.04)
1359 (COD)	1994	42	0.01 [0.01 0.01]	0.36 (0.03)
934 (COD)	2796	40	0.04 [0.04 0.04]	0.36 (0.04)
1052 (COG)	927	46	-0.01 [-0.01 0]	0 (0)
856 (COL)	107	31	0.04 [0.03 0.04]	0.22 (0.05)
1399 (COL)	405	47	-0.04 [-0.05 -0.03]	0.06 (0.04)
1391 (COL)	542	45	-0.05 [-0.06 -0.05]	0.06 (0.03)
1396 (COL)	611	58	-0.19 [-0.2 -0.17]	0.12 (0.06)
1389 (COL)	554	49	-0.61 [-0.63 -0.6]	0.38 (0.06)
1392 (COL)	658	47	-0.08 [-0.08 -0.07]	0.05 (0.02)
1395 (COL)	920	41	-0.21 [-0.23 -0.19]	0.09 (0.06)
1400 (COL)	650	36	-0.1 [-0.11 -0.09]	0.05 (0.01)
1390 (COL)	1342	20	-0.05 [-0.06 -0.04]	0.17 (0.06)
1566 (COL)	15572	39	0.02 [0.01 0.02]	0.24 (0.08)
1477 (IDN)	1507	24	-0.31 [-0.34 -0.29]	0.23 (0.06)
PL1748 (KHM)	5005	35	-0.12 [-0.14 -0.11]	0.55 (0.1)
1215 (MDG)	3747	46	-0.2 [-0.2 -0.19]	0.9 (0.13)
1311 (MDG)	3926	57	-0.14 [-0.14 -0.13]	0.44 (0.09)
1047 (MDG)	1443	46	-0.2 [-0.23 -0.18]	0.53 (0.16)
PL1218 (PER)	87.2	9	-0.14 [-0.18 -0.1]	0.33 (0.48)
1360 (PER)	1293	53	0.03 [0.03 0.03]	0.16 (0.02)
944 (PER)	1787	66	-0.03 [-0.03 -0.03]	0.38 (0.05)
958 (PER)	3003	83	-0.01 [-0.01 -0.01]	0.15 (0.02)
1067 (PER)	5658	56	-0.04 [-0.04 -0.03]	0.15 (0.02)
985 (PER)	13627	56	-0.01 [-0.01 -0.01]	0.09 (0.03)
1122 (PNG)	6539	44	0 [-0.01 0]	0.05 (0.01)

METRICS OF MATCHING QUALITY AND MATCHING PARAMETERS

We conducted post-hoc tests to evaluate how matching specifications performed across the four specified components of matching quality: 1) post-matching covariate balance between treated and selected control samples, 2) proportion of units matched, 3) similarity of deforestation rates in the period prior to REDD+ (i.e. "parallel trends"), and 4) sensitivity of effect sizes to hidden confounders. We developed a quantitative quality metric for each component, which is developed in the following sections. To determine differences in quality metric due to matching specifications, we ran post-hoc Kruskal–Wallis test on matching runs grouped by algorithm, covariates and calipers thresholds. In all cases, the critical difference among groups was established at $\alpha = 0.05$.

Covariate balance

We derived a single numerical score of covariate balance, by computing the mean of the standardised differences across all the covariates used in each matching run, where values close to or equal to zero indicate the most optimal balance between treated and control covariate distributions.

Variations in covariate balance scores were explained by the choice of algorithm (Kruskal-Wallis rank sum test, $p < 0.001$, H=984.73). Comparisons between groups of model runs showed that RFM models attained the lowest average balance scores (median= 0.965), and were significantly different from PSM models (median=0.897). MDM runs showed the highest post-matching balance scores (median=0.846), and were significantly better than PSM models (Fig. 1.2).

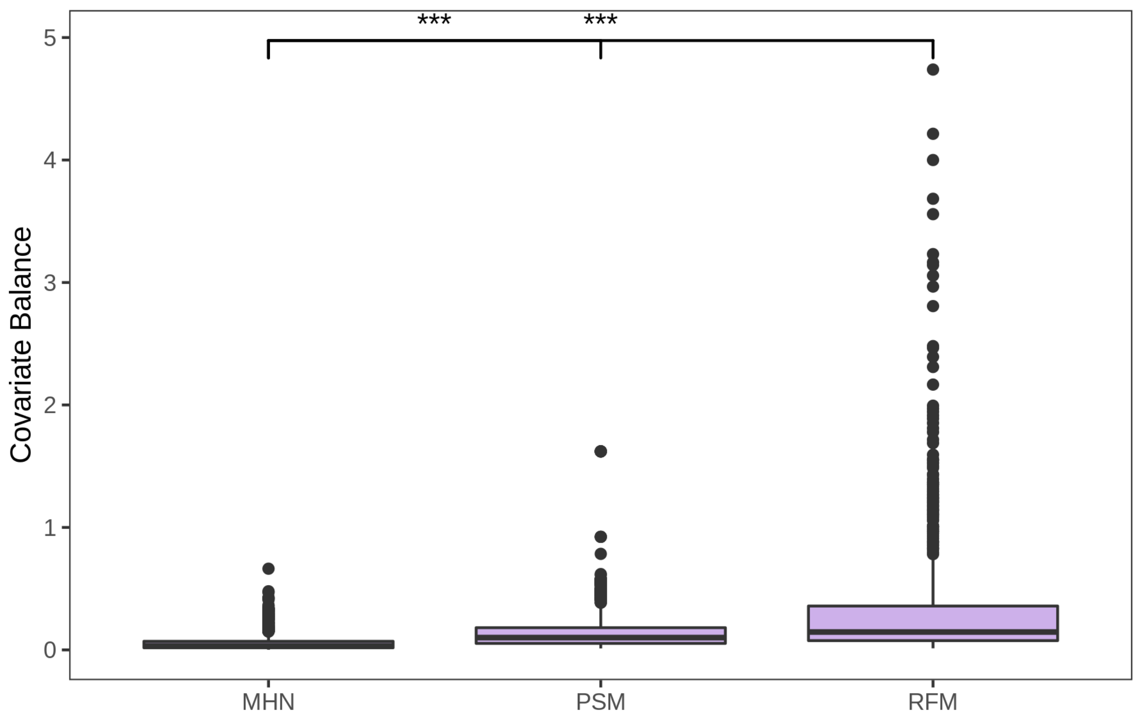

Figure B.2: Balance scores by matching algorithms. Asterisks indicate significant differences between matching algorithms (*p < 0.05; ***p < 0.001). Values close to zero indicate better covariate balance.

Proportion of matched units

Variations in matched sample sizes were explained by the choice of matching algorithm (Kruskal-Wallis rank sum test, $p < 0.001$, H=680.67). Comparisons between model groups showed that the proportion of matched observations using MHN and PSM were virtually similar (median=0.787). RFM models showed significantly lower proportion of matched observation than their counterparts (median = 0.667) (Fig. 1.3). Moreover, we found a moderate association between the proportions of observations matched and the choice of caliper. Comparisons showed a significant difference between models with the lowest (0.1, median=0.742) and the highest caliper threshold (0.9, median=0.775, Fig. 1.4).

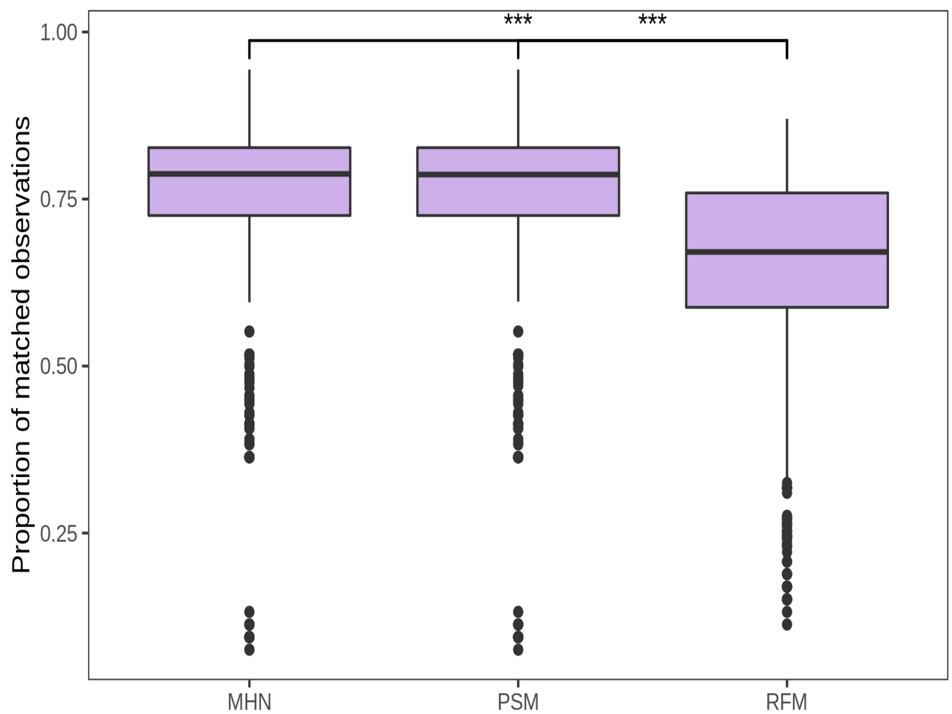

Figure B.3: Proportion of treated observations matched by matching algorithms. Asterisks indicate significant differences between matching algorithms (*p < 0.05; ***p < 0.001). Values closer to 1 indicate better quality scores.

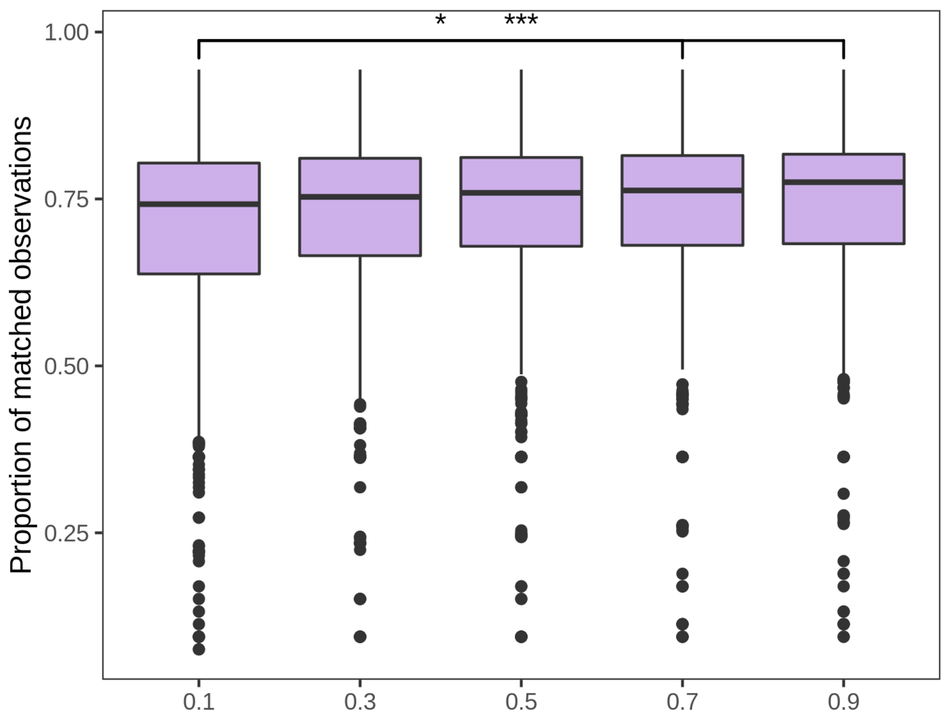

Figure B.4: Proportion of treated observations matched by caliper selection. Asterisks indicate significant differences between matching algorithms (*p < 0.05; ***p < 0.001). Values closer to 1 indicate better quality scores.

Similarity of outcomes prior to the intervention

The similarity outcomes in the pre-treatment period was determined by fitting a mixed effect model: PropAreaforest$_{\text{it}} = \beta_0 + \beta_1$Year$_t + \beta_2 CI_i + \beta_3 Year \times CI_{\text{it}} + (1|t) + \varepsilon$, where the $\beta_3 Year \times CI_{\text{it}}$ term is the parameter that allows us to test for significant differences in the trend lines of treated and control plots prior to treatment. We used the significance value of β_3 as a quantitative metric of quality, where values further away from zero represent matched sets that more credibly meet the parallel trends condition.

Variations in the significance of the β_3 term were explained by the choice of matching algorithm (Kruskal-Wallis rank sum test, p < 0.001, *H*=680.67). Comparisons between model groups showed that MHN a resulted in models with similar magnitudes in the interaction term (median=0.173) and were not significantly different from RFM (median=0.171). However, PSM models resulted in sets whose interaction terms were more likely to be significant (median = 0.039, Fig. 1.5). Moreover, comparisons showed a significant difference between models matched with the 'Base' set of covariates (median=0.008) and models that included a measurement of forest area at the time of project implementation (base + forest area, median=0.206; and base + forest area + forest 1990, median=0.452; Fig. 1.6). We found no association between caliper choice and parallel trends scores.

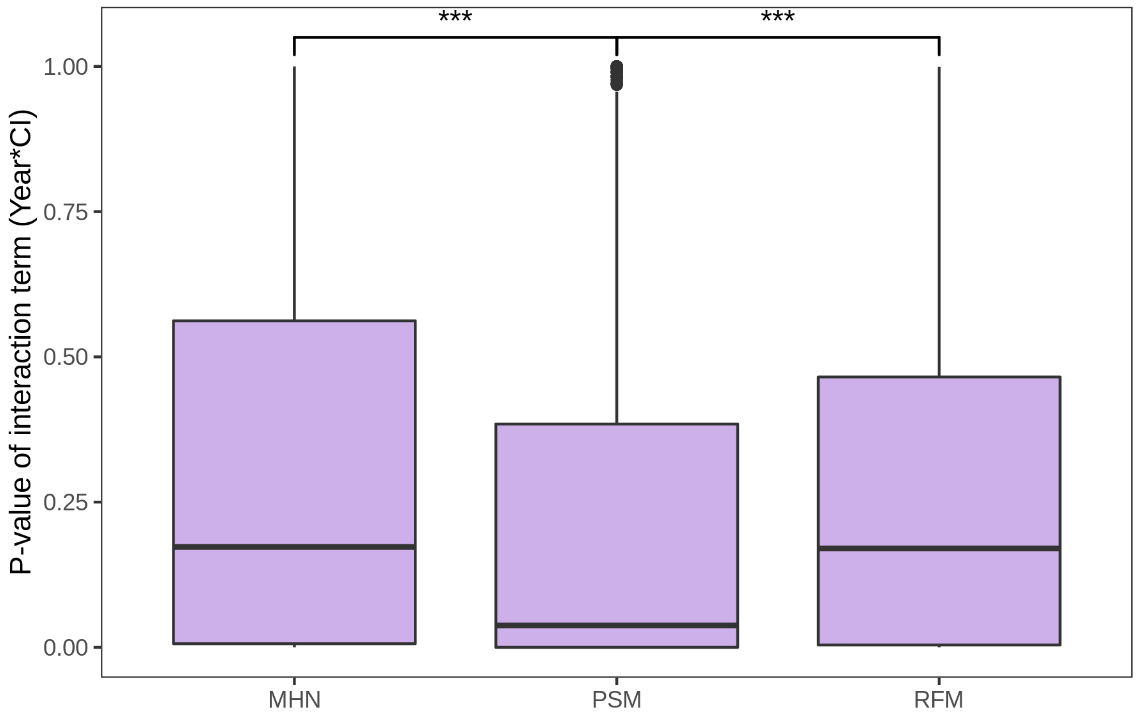

Figure B.5: Significance of the parallel trends' interaction term by matching algorithms. Asterisks indicate significant differences between matching algorithms (*p < 0.05; ***p < 0.001). Values further away from zero indicate better quality scores.

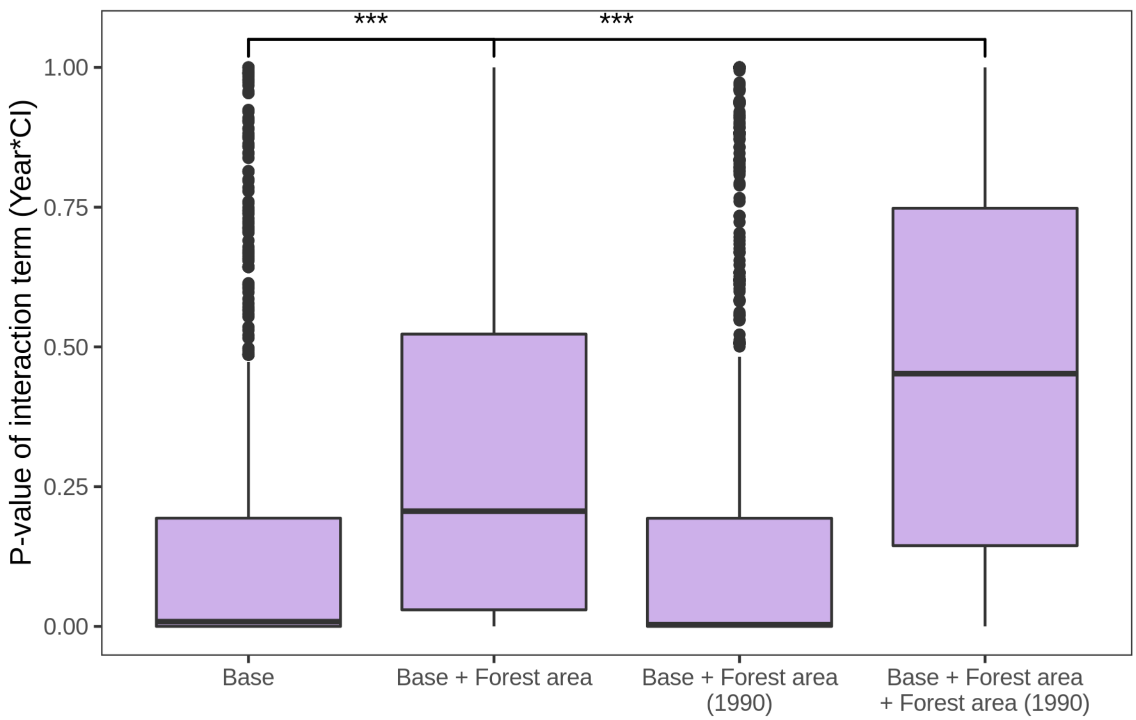

Figure B.6: Significance of the parallel trends interaction term by sets of covariates. Asterisks indicate significant differences between matching algorithms (*p < 0.05; ***p < 0.001). Values further away from zero indicate better quality scores.

Robustness to hidden confounders

We used the residual variation value (RV) as a quantitative metric of robustness to hidden confounders, where values further away from zero represent more robust matched sets.

Variations in thresholds to unobserved confounders RV were explained by the choice of matching algorithm (Kruskal-Wallis rank sum test, $p < 0.001$, H=60.38). Comparisons between model groups showed that RFM models attained the lowest average RV thresholds (median=0.021), and were significantly different from MHD models (median=0.024). PSM runs showed the highest average robustness scores (median=0.031), and were significantly better than MHD models (Fig. 1.7). Moreover, variations in the RV values were observed by the combination of covariates. Comparisons showed a significant difference between models fitted with the core matching covariates (‘Base’, median=0.024) and model runs that included a measurement of the area of forest at the time of project implementation (‘Base + Forest area’, median=0.032; Fig. 1.8).

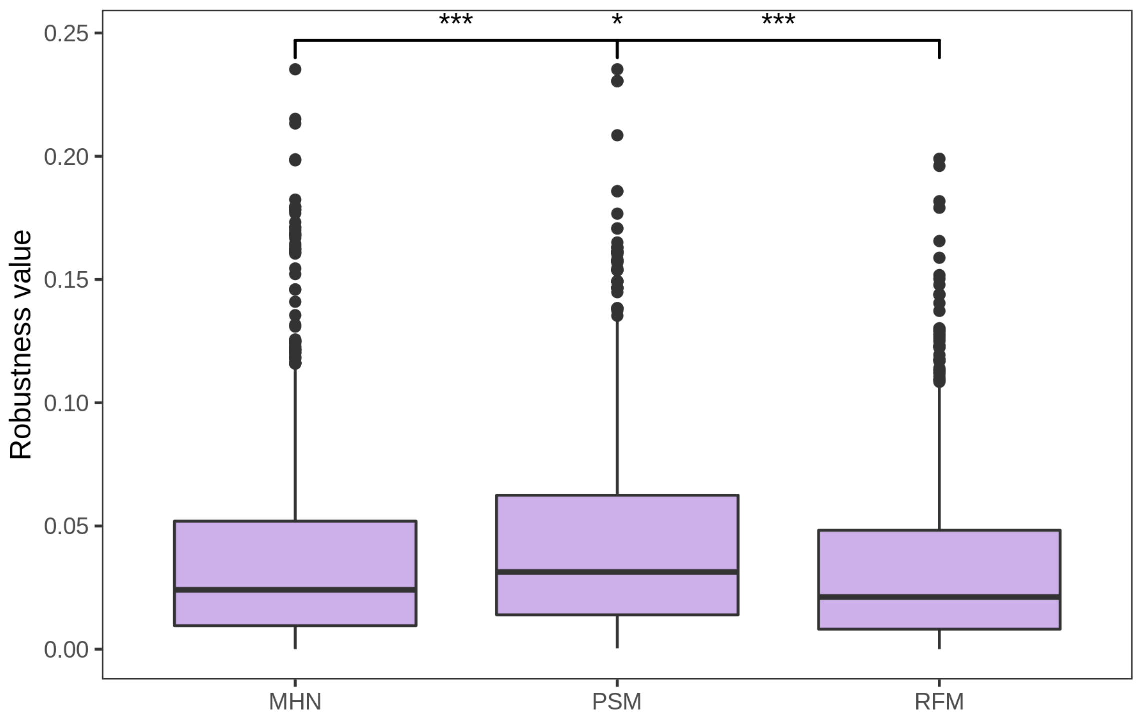

Figure B.7: RV by matching algorithms. Asterisks indicate significant differences between matching algorithms (*p < 0.05; ***p < 0.001). Values further away from zero represent better quality sets.

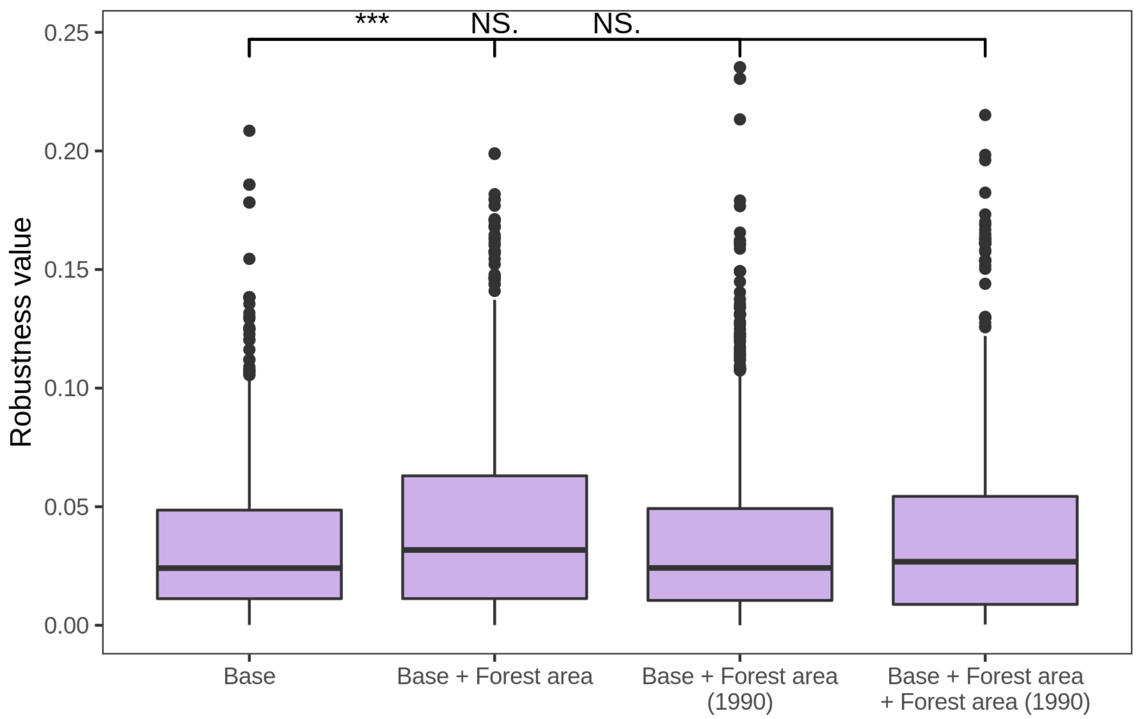

Figure B.8: RV by covariate selection. Asterisks indicate significant differences between sets of matching covariates (*p < 0.05; ***p < 0.001). Values further away from zero represent better quality sets.

APPENDIX C

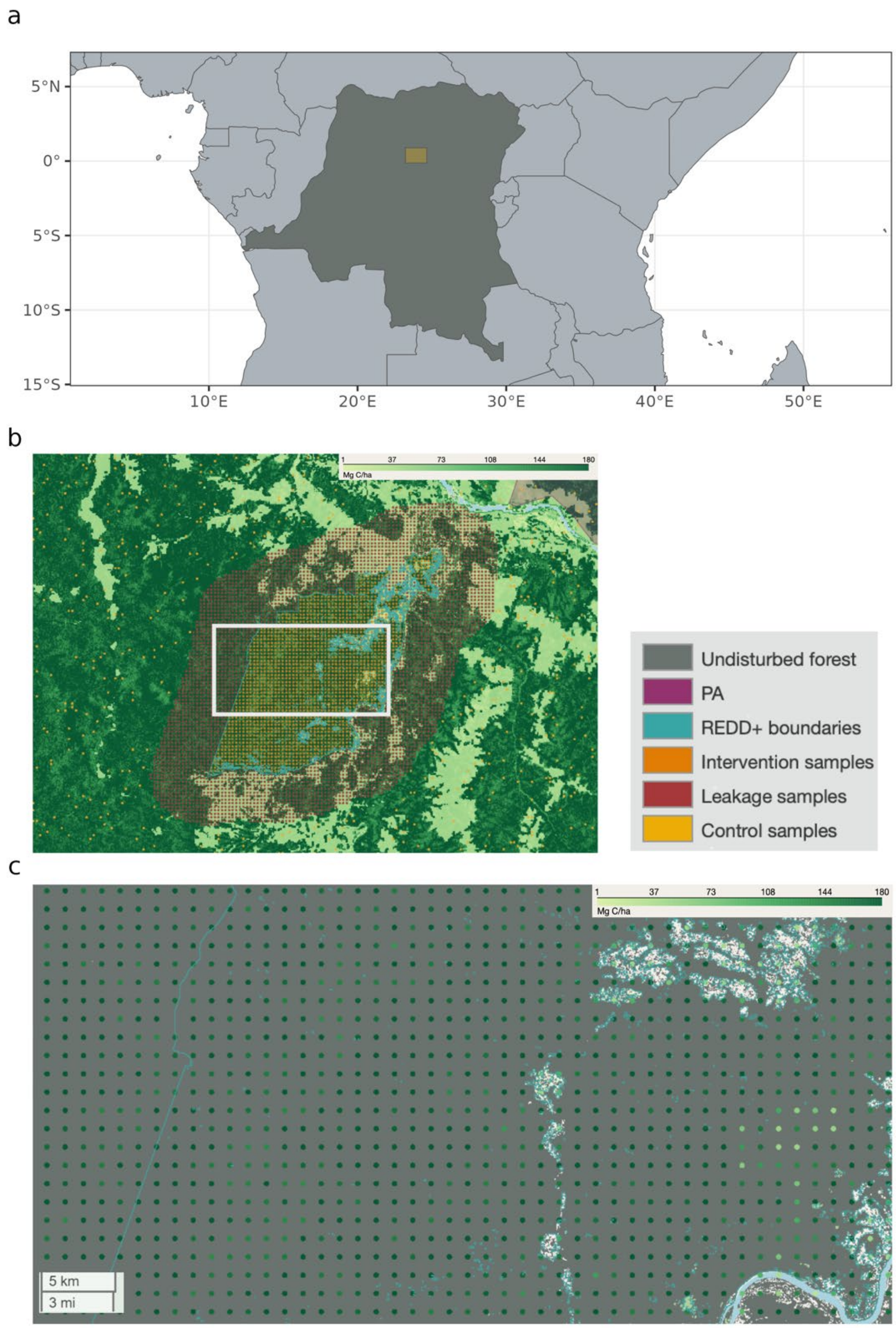

Figure C.1: Schematic of the sampling design. Inner rectangles in **a** and **b** represent the reference regions for panels **b** and **c**, respectively. Features shown in the maps include protected areas (purple), REDD+ area (VCS 1359, cyan), treatment samples (orange), landscape samples (yellow) and leakage samples (red). Biomass as in Soto-Navarro *et al.* 2020 (**b**) and aggregated at the plot level (**c**) shown with reference scale bars (MgC ha^{-1}).

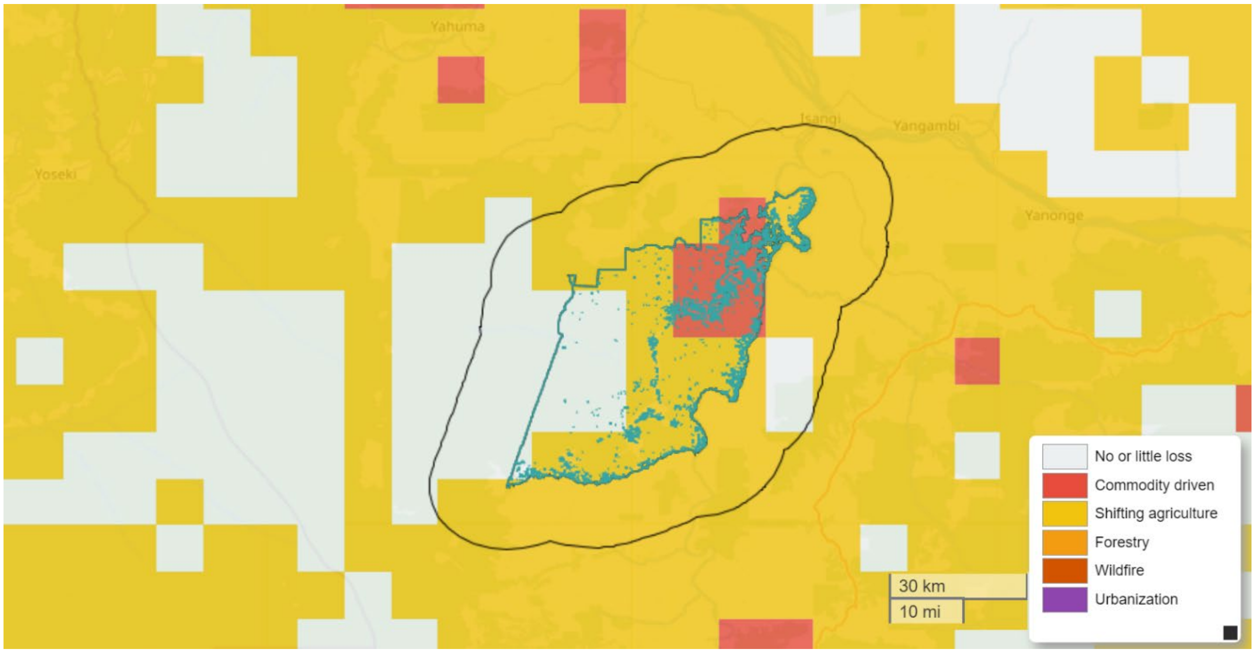

Figure C.2: Analysis of drivers of deforestation. Spatial patterns of drivers of deforestation as represented by Curtis *et al.* 2018. REDD+ boundaries (VCS 1359, green) and leakage belt (black) delineated for reference.

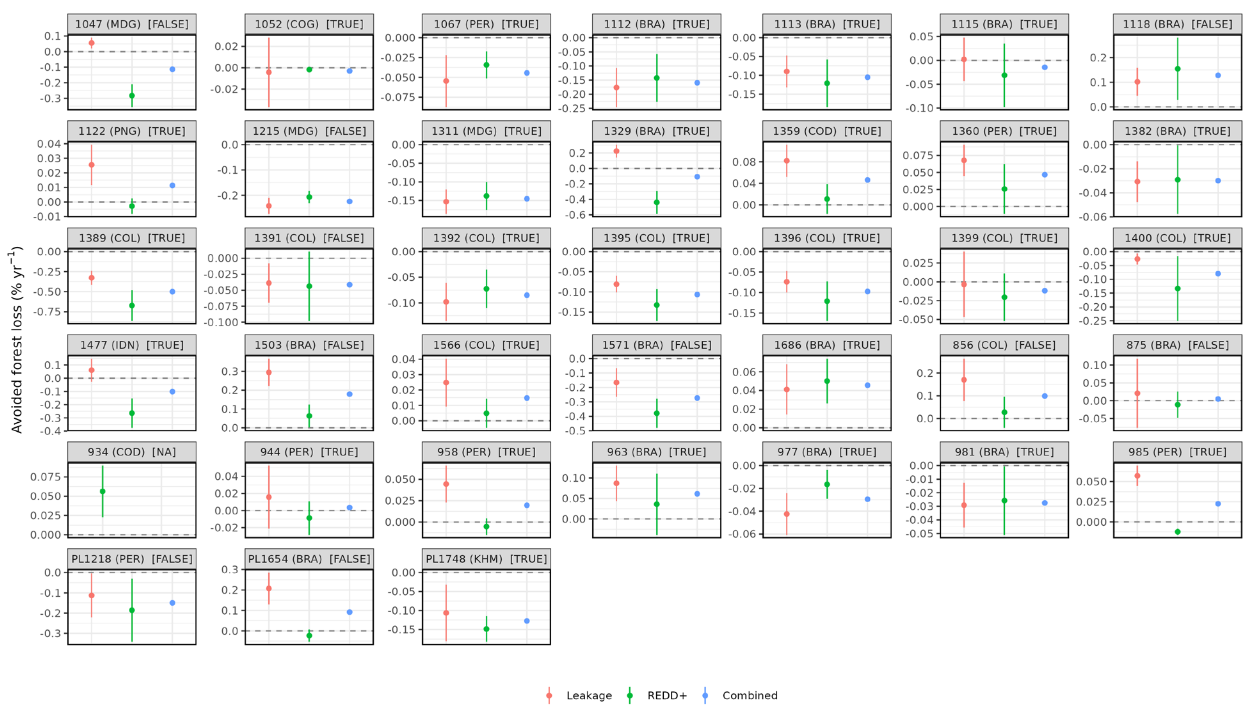

Figure C.3: Differences in annual forest loss rates. Differences in annual rates of deforestation (% yr^{-1}) observed in the 38 matched project sites and their controls (green), and leakage belts and their controls (red), five years after the implementation of REDD+. Error bars are for 95% CI where standard errors were clustered at the treatment-control pair level. The combined effect sizes of the differences in deforestation in rates in REDD+ and leakage belts provided in blue. Banner indicates the project ID followed by the ISO-3 country code and an indicator of whether the matched set in the leakage belt met all the quality criteria.

TABLE C.1: OVERLAP BETWEEN MATCHED PLOTS AND OBSERVATIONS FROM THE GEDI MEAN BIOMASS GRIDDED PRODUCT. THE TABLE SHOWS SEVEN PROJECTS WITH THE HIGHEST GEDI DATA COVERAGE (BUT ALL PROJECTS WERE EXAMINED).

Project iso3	Project ID	Available GEDI observations	Sample plots	Percent of overlap
BRA	981	1210	25593	5%
BRA	977	1453	36195	4%
BRA	1094	480	11986	4%
PER	1360	895	24486	4%
PER	944	1006	28584	4%
BRA	1382	292	8303	4%
BRA	PL1654	324	9216	4%

TABLE C.2: PREDICTORS OF FOREST LOSS

term	estimate	std.error	df	statistic	p.value
Intercept	0.0085	1e-03	104712	8.2844	0.0000
Elevation	-0.0001	1e-04	104712	-1.2295	0.2189
Slope	-0.0001	1e-04	104712	-1.3813	0.1672
Dist. to cities	-0.0032	1e-04	104712	-53.2354	0.0000
Dist. to REDD+	0.0002	1e-04	104712	2.8651	0.0042

TABLE C.3: TOTAL ESTIMATES OF AREA (HA) AND BIOMASS (GGC)

Observational group	Control	Treatment
	Area (Ha)	
REDD+	338940.2	510757.5
Leakage	515185.0	661831.9
	Carbon (GgC)	
REDD+	54289.44	79697.90
Leakage	78151.85	99563.09