Sensitivity of estimates of the effectiveness of REDD+ projects to matching specifications and moving from pixels to polygons as the unit of analysis

Alejandro Guizar-Coutiño[1,2*], David Coomes[1], Tom Swinfield[4], Julia P G Jones[5,6]

[1]Department of Plant Sciences, University of Cambridge.

[2]The Biodiversity Consultancy, Cambridge, UK

[3]Department of Zoology, University of Cambridge

[4]Cambridge Centre for Carbon Credits, University of Cambridge

[5]School of Environmental and Natural Sciences, Bangor University.

[6]Department of Biology, Utrecht University

* Corresponding author: ag997@cam.ac.uk

ABSTRACT

There is a substantial interest in the potential of carbon credits generated by Reducing Emissions from tropical Deforestation and Degradation (REDD+) and traded on the voluntary carbon market for generating the finance needed to slow forest loss. However, such credits have become marred in controversy. Recent global-scale analysis using a range of methods for estimating the counterfactual rate of deforestation *ex post* suggest that many REDD+ projects have overestimated their effectiveness at reducing deforestation and consequently issued more credits than can be justified. All such methods include potentially arbitrary choices which can affect the estimate of the treatment effect. In addition, using pixels as the sampling unit, as some of the studies do, can introduce biases. One study which has been widely cited in the debate (Guizar-Coutino et al. 2022) estimated avoided deforestation using statistical matching of pixels and a single set of matching options. We estimate avoided deforestation from the same set of projects using 7-hectare plots rather than pixels to sample deforestation and explore the sensitivity of the results to matching choices (exploring 120 matched sets in total). We filtered the results on three criteria: 1) post-matching covariate balance, 2) proportion of REDD+ samples that were successfully matched, and 3) similarity of trends in deforestation rates prior to REDD+ implementation (parallel trends). While one of the 44 REDD+ projects failed these quality control process, we estimate treatment effects for the remaining 43 projects. There was a substantial

correlation between our new estimates and those published in 2022 (0.72 measured in annual percent change, and 0.9 measured in total area change) and our headline estimate of 0.22% per yr (95% CI: 0.13 to 0.34) is essentially unchanged. At a time when confidence in the voluntary carbon markets is low, we hope these results provide reassurance that ex-post counterfactual estimates of avoided deforestation are consistent, helping accelerate their widespread adoption and rebuild trust in nature-based climate solutions.

INTRODUCTION

Forests sequester and store enormous amounts of carbon, making forest conservation an important nature-based solution to climate change [1]. REDD+ credits (Reducing Emissions from Deforestation and forest Degradation) are the most numerous of credits traded in the voluntary market [2] and there is very substantial interest in the potential of such credits to provide much-needed finance for tropical forest conservation [3]. However, the extent to which REDD+ projects have provided the reductions in deforestation or forest degradation they have claimed has become hotly debated [4–7].

To sell carbon credits on the voluntary carbon market, a REDD+ project needs to demonstrate additionality: i.e. that any reduction in deforestation is due to the project [8]. This requires an estimate of the counterfactual: how much deforestation would have occurred without the project [4]. REDD+ projects do this by using a variety of methodologies, the majority published and administered by the nonprofit organization VERRA, to estimate 'baseline' deforestation. Baselines are based on historical deforestation in project reference areas which are selected to be as similar as possible to project areas in terms of drivers of deforestation [9]. Deforestation in project areas is then compared with the baseline and, following verification, VCUs (voluntary carbon units) are issued which can then be traded on the voluntary market. This approach has advantages as allows a project to have some certainty up-front as to the number of credits a project can produce, however it has been criticized for putting a lot of weight on the ex-ante baseline estimates, rather than using information on real rates of deforestation in comparable areas during the project period to construct counterfactuals [9,10].

A suite of independent global-scale analyses using a variety of approaches have suggested that many REDD+ projects provide limited additional avoided deforestation. For example, Guizar-Coutiño et al. 2022 found that 33 out of 40 projects studied did not avoid any deforestation. Even more concerning, there have been claims that projects have issued more voluntary carbon units than can be justified[10,12]. For example West et al.[10], found that across 18 REDD+ projects for which they had data, projects had

been used to offset almost three times more carbon emissions than their contribution to climate change mitigation, though these estimates are disputed [7].

While many analyses of the impact of conservation interventions on deforestation uses pixels as the unit of analysis (e.g. [13–15]) there is increasing recognition that analyzing deforestation as binary, unrepeatable data (a pixel is deforested or not) can result in biases which are reduced by aggregating binary pixels spatially [16]. Where appropriate polygons exist, the unit of analysis can be pixels aggregated within land holdings [17], or administrative boundaries [18,19]. However in many contexts appropriate units may not exist, or polygons reflecting such units may not be available. In such cases polygons of a given area placed at random can be used to sample the landscape [10].

Given that forest conservation projects, including REDD+, are never randomly placed in a landscape, evaluating their impact on deforestation requires causal inference from observational data [20]. Areas where REDD+ projects are set up (treatment units) will inevitably differ in terms of confounders which influence both exposure to the intervention (in this case the location of REDD+ projects) and the outcome of interest (deforestation). Statistical matching is often used to ensure treatment and selected control units are as similar as possible with respect to these observable confounders [21,22]. However, there are a range of somewhat arbitrary modelling choices made when applying statistical matching which can influence the estimated treatment effect [23,24].

In this paper we revisit the same set of REDD+ projects as were presented in Guizar-Coutiño et al. 2022. While Guizar-Coutiño et al. 2022 sample deforestation using pixels, we instead sample deforestation using circular polygons of 150 m radius, or 7.1-ha plots. We also explore how the choices made during statistical matching (applying 120 different combinations of matching options to each project) influence the estimated treatment effect.

METHODS

The research pipeline is shown in Fig. 1 and developed in the following sections. Our outcome of interest is deforestation as reported in the Tropical Moist Forest (TMF) maps which present deforestation and degradation across the humid tropics at 30 m resolution from 1990 to 2019 [25].

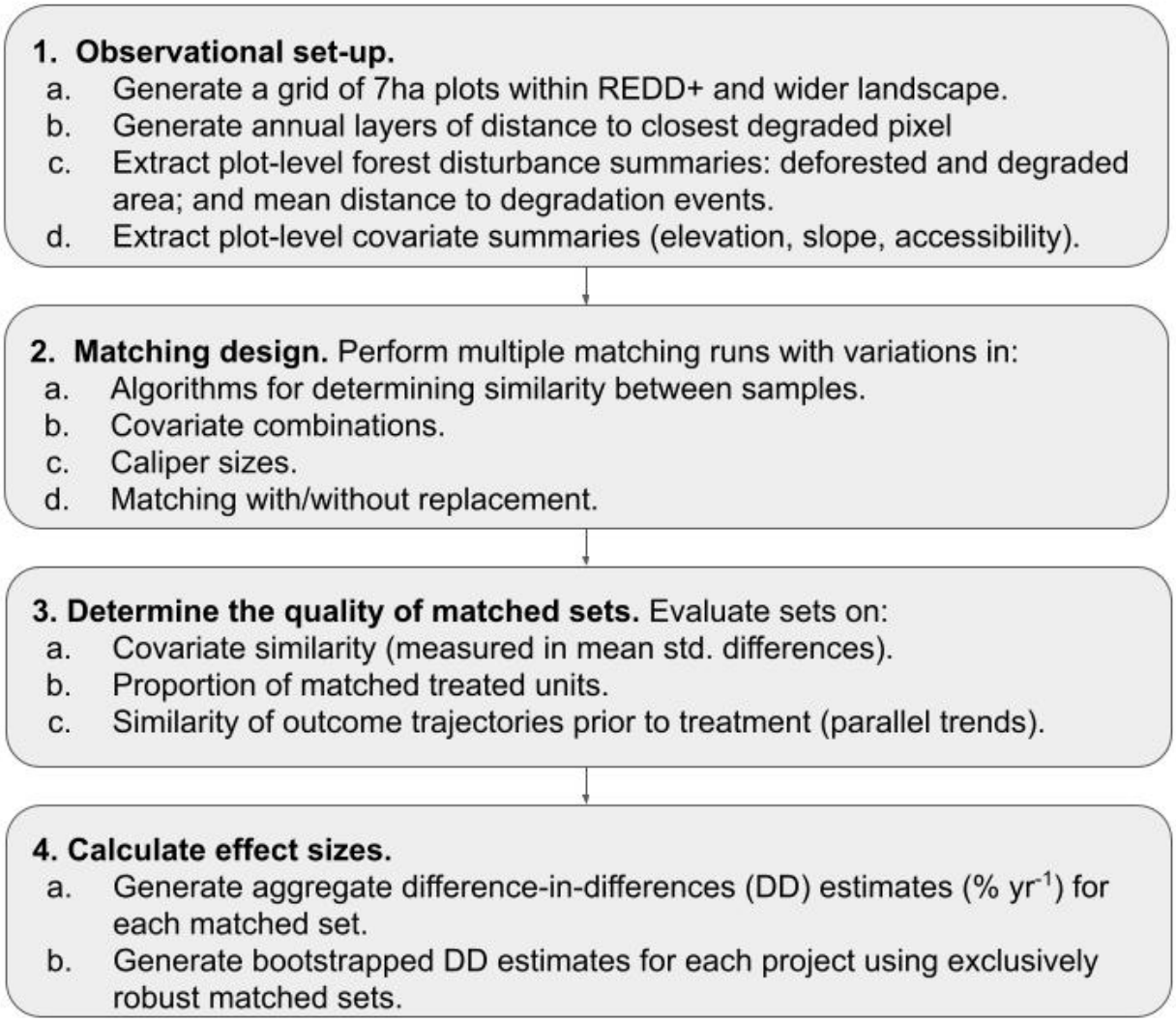

Figure 1. Schematic of the pipeline use to explore the impact of matching decisions on estimates of REDD+ effectiveness. Blocks indicate components of the analysis and are described in more detail in the methods.

1. Observational set-up

Selection of REDD+ projects: The VCS database is the leading accreditation registry for voluntary REDD+ projects [2] and the largest sources of data on voluntary REDD+ projects locations. We used the REDD+ database compiled by Guizar-Coutiño[1] [11] which identified 81 "Reducing Deforestation and Degradation" projects established in the tropics [26] and successfully found boundary data for 71 of these. While new projects have been introduced since 2019, and others withdrawn, for consistency with [11] we analyze the same set of projects. We excluded projects with <80% evergreen forest cover at the start (because the TMF dataset only covers evergreen forest), and projects which operated for fewer than 5 years or commenced before 2000. In Guizar-Coutino 2022 we included project PL1748 in the analyses despite

not meeting the minimum number of years for our criteria. This is because we recorded its start date incorrectly during our original analysis. We include it here for consistency despite it only running for four years.

Sampling within REDD+ areas and forested landscapes. To assess the extent of reductions in deforestation due to REDD+, we compared changes in forest cover within treated areas with those taking place in similar forest patches within the wider landscape. We generated plots with a radius of 150 m on a 1 km grid within forested areas of the REDD+ projects and surrounding regions. Each plot was 7.1 hectares in size and contained up to 78 TMF pixels, which we considered adequate for characterising local drivers of deforestation and our outcome as continuous variables [27]. We used the extent of forest cover in 1990 as delineated by the TMF [25] to position sampling plots, ensuring that all samples contained at least some portion of undisturbed moist forest cover when building the time-series of forest cover. We used the same approach to sample tropical moist forests plots in the wider landscape: we removed sampling locations that were not forested or that lay within a 15-km buffer zone around the project area, which may have been affected by local leakage of economic activities. We then random subsampled, until we had up to five samples in the wider landscape for each sampling plot located inside the REDD+ project, each separated by at least 1 km to reduce the spatial dependence [28]; this created a pool of samples which was subsequently used to find matches for REDD+ project samples (Fig. 2).

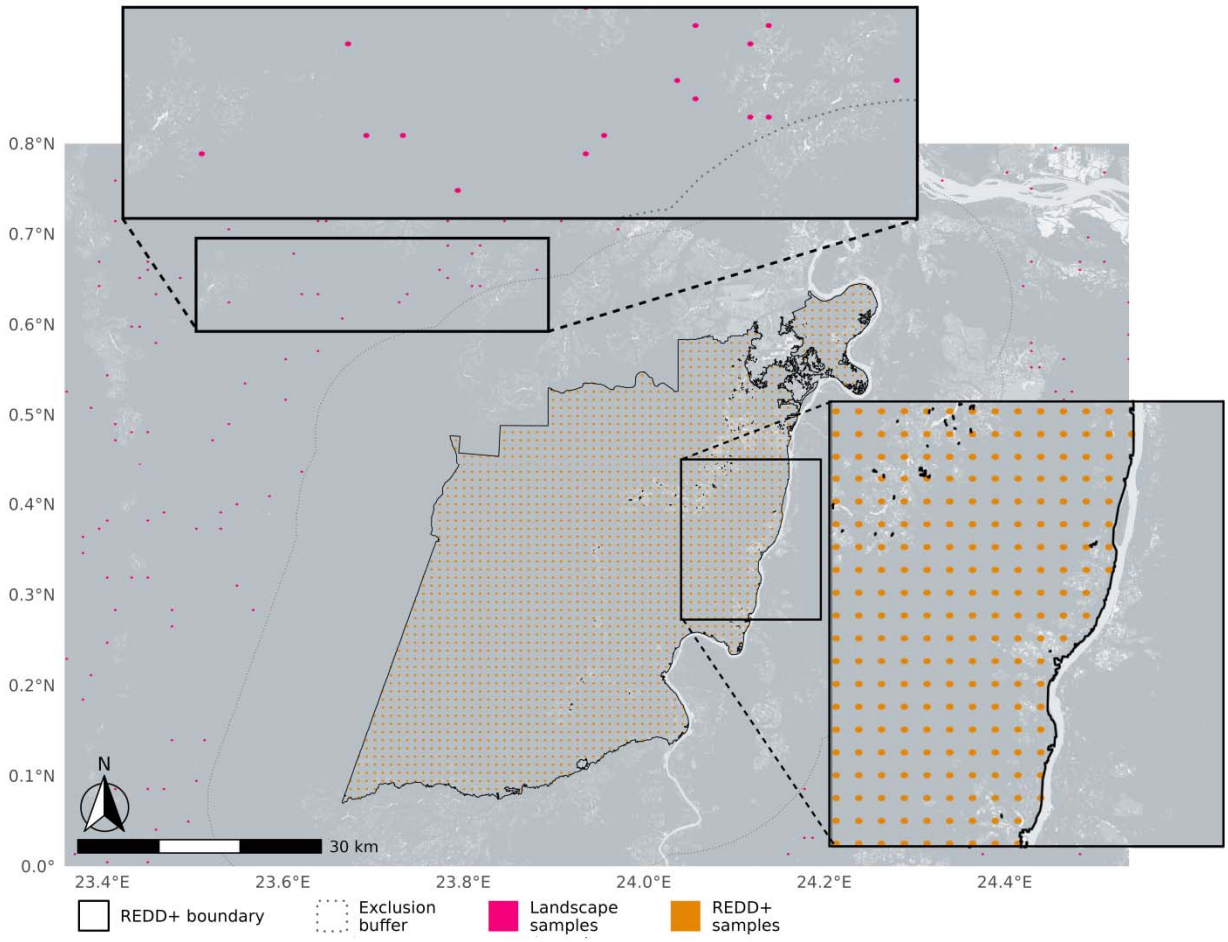

Figure 2. Illustration of the sampling strategy, for a REDD+ project in Democratic Republic of the Congo (VCS 1359). Features shown include the boundary of the REDD+ area (solid boundary) and the 7.1-ha plots collected from treatment (orange) and landscape (pink) areas. A buffer surrounding the project boundaries (dotted boundary) is shown within which we did not potential controls. The extent of undisturbed forest cover (dark grey) was used as a reference for distributing sample plots.

Plot-level forest disturbance estimates. Annual estimates of forest cover and forest disturbances were extracted from the TMF database. This database provides a long-term characterisation of forest disturbances and land cover change, 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, characterised as short-term disturbance (less than 2.5 years); and the deforested class, representing a long-term removal of forest cover [25]. For each year during the 1990 - 2019 period, we calculated the extent of undisturbed, and disturbed forests classes within sampling plots. From these metrics we derived our outcome of interest - the proportion of forest lost - defined as ,

where Δp_t is the total area deforested in year t and p_d is the area of undisturbed forest in the baseline or reference year (see 4. Exploring the impact of matching choices on estimated treatment effect)

Drivers of deforestation: We identified key covariates associated with deforestation [29], some of which were also likely to have been associated with assignment to the treatment [30,31]. To characterise deforestation risks factors, we collected data on key and observable bioclimatic and socio-demographic determinants of deforestation (as in [29] Table S1). For each circular plot, we obtained mean estimates for elevation and slope [32], distance to the nearest urban center [33]and distance to forest edge. 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 time-series. First, for each pixel covered by the circular plot, 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 design

We used statistical matching to select a subset of samples in the wider landscape that were as similar as possible to samples in the project areas in terms of observable confounders: initial forest cover density, exposure to drivers of deforestation and trajectories of deforestation in the period prior to REDD+ (Table S1). We exact matched on country. We used a "base" set of covariates for all runs, which included: 1) elevation [32], 2) slope [32], 3) distance to population centers [33], 4) distance to degraded areas over the five years prior to project commencement, and 5) extent of degraded area at the time of project implementation.

To evaluate the sensitivity of estimates of avoided deforestation to matching specifications [24], we ran matching runs with alternative specifications of:

Algorithms: We used propensity score (PSM), Mahalanobis distance (MHN) and random forest (RFM) algorithms;

Covariates: We first used the "base" set of covariates alone, then iteratively added the following covariates to that base set: extent of undisturbed forest in 1990 (forest area 1990), extent of undisturbed forest in the project implementation year (forest area), and extent of undisturbed forest in

1990 and implementation year (forest area + forest area 1990); resulting in 4 distinct covariate combinations.

Calipers: We applied calipers of different sizes to constrain the selection of landscape observations prior to matching, so that the covariate distribution lay within 0.1, 0.3, 0.5, 0.7 and 0.9 SD of the covariate distribution in the treatment group.

Matching with and without replacement: Matching iterations allowed for both, matching with replacement (e.g. allowing control observations to be paired with multiple treatment units), or without replacement (e.g. constraining one control observation per treatment unit).

The combination of matching criteria resulted in 120 matching runs for each of the 44 selected REDD+ sites, giving a total of 5280 runs.

3. Assessing the quality of matched sets

Matching runs were assessed against three criteria:

Covariate balance between treated and control samples. Similarity between treated and control samples was determined by the overlap of the distributions of the covariate values. Matching runs were considered adequate if the absolute standardised mean difference between treated and matched control samples was less than 0.25 across all covariates included in the matching trial[22].

Proportion of project-area samples that were matched. Estimates of additionality can be biased unless a high proportion of treatment samples are successfully matched: only then is the "average treatment effect on the treated" estimated accurately[22]. Matching was considered adequate if at least 80% of treatment samples were paired with a control.

Parallel trends. A key condition of difference-in-differences designs is that inferences cannot be drawn unless treatment and control groups follow similar outcome trajectories in the period prior to the intervention; this is known as the parallel trends condition [34]. Since our sampling strategy involves characterising historical deforestation within the sampling plots, counterfactuals would be credible only if deforestation trends between treated and control plots prior to the introduction of REDD+ are comparable. We established the existence of parallel trends by ruling out significant differences in rates of forest loss prior to project implementation (exemplified in Fig. S1). To assess parallel trends, a linear model was fitted for each project:

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

Where P_{it} is the proportion of undisturbed forest area in the plot *i* in year *t*, and CI is the treatment group indicator (*i* =1 for treated or 0 for control), relative to forest extent in 1990. Parameter β_3 allows us to test for significant differences in deforestation trends between treated and control plots. We performed these tests using all available years before the project starting date. Departure from the parallel trends was assessed by testing whether the slope of the treatment and control groups were significantly different at $p < 0.05$; and matching was considered adequate if no departure was detected.

4. Calculating effect sizes.

We used the same simple difference-in-difference model we used to estimate the treatment effect in Guizar et al., 2022 [11] but explored how this varied with different matched sets.

Matched model effect sizes. To estimate additionality for each matched set, we calculated annual forest loss rates by aggregating the extent of deforested area across all sampling plots for each group (e.g. treatment or control) in the five years before and after project implementation, divided by the extent of undisturbed forest across all sampling plots in each group in the 6th year prior to project implementation (i.e., reference year). We then estimated a mean difference-in-differences in forest loss rates as: DD = (AT + AC) – (BT + BC), where DD is difference-in-differences, A and B refer to the before and after period and T and C refer to the treatment and control group rates of deforestation rates (% yr^{-1}).

Bootstrapped effect sizes: We computed project-level mean difference-in-difference with 95% confidence intervals by bootstrapping difference-in-difference estimates of robustly matched sets. 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

Sites included in the analysis

Of the 71 mapped REDD+ projects, 22 had less than 80% evergreen forest cover and 5 did not meet our time frame criteria. We applied matching to the remaining 44 projects (Fig. 3). Only 22% of all matched sets met all three of our quality-control criteria (1142 of 5280 matched sets). One of the 44 REDD+

projects failed to produce any high-quality matched set (project 1201). The remaining 43 REDD+ sites were located in America (n=34), Africa (n=6) and Asia-Pacific (n=3).

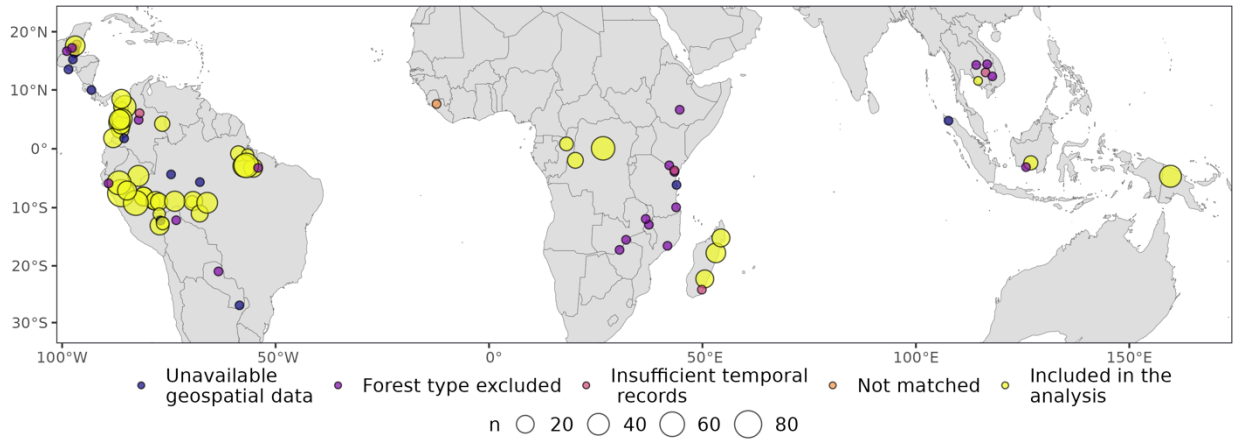

Figure 3. Location of the REDD+ projects included in the analysis. Of the 71 mapped sites, 22 had less than 80% evergreen forest cover at the start of projects (purple), 5 did not meet our time frame criteria, leaving 44 projects that were examined with multiple matching approaches (34 in the Americas, 7 in Africa and 3 in Asia-Pacific). Successful matching models were returned for 43 projects (yellow), with plots size indicate the number of robust models available for each of them. Project PL1748 did not met the time frame criteria but was included in the analyses for consistency with Guizar-Coutino et al. 2022.

Matching parameters and quality of matched sets

The quality of matched sets was more influenced by the choice of algorithms and by covariate selection, than by caliper size or matching with or without replacement. We note varying success rates across the three matching quality criteria (Fig. S2): 70% of sets failed the parallel trends criteria (i.e. they did not display similar outcome trajectories before the REDD+ project was established; Fig. S1), while 54% failed to achieve covariate balance between treated and control observations, and 31% failed to meet the criteria that 80% of treatment plots were successfully matched. Matched sets using the Mahalanobis distance method were much more likely to pass the three matching quality tests than sets matched with other methods (Fig. 4). Moreover, including measures of forest area on different time periods alongside the base covariates increased the quality of matched sets. Neither the caliper size nor the choice of whether to match with replacement had much effect on whether a matched set passed the quality control tests.

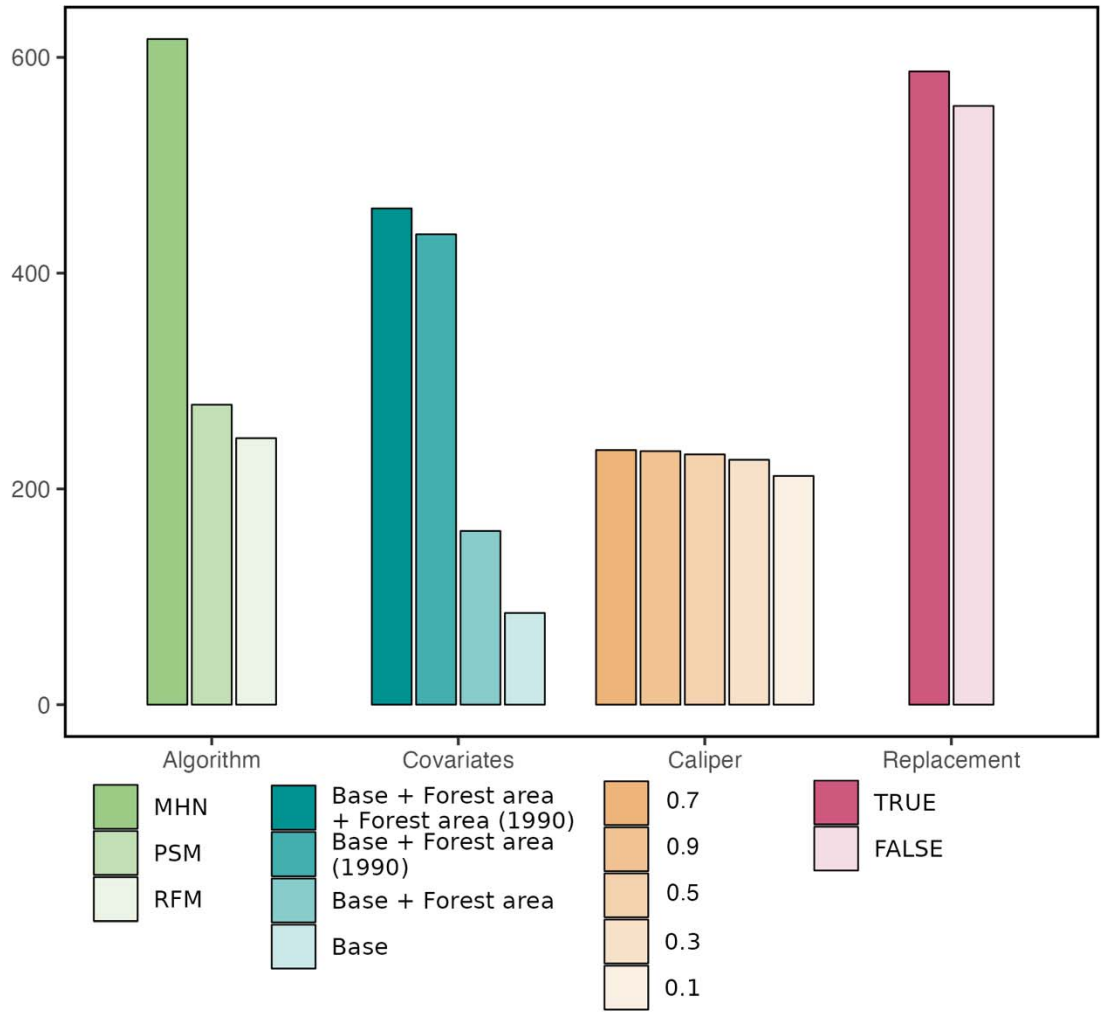

Figure 4. The number of matched sets that passed all three quality control criteria, subdivided by specifications made when matching. A total of 120 alternative matching specifications were made in each of 44 REDD+ sites, giving a total of 5280 matching sets; of these 1142 met all three quality control criteria. The specifications included: a) four choice of covariates combinations; base covariates only (i.e. elevation, slope, accessibility, distance to degradation, country and biome), base covariates plus forest cover at the time of project implementation, base covariates plus forest cover in 1990, and base covariates + forest cover in 1990 and implementation year; b) choice of matching algorithm between Mahalanobis (MHN), propensity score (PSM), and random forest (RFM); c) choice of filtering regime (caliper sizes 0.1, 0.3, 0.5, 0.7 and 0.9 SD); and d) choice of whether to allow matching with replacement or not.

Revised estimates of effectiveness of REDD+ in reducing deforestation

Our analyses highlights the variability in the avoided deforestation delivered by REDD+ projects over their first five years of operation, both in terms of relative reductions in forest loss (Fig. 5a), and absolute deforestation avoided (Fig. 5b). While some projects have delivered substantial avoided deforestation (those to the left hand side of the Figure), the majority have effect sizes close to zero. Our

project-level difference-in-differences estimates of robust matched sets were well correlated with the estimates produced by pixel-matching in Guizar-Coutiño et al., 2024 [11]. The correlation is 0.72 when measured in annual percent change, and 0.9 when measured in total area change.

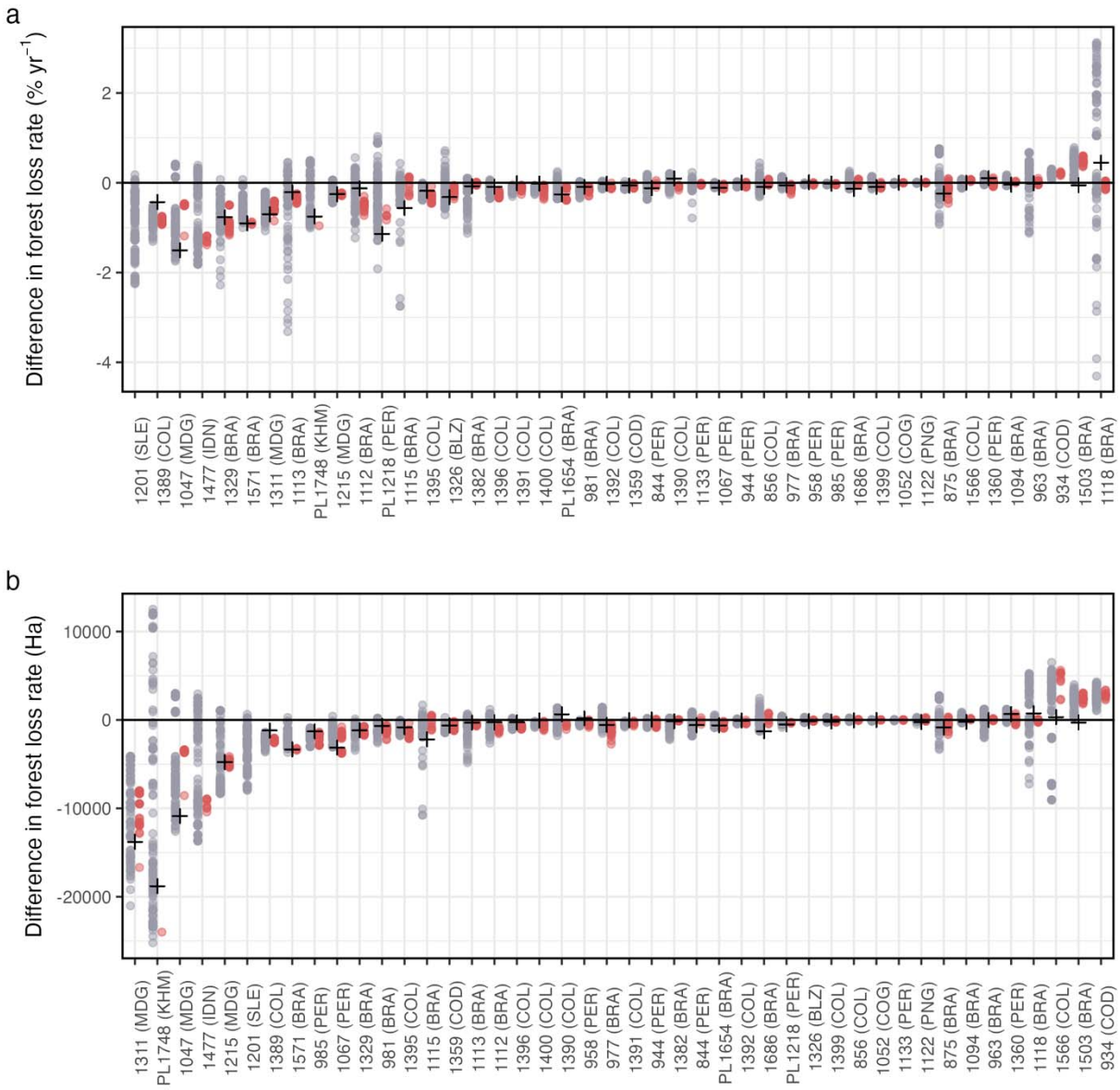

Figure 5. Annual differences in (a) deforestation rates for the 44 REDD+ projects included in this study and (b) total area loss using all matched models, with estimates below the zero line indicating avoided deforestation. Dots indicate difference-in-differences estimates coloured according to whether the matching met our quality control criteria (red=robustly matched,

grey=rejected). There are no valid models for project 1201 in Sierra Leone. For projects included in [11], estimates from that analysis are indicated with a '+' symbol.

Table S2 presents the bootstrapped estimates of the deforestation avoided by each project (both in terms of percentage per year and hectares per year) with 95% confidence intervals. For 30 of the 43 projects, our best estimate of avoided deforestation as a percentage and total ha is negative and does not overlap zero (Table S2). However, for many of those projects the reduction in deforestation is modest.

Combining the 43 site-level estimates, we observe a mean reduction of 0.22% yr^{-1} (95% CI: 0.13–0.36) in the rates of forest loss compared to controls, which amounted to an approximate area of avoided deforestation of 64903 Ha. This is essentially unchanged from the figures presented in Guizar-Coutiño et al., 2024 which similarly calculated a mean reduction of 0.22%/year (95% CI: 0.13–0.34) amounting to approximately 66754 ha of avoided deforestation.

DISCUSSION

Given heated debate about the extent to which REDD+ projects have effectively avoided deforestation [4,7,10,12,35], the inevitable uncertainty in estimates using a single matching approach [22,24], and concerns about analyzing deforestation as a binary process [16], revisiting and updating the estimates of avoided deforestation provided in Guizar-Coutiño et al. (2022) is valuable. We find that most projects reduce deforestation. However effectiveness is highly variable; for many projects the reductions in deforestation are modest and for some they are non-existent. These revised estimates add to the growing evidence base exploring the impacts of site-based REDD+ project effectiveness using ex post counterfactuals [10–12,36].

Our previous study [11] used ~30m pixels to sample deforestation. Our new approach, using 7.1-ha plots all separated by at least 1 km, has the advantage of reducing the spatial dependence of samples [28]. By using plots as our sampling unit (within which deforestation is a continuous variable) rather than pixels (where deforestation is a binary outcome), we were also able to select controls with similar trends in deforestation. There are inevitably a range of somewhat arbitrary choices which need to be made when selecting matching approaches [21,22]. By trialing 120 combinations of matching options and estimating the treatment effect from the full set which meet quality-control criteria, we get a valuable estimate of the uncertainty in estimates of ex post estimates of avoided deforestation.

This work explores sensitivity to how matching on observed confounders was carried out. However, there is a risk that important confounders remain which cannot be accounted for [37]. Despite the importance of testing for sensitivity of results to unobserved confounding, this is not yet standard practice in conservation impact evaluation [38]. In the context of REDD+, hidden confounders might include institutional factors such as tenure security and presence of local conservation initiatives [39–41] or local drivers of agricultural activities [29]. Some of the critiques of the recent published estimates of the avoided deforestation from REDD+ projects [7,42,43] suggest that such large-scale analysis (covering multiple REDD+ projects in multiple sites using the same methodology) inevitably fails to account for key local-level drivers of deforestation. Further work exploring the sensitivity of these results to hidden confounders would be valuable.

Climate change is advancing faster than many climate scientists had predicted. The target of keeping warming within 1.5 degrees of pre-industrial levels has already been missed [44], or will be soon [45]. Given the contribution of tropical forests to both sequestering and storing carbon[46,47], it is vital that the global community find ways to stop further loss of tropical forests [4]. The voluntary carbon market is not the only way that forest conservation can be funded. International policy is focused on developing a UN-governed schemed under article 6 of the Paris Climate Agreement which will focus on lowering deforestation over entire jurisdictions [40] rather than in specific REDD+ project locations. There are also many bilateral agreements where countries support other countries to lower their deforestation-related emissions.

However, selling emissions reductions from avoiding deforestation on the voluntary carbon market is generating substantial, and hard-to-replace, finance for forest conservation and will remain important at least into the medium term. The lack of credibility of the credits being issued poses a threat to this important revenue stream [35,48]. There is growing recognition that conservative credits issuance [49] can help address issues such as impermanence [50], and leakage [51]. At time when confidence in the voluntary carbon markets is low, we hope our results provide reassurance that *ex-post* counterfactual estimates of avoided deforestation can deliver relatively consistent estimates of additionality, helping accelerate their widespread adoption and rebuild trust in nature-based climate solutions.

ACKNOWLEDGMENTS

A.G.C. received a scholarship from Consejo Nacional de Ciencia y Tecnología México, CONACYT, and the Frank Jackson Foundation through a grant to Wolfson College, as well as funding from the Tezos

Foundation gift to the Cambridge Centre for Carbon Credits (4C). J.P.G.J. was an Oxford Martin School visiting fellow and short-term visiting fellow at Jesus College Oxford when working on this paper.

DATA AND CODE

Data summaries and code to reproduce the main figures provided at: https://github.com/guizar/redd-sens

LITERATURE CITED

1. Griscom, B. W. *et al.* Natural climate solutions. *Proc Natl Acad Sci U S A* **114**, 11645–11650 (2017).

2. Donofrio, S., Patrick–Daley, C. & Ciro–Lin, K. The art of integrity: State of the voluntary carbon markets 2022 Q3. *Forest Trends' Ecosystem Marketplace* (2022).

3. TSVCM. *Taskforce on Scaling Voluntary Carbon Markets*. www.iif.com/Portals/1/Files/TSVCM_Report.pdf (2021).

4. Jones, J. P. G. & Lewis, S. L. Forest carbon offsets are failing. *Science (1979)* **381**, 830–831 (2023).

5. Greenfield, P. Revealed: more than 90% of rainforest carbon offsets by biggest certifier are worthless, analysis shows. *The Guardian* (2023).

6. Jones, J. P. G. Scandal in the voluntary carbon market must not impede tropical forest conservation. *Nat Ecol Evol* **in press**, (2024).

7. Mitchard, E. *et al.* Serious errors impair an assessment of forest carbon projects: A rebuttal of West et al. (2023). *SSRN Electronic Journal* (2023) doi:10.2139/SSRN.4661873.

8. Delacote, P., Le Velly, G. & Simonet, G. Revisiting the location bias and additionality of REDD+ projects: the role of project proponents status and certification. *Resour Energy Econ* **67**, 101277 (2022).

9. West, T. A. P., Bomfim, B. & Haya, B. K. Baselines. in *Quality assessment of REDD+ carbon credit projects* 28–50 (Berkley Carbon Trading Project, 2023).

10. West, T. A. P. *et al.* Action needed to make carbon offsets from forest conservation work for climate change mitigation. *Science (1979)* **381**, 873–877 (2023).

11. Guizar-Coutiño, A., Jones, J. P. G., Balmford, A., Carmenta, R. & Coomes, D. A. A global evaluation of the effectiveness of voluntary REDD+ projects at reducing deforestation and degradation in the moist tropics. *Conservation Biology* **n/a**, e13970 (2022).

12. West, T. A. P., Börner, J., Sills, E. O. & Kontoleon, A. Overstated carbon emission reductions from voluntary REDD+ projects in the Brazilian Amazon. *Proc Natl Acad Sci U S A* **117**, 24188–24194 (2020).

13. Baehr, C., Benyishay, A. & Parks, B. Linking Local Infrastructure Development and Deforestation: Evidence from Satellite and Administrative Data. *https://doi.org/10.1086/712800* **8**, 375–409 (2021).

14. Anderson, C. M., Asner, G. P., Llactayo, W. & Lambin, E. F. Overlapping land allocations reduce deforestation in Peru. *Land use policy* **79**, 174–178 (2018).

15. Alix-Garcia, J. & Gibbs, H. K. Forest conservation effects of Brazil's zero deforestation cattle agreements undermined by leakage. *Global Environmental Change* **47**, 201–217 (2017).

16. Garcia, A. & Heilmayr, R. Impact evaluation with nonrepeatable outcomes: The case of forest conservation. *J Environ Econ Manage* **125**, 102971 (2024).

17. Carlson, K. M. *et al.* Effect of oil palm sustainability certification on deforestation and fire in Indonesia. *Proc Natl Acad Sci U S A* **115**, 121–126 (2018).

18. Tabor, K. *et al.* Evaluating the effectiveness of conservation and development investments in reducing deforestation and fires in Ankeniheny-Zahemena Corridor, Madagascar. *PLoS One* **12**, e0190119 (2017).

19. Koch, N., Zu Ermgassen, E. K. H. J., Wehkamp, J., Oliveira Filho, F. J. B. & Schwerhoff, G. Agricultural Productivity and Forest Conservation: Evidence from the Brazilian Amazon. *Am J Agric Econ* **101**, 919–940 (2019).

20. Börner, J., Schulz, D., Wunder, S. & Pfaff, A. The effectiveness of forest conservation policies and programs. **12**, 45–64 (2020).

21. Schleicher, J. *et al.* Statistical matching for conservation science. *Conservation Biology* **34**, 538–549 (2019).

22. Stuart, E. A. Matching Methods for Causal Inference: A Review and a Look Forward. *Statistical Science* **25**, 1–21 (2010).

23. Devenish, K., Desbureaux, S., Willcock, S. & Jones, J. P. G. On track to achieve no net loss of forest at Madagascar's biggest mine. *Nature Sustainability 2022 5:6* **5**, 498–508 (2022).

24. Desbureaux, S. Subjective modeling choices and the robustness of impact evaluations in conservation science. *Conservation Biology* **35**, 1615–1626 (2021).

25. Vancutsem, C. *et al.* Long-term (1990–2019) monitoring of forest cover changes in the humid tropics. *Sci Adv* **7**, eabe1603 (2021).

26. Verified Carbon Standard (VCS). VCS Project Database. *http://www.vcsprojectdatabase.org/* (2019).

27. Avelino, A. F. T., Baylis, K. & Honey-Rosés, J. Goldilocks and the Raster Grid: Selecting Scale when Evaluating Conservation Programs. *PLoS One* **11**, e0167945 (2016).

28. Robalino, J. A. & Pfaff, A. Contagious development: Neighbor interactions in deforestation. *J Dev Econ* **97**, 427–436 (2012).

29. Busch, J. & Ferretti-Gallon, K. What Drives and Stops Deforestation, Reforestation, and Forest Degradation? An Updated Meta-analysis. *Rev Environ Econ Policy* 0 (2023) doi:10.1086/725051.

30. Lin, L., Pattanayak, S. K., Sills, E. O. & Sunderlin, W. D. Site selection for forest carbon projects. in *Analysing REDD+: Challenges and choices* 209–230 (CIFOR, Bogor, Indonesia, 2012).

31. Sunderlin, W. D. & Sills, E. O. REDD+ projects as a hybrid of old and new forest conservation approaches. in *Analysing REDD+: Challenges and choices* 177–191 (CIFOR, Bogor, Indonesia, 2012).

32. Jarvis, A., Reuter, H. I., Nelson, A. & Guevara, E. *Hole-Filled SRTM for the Globe. V4*. http://srtm.csi.cgiar.org (2008).

33. Weiss, D. J. *et al.* A global map of travel time to cities to assess inequalities in accessibility in 2015. *Nature* **553**, 333–336 (2018).

34. Angrist, J. D. & Pischke, J.-S. *Mostly Harmless Econometrics: An Empiricist's Companion*. (Princeton University Press, 2008).

35. Balmford, A. *et al.* Credit credibility threatens forests. *Science (1979)* **380**, 466–467 (2023).

36. Correa, J. *et al.* Evaluating REDD+ at subnational level: Amazon fund impacts in Alta Floresta, Brazil. *For Policy Econ* **116**, 102178 (2020).

37. Cinelli, C. & Hazlett, C. Making sense of sensitivity: Extending omitted variable bias. *J R Stat Soc Series B Stat Methodol* **82**, 39–67 (2020).

38. Jones, J. P. G. *et al.* Quantifying uncertainty about how interventions are assigned would improve impact evaluation in conservation: Reply to Rasolofoson 2022. *Conservation Biology* **n/a**, e14007 (2022).

39. Lin, L., Pattanayak, S. K., Sills, E. O. & Sunderlin, W. D. Site selection for forest carbon projects. in *Analysing REDD+: Challenges and choices* 209–230 (CIFOR, Bogor, Indonesia, 2012).

40. Wunder, S. *et al.* REDD+ in Theory and Practice: How Lessons From Local Projects Can Inform Jurisdictional Approaches. **3**, (2020).

41. Usmani, F., Jeuland, M. & Pattanayak, S. K. NGOs and the Effectiveness of Interventions. *Rev Econ Stat* 1–45 (2022) doi:10.1162/rest_a_01217.

42. VERRA. Technical Review of West et al. 2020 and 2023, Guizar-Coutiño 2022, and Coverage in Britain's Guardian. *Verra* Preprint at https://verra.org/technical-review-of-west-et-al-2020-and-2023-guizar-coutino-2022-and-coverage-in-britains-guardian/ (2023).

43. VERRA. non-dogmatic critique of West et al. (2023), "Action needed to make carbon offsets from forest conservation work for climate change mitigation," Science 381: 873-877. (2023).

44. McCulloch, M. T., Winter, A., Sherman, C. E. & Trotter, J. A. 300 years of sclerosponge thermometry shows global warming has exceeded 1.5°C. *Nature Climate Change 2024 14:2* **14**, 171–177 (2024).

45. Diffenbaugh, N. S. & Barnes, E. A. Data-driven predictions of the time remaining until critical global warming thresholds are reached. *Proc Natl Acad Sci U S A* **120**, e2207183120 (2023).

46. Hubau, W. *et al.* Asynchronous carbon sink saturation in African and Amazonian tropical forests. *Nature* **579**, 80–87 (2020).

47. IPCC AR6 Working Group 1: Summary for Policymakers | Climate Change 2021: The Physical Science Basis. https://www.ipcc.ch/report/ar6/wg1/chapter/summary-for-policymakers/.

48. Delacote, P. *et al.* Strong transparency required for carbon credit mechanisms. *Nature Sustainability 2024* 1–8 (2024) doi:10.1038/s41893-024-01310-0.

49. Swinfield, T. *et al.* Nature-based credit markets at a crossroads. (2023) doi:10.33774/COE-2023-PL9XV-V2.

50. Balmford, A. *et al.* Realizing the social value of impermanent carbon credits. *Nature Climate Change 2023 13:11* **13**, 1172–1178 (2023).

51. Filewod, B. & McCarney, G. Avoiding carbon leakage from nature-based offsets by design. *One Earth* **6**, 790–802 (2023).

SUPPLEMENTARY INFORMATION

SUPPLEMENTARY FIGURES

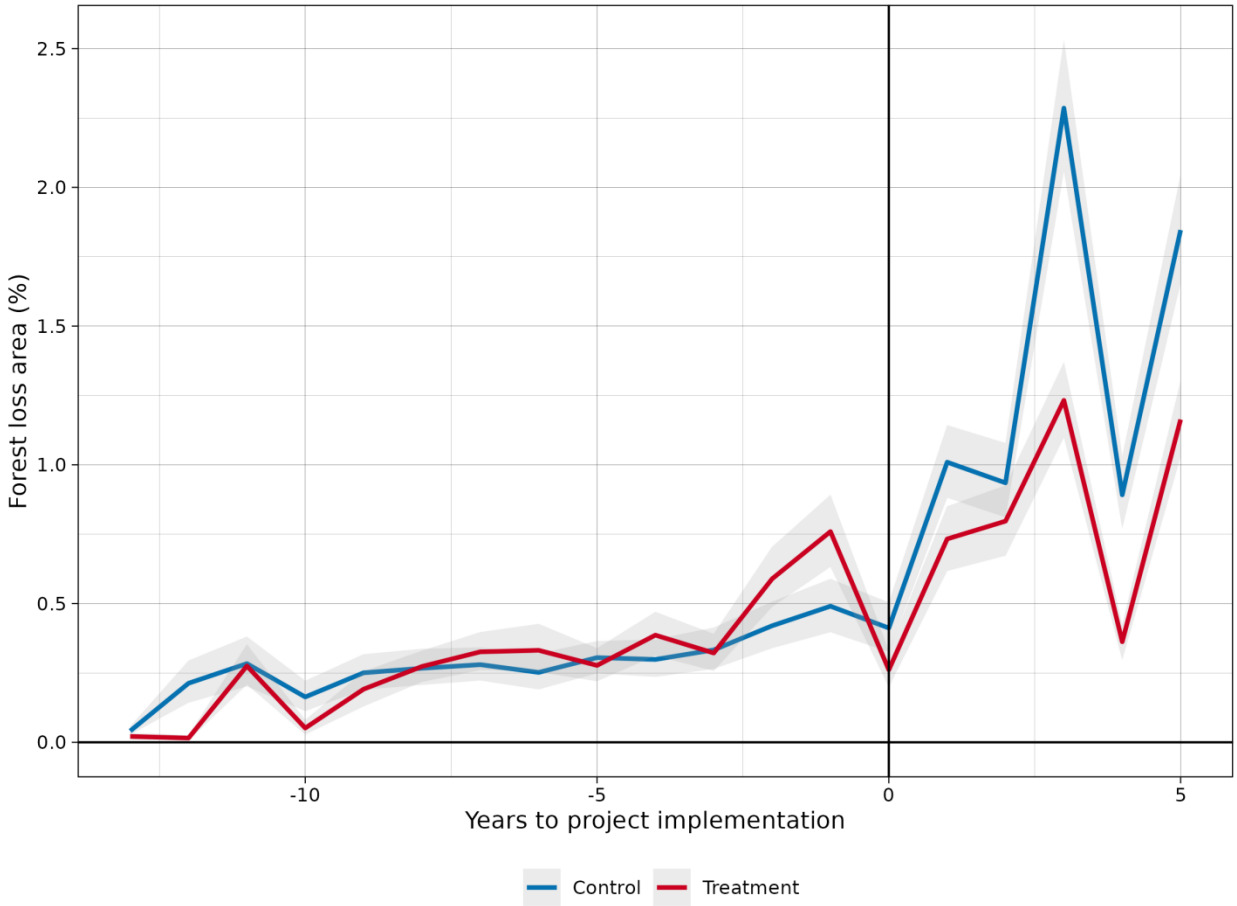

Figure S1. Patterns of deforestation, before and after project implementation, for the Ankeniheny-Zahamena (CAZ) Protected Area project in Madagascar (VCS 1311). Plot-level mean estimates of forest loss (Ha) in treated (blue) and control (red) samples. 95% confidence intervals (grey) generated by bootstrapping. This particular time-series comes from a matched set that meets all quality criteria and was used to assess the sensitivity of estimates to unobserved confounders.

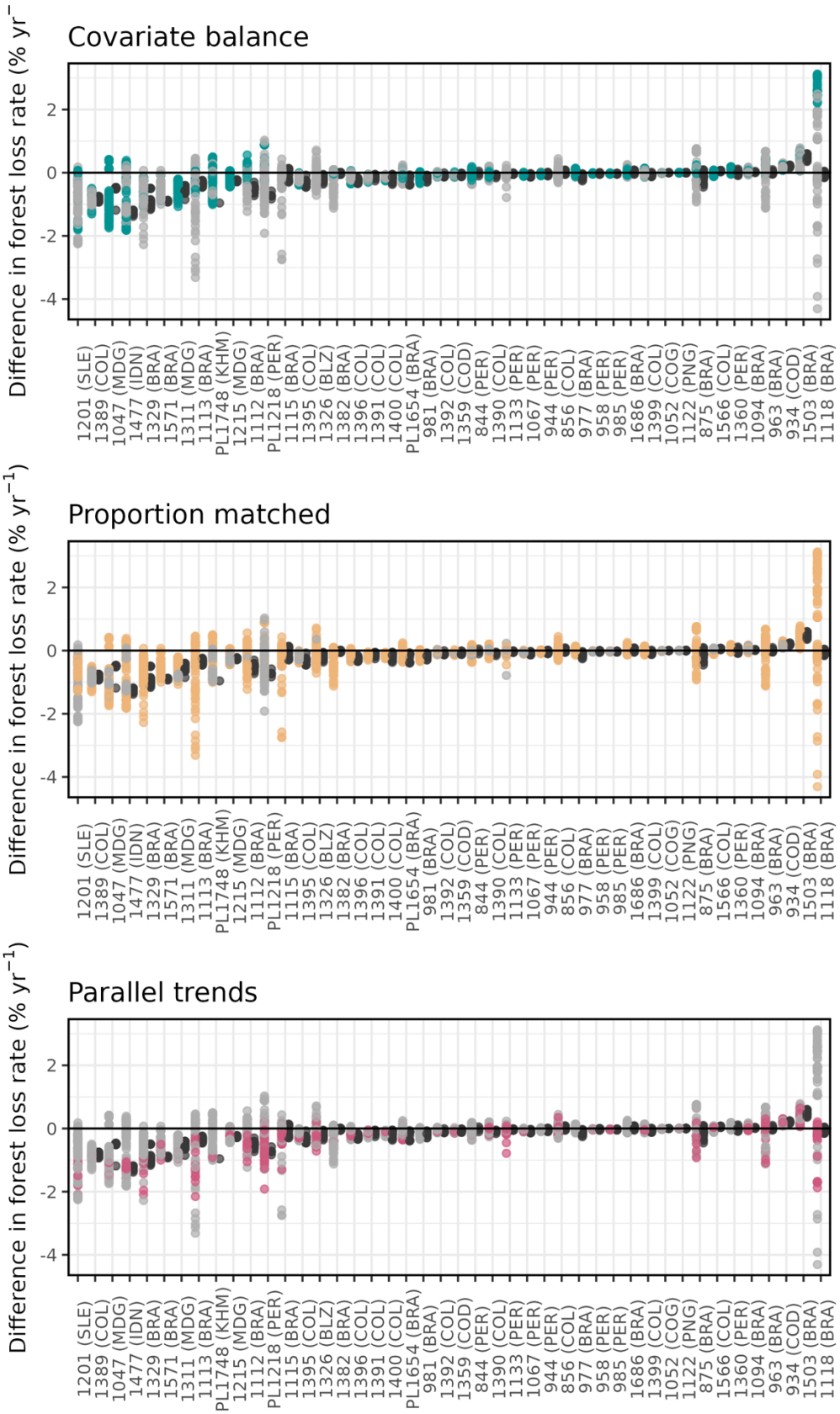
Covariate balance
Proportion matched
Parallel trends
Difference in forest loss rate (% yr^{-1})
2
0
-2
-4
1201 (SLE)
1389 (COL)
1047 (MDG)
1477 (IDN)
1329 (BRA)
1571 (BRA)
1311 (MDG)
1113 (BRA)
PL1748 (KHM)
1215 (MDG)
1112 (BRA)
PL1218 (PER)
1115 (BRA)
1395 (COL)
1326 (BLZ)
1382 (BRA)
1396 (COL)
1391 (COL)
1400 (COL)
PL1654 (BRA)
981 (BRA)
1392 (COL)
1359 (COD)
844 (PER)
1390 (COL)
1133 (PER)
1067 (PER)
944 (PER)
856 (COL)
977 (BRA)
958 (PER)
985 (PER)
1686 (BRA)
1399 (COL)
1052 (COG)
1122 (PNG)
875 (BRA)
1566 (COL)
1360 (PER)
1094 (BRA)
963 (BRA)
934 (COD)
1503 (BRA)
1118 (BRA)

Figure S2 The relative importance of the three validation criteria as the reasons matched sets were rejected. The three panels show differences in deforestation rates for the 44 REDD+ projects (% yr^{-1}) during the first five years of project implementation, arranged by project-level average differences (x-axis). Grey indicates that the run failed to meet the criteria covered by the panel, colour shows that it met the criteria but failed to meet other criteria, black indicates that all criteria were met.

SUPPLEMENTARY TABLES

Table S1: Covariates selected for matching. All the covariates listed in the table are included as part of our core set of matching variables, except for 'Temporal estimates of forest cover' variables, which we used to iterate in the matching runs (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	SRTM Digital Elevation	Elevation (meters)	Raster (90m2)
	steeper slopes consistently associated with a lower risk of deforestation	Data Version 4	Slope (degrees)	
	(Joppa & Pfaff 2009).	(Jarvis et al. 2008)		
Accessibility	Accessibility to forest lands from urban centres is associated with a higher risk of deforestation (Chomitz et al. 1999; Rudel et al. 2009).	Global accessibility map (Weiss et al. 2018)	Time travel to population centers (seconds)	Raster (300m2) Temporal res: 2015
Bioclimatic	The biophysical characteristics of biomes, such as temperature, precipitation and agricultural suitability (Busch & Ferretti-Gallon 2023). Our sampling occurred in tropical moist forest regions for climatic comparability	JRC Tropical Moist Forest database (Vancutsem et al. 2021)	Biome (Tropical moist forest)	Raster (30m2)
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
Proximity to deforestation	Proximity to cleared lands is consistently associated with higher deforestation risk (Busch & Ferretti-Gallon 2023). We computed the average distance to the closest deforested pixel in sampled plots used for matching	Own analysis using the JRC Tropical Moist Forest database (Vancutsem et al. 2021)	Euclidean distance to deforested areas (meters)	Raster (30m2)
Are of forest disturbance	The risk of deforestation increases when there are cleared or disturbed patches of forest (Busch & Ferretti-Gallon 2023). We calculated the area of degraded forests within plots by the year the project was started.	Own analysis using the JRC Tropical Moist Forest database (Vancutsem et al. 2021)	Degraded area (Ha)	Raster (30m2)

Category	Rationale for inclusion	Source	Variables	Format and resolution
Temporal estimates of forest cover	**We used estimates of plot-level forest cover at the beginning of the TMF time series (1990) and at the time of project implementation to ensure temporal dynamics of deforestation were captured in the matching**	**JRC Tropical Moist Forest database (Vancutsem et al. 2021)**	**Undisturbed forest cover extent (Ha)**	**Raster (30m2)**

463

Table S2: Project-level treatment effect estimates (bootstrapped robustly matched sets only) for the 43 REDD+ projects for which valid matched sets are available. Estimated treated effects are averaged over the first 5 years. Projects classified as "Avoided Planned Deforestation and forest Degradation" indicated with asterisks.

VCS ID (ISO 3)	Project Name	Area (km^2)	No. of matched sets	Bootstrapped effect size [95% CI] (% yr^{-1})	Avoided deforestation [95% CI] (ha)
1326* (BLZ)	Laguna Seca Forest Carbon Project	87.2	29	-0.19 [-0.22 -0.16]	-83.47 [-95.18 -70.94]
1571 (BRA)	Manoa REDD+ Project	736	4	-0.9 [-0.93 -0.88]	-3318.66 [-3413.23 -3227.12]
1329 (BRA)	Maísa REDD+ Project	302	25	-0.89 [-0.96 -0.79]	-1340.19 [-1446.6 -1192.02]
1112 (BRA)	The Russas Project	428	25	-0.5 [-0.54 -0.47]	-1070.02 [-1149.51 -1001.8]
1113 (BRA)	The Valparaiso Project	292	24	-0.36 [-0.38 -0.33]	-520.78 [-554.02 -485.26]
PL1654 (BRA)	Fortaleza Ituxi REDD Project	473	32	-0.2 [-0.24 -0.16]	-467.5 [-583.37 -373.65]
981 (BRA)	ADPML PORTEL-PARA REDD PROJECT	1502	46	-0.15 [-0.17 -0.14]	-1133.21 [-1240.19 -1057.73]
875 (BRA)	FLORESTAL SANTA MARIA PROJECT	721	35	-0.12 [-0.16 -0.08]	-419.68 [-572.16 -296.31]
977 (BRA)	RMDLT PORTEL-PARA REDD PROJECT	2111	60	-0.1 [-0.11 -0.09]	-1082.68 [-1211.82 -977.59]
1118 (BRA)	Suruí Forest Carbon Project	336	17	-0.05 [-0.08 -0.02]	-85.8 [-137.66 -35.38]
1115 (BRA)	JARI/AMAPÁ REDD+ PROJECT	782	12	-0.04 [-0.14 0.04]	-160.56 [-558.82 178.31]
1382* (BRA)	The Envira Amazonia Project - A Tropical Forest Conservation Project in Acre, Brazil	397	25	-0.02 [-0.03 -0.01]	-40.71 [-60.06 -22.76]
963 (BRA)	The Purus Project	357	18	0.02 [0 0.03]	28.42 [-4.85 58.45]
1094 (BRA)	Ecomapua Amazon REDD Project	991	4	0.02 [0 0.03]	97.95 [24.69 146.95]
1686 (BRA)	Agrocortex REDD Project	1877	9	0.02 [-0.01 0.05]	221.17 [-96.31 506.78]
1503 (BRA)	Resex Rio Preto-Jacundá REDD+ Project	1016	26	0.48 [0.46 0.5]	2439.17 [2326.51 2563.96]
1359* (COD)	Isangi REDD+ Project	1994	52	-0.07 [-0.07 -0.06]	-671.94 [-750.74 -594.03]
934* (COD)	The Mai Ndombe REDD+ Project	2796	13	0.2 [0.19 0.21]	2825.1 [2666.45 2994.18]
1052 (COG)	North Pikounda REDD+	927	7	0	0
1389 (COL)	Acapa - Bajo Mira Y Frontera REDD+ Project	554	27	-0.85 [-0.87 -0.83]	-2360.74 [-2415.12 -2297.13]

VCS ID (ISO 3)	Project Name	Area (km^2)	No. of matched sets	Bootstrapped effect size [95% CI] (% yr^{-1})	Avoided deforestation [95% CI] (ha)
1395 (COL)	Bajo Calima y Bahía Málaga (BCBM) REDD+ Project	920	21	-0.33 [-0.36 -0.28]	-1510.07 [-1674.47 -1315.22]
1396 (COL)	Rio Pepe y ACABA REDD+ Project	611	33	-0.27 [-0.29 -0.25]	-822.68 [-875.51 -768.41]
1400 (COL)	Concosta REDD+ Project	650	41	-0.14 [-0.17 -0.12]	-461.1 [-548.38 -402.51]
1391 (COL)	SUPP REDD+ Project	542	34	-0.12 [-0.13 -0.1]	-311.84 [-357.23 -278.15]
1392 (COL)	Cajambre REDD+ Project	658	20	-0.1 [-0.11 -0.09]	-342.11 [-369.02 -281.52]
1390 (COL)	Carmen del Darién (CDD) REDD+ Project	1342	6	-0.09 [-0.13 -0.07]	-630.17 [-891.73 -474.07]
1399 (COL)	Mutatá REDD+ Project	405	44	-0.07 [-0.08 -0.06]	-138.13 [-154.74 -120.16]
856 (COL)	The Chocó-Darién Conservation Corridor REDD Project	107	26	-0.01 [-0.02 0.01]	-3.67 [-10.48 5.45]
1566 (COL)	REDD+ Project Resguardo Indigena Unificado Selva de Mataven (RIU SM)	15572	11	0.06 [0.05 0.07]	4556.63 [3670.79 5073.51]
1477* (IDN)	Katingan Peatland Restoration and Conservation Project	1507	8	-1.26 [-1.31 -1.22]	-9507.73 [-9908.11 -9170.93]
PL1748 (KHM)	Southern Cardamoms REDD+ Project	5005	1	-0.96	-23994.31
1311 (MDG)	Carbon Emissions Reduction Project in the Corridor Ankeniheny-Zahamena (CAZ) Protected Area, Madagascar	3926	29	-0.53 [-0.57 -0.49]	-10334.87 [-11195.7 -9663.48]
1047 (MDG)	Carbon Emissions Reduction Project in the Forest Corridor Ambositra-Vondrozo (COFAV), Madagascar	1443	21	-0.52 [-0.66 -0.49]	-3764.09 [-4750.5 -3510.7]
1215 (MDG)	The Makira Forest Protected Area in Madagascar	3747	21	-0.26 [-0.27 -0.25]	-4817.44 [-4969.3 -4672.71]
PL1218 (PER)	Evio Kuiñaji Ese´Eja Cuana, To Mitigate Climate Change, Madre de Dios - Perú	87.2	5	-0.72 [-0.78 -0.64]	-314.31 [-338.39 -278.57]
1067 (PER)	Reduction of deforestation and degradation in Tambopata National Reserve and Bahuaja-Sonene National Park within the area of Madre de Dios region –Peru	5658	26	-0.08 [-0.1 -0.07]	-2400.25 [-2814.02 -2027.18]
1133 (PER)	Yacumama Forest Carbon Project	30.2	36	-0.04 [-0.04 -0.04]	-5.91 [-6.31 -5.33]
944 (PER)	Alto Mayo Conservation Initiative	1787	55	-0.03 [-0.04 -0.03]	-297.55 [-367.55 -233.76]
985 (PER)	Cordillera Azul National Park REDD project	13627	26	-0.03 [-0.03 -0.03]	-2144.26 [-2374.04 -1931.25]
958 (PER)	BIOCORREDOR MARTIN SAGRADO REDD+ PROJECT	3003	80	-0.03 [-0.03 -0.03]	-433.83 [-462.61 -402.1]
844 (PER)	Madre de Dios Amazon REDD Project	985	4	-0.03 [-0.09 0.02]	-124.8 [-425.11 101.08]
1360 (PER)	Forest Management to reduce deforestation and degradation in Shipibo Conibo and	1293	62	0 [-0.01 0.01]	9.42 [-44.97 61.84]

VCS ID (ISO 3)	Project Name	Area (km^2)	No. of matched sets	Bootstrapped effect size [95% CI] (% yr^{-1})	Avoided deforestation [95% CI] (ha)
	Cacataibo Indigenous communities of Ucayali region				
1122* (PNG)	April Salumei REDD Project	6539	42	0 [0 0]	34.57 [5.93 63.67]

468