IOP Publishing Environ. Res. Lett. 20 (2025) 114051 https://doi.org/10.1088/1748-9326/ae0f44

ENVIRONMENTAL RESEARCH
LETTERS

OPEN ACCESS

RECEIVED
19 June 2025

REVISED
21 September 2025

ACCEPTED FOR PUBLICATION
3 October 2025

PUBLISHED
28 October 2025

LETTER

Strengthening the integrity of REDD+ credits: objectively assessing counterfactual methods using placebos

E-Ping Rau[1,2,*], Jody Holland[1,3], Thomas Swinfield[1,3], Abby Williams[1,3,4], Srinivasan Keshav[1,5] and David A Coomes[1,2]

1 Conservation Research Institute, University of Cambridge, Pembroke Street, Cambridge, United Kingdom
2 Department of Plant Sciences, University of Cambridge, Downing Street, Cambridge, United Kingdom
3 Department of Zoology, University of Cambridge, Downing Street, Cambridge, United Kingdom
4 Department of Biology, University of Oxford, Life and Mind Building, South Parks Road, Oxford, United Kingdom
5 Department of Computer Science and Technology, University of Cambridge, William Gates Building, 15 JJ Thomson Avenue, Cambridge, United Kingdom
* Author to whom any correspondence should be addressed.

E-mail: epr26@cam.ac.uk

Keywords: carbon credit, placebo test, counterfactual analysis, conservation finance, forest conservation, carbon market, emissions reductions

Supplementary material for this article is available online

Abstract

The financing of tropical forest conservation projects through the sale of carbon credits remains a key opportunity to curb forest loss. Reducing emissions from deforestation and forest degradation (REDD)+ projects generate carbon credits by reducing forest loss within the project area compared with a counterfactual area that faces similar pressures (known as 'additionality'). Several methods are available for constructing counterfactuals, but comparing their reliability is challenging. Here, we present an evaluation approach based on the creation of placebo projects, where there are no REDD+ activities and in which we would not expect project and counterfactual outcomes to diverge. We compare four methods based on pixel matching that estimate counterfactual deforestation rates. Using 27 placebo projects spread across the tropics, we found that pixel-matching is a reliable way of estimating a key element of additionality (i.e. deforestation in counterfactual areas) when based on data gathered at the end of an evaluation period after project start (i.e. *ex-post* estimation). However, forecasting counterfactual deforestation rates from information available at the start of a project (i.e. *ex-ante* estimation) is much less reliable, reinforcing existing concerns about *ex-ante* crediting mechanisms. We argue that systematic application of the placebo approach can accelerate the development and adoption of more credible counterfactual-estimating methods. As counterfactuals are the basis which underpins the validity claims of most nature credits, strengthening the credibility of counterfactuals will enhance the effectiveness of conservation finance, helping REDD+ and other nature-based solutions realise their full potential in delivering real, measurable benefits.

1. Introduction

The escalating deforestation and forest degradation across the tropics (World Resources Institute 2025) have devastating consequences on biodiversity, carbon balance, and climate change (Barlow *et al* 2016, Baccini *et al* 2017, Betts *et al* 2017). The reducing emissions from deforestation and forest degradation (REDD) programme, as well as the expanded REDD+ framework, aims to reduce carbon emissions by incentivising alternative activities to deforestation while promoting biodiversity and livelihood co-benefits: it is therefore an indispensable tool to mitigate the climate and biodiversity crises (Roe *et al* 2019, Girardin *et al* 2021, Seddon *et al* 2021). To ensure that REDD+ projects provide genuine benefits, emissions reductions must be quantified following the principle of *additionality* as the difference between carbon losses in the project area and its *counterfactual scenario*—i.e. what would have occurred without the project (ICVCM 2023). Verified emissions reductions are issued as carbon credits and traded in the voluntary carbon markets, providing a critical financing mechanism for REDD+ (Norman and Nakhooda 2015, Seymour and Langer 2021).

However, recent studies showed evidence of widespread over-crediting in REDD+ projects, with only 7%–25% of issued credits representing genuine emissions reductions (West *et al* 2023, Probst *et al* 2024). The ensuing controversy has severely undermined market confidence in REDD+ credibility and shrunk the carbon market value by 61% in 2023 (Procton 2024, Greenfield 2024), further fragilising the already insufficient REDD+ finance (Morita and Matsumoto 2023). Apart from the issue of inadequate carbon accounting and governance, which current development of jurisdictional approaches attempt to address (von Essen and Lambin 2021, Zhao *et al* 2025), the phenomenon of over-crediting also underscores the major challenge of estimating the inherently unobservable counterfactual outcomes. Current methodologies typically rely on historical deforestation trends in expert-defined reference areas to generate *ex-ante* forecasts of counterfactual outcomes before the intervention starts. However, this has been criticised for being prone to biases by failing to adequately account for temporally varying environmental variables that drive deforestation. This creates discrepancies between variables considered important before project start and those that are identified as important after project start, resulting in misestimation of the counterfactual outcomes (Swinfield *et al* in review, Swinfield *et al* 2024, West *et al* 2024). In contrast, counterfactual areas identified by matching their environmental variables to project areas allow for both project and counterfactual outcomes to be evaluated *ex post* (after interventions started) during the same period, which could mitigate misalignment and biases caused by time-varying deforestation drivers. *Ex-post* analyses are increasingly used in independent studies that evaluate claims of emissions reductions in REDD+ projects (Guizar-Coutiño *et al* 2022, Delacote *et al* 2025), but they are currently not employed by any official certification framework, and the adequacy of the statistical matching process has also come under question (Pauly *et al* 2024). Faced with this debate, a standardised and objective assessment of the robustness and accuracy of counterfactual-estimating methods is key to improving the credibility and viability of REDD+ credits (Delacote *et al* 2024, West *et al* 2024), as well as other nature credits based on counterfactuals.

One approach to assess counterfactual-estimating methods is by applying them to *placebo* projects, areas designated to be representative of the deforestation risks of existing REDD+ sites but not yet subject to any REDD+ activities, to examine if the methods correctly capture the absence of effect. The placebo approach has been widely used in the social sciences (Abadie *et al* 2010): for example, Abadie and Gardeazabal (2003) estimated the effect of a treatment on economic activity in the treated region by comparing it to similar but untreated regions as counterfactuals, and they tested the robustness of this analysis by applying it to a comparable untreated region, finding 'counterfactuals' for this region. This approach was recently applied to quantify variations in counterfactual deforestation rates in the absence of REDD+ activities, which were then used to construct confidence intervals for the observed REDD+ effects (West *et al* 2020, 2023). We propose that placebos could also allow us to evaluate competing counterfactual-estimating methods, via performance statistics that quantify overall uncertainty and bias of predictions of counterfactual deforestation rates.

In this study, we designate 27 placebo projects within the tropical moist forest (TMF) biome with comparable environmental characteristics to existing REDD+ projects, and track deforestation trends using remotely-sensed data on land class transition. Setting a putative year of project start, we define three *ex-ante* methods that forecast counterfactual deforestation rates after project start for the placebo projects using historical deforestation rates before project start (figure 1(A)). We also generate *ex-post* estimates of counterfactual deforestation rates after project start, using pixel matching to identify counterfactual pixels with similar characteristics to those of project pixels, and estimating deforestation rates in counterfactual pixels after project start (figure 1(B)). Since project and counterfactual deforestation rates should follow the same trend in placebo projects, both *ex-ante* forecasts and *ex-post* estimates can be compared against post-start observed deforestation rates within the placebo projects (figure 1(B)): the most robust counterfactual-estimating method can be identified as the one producing predictions with the lowest uncertainties and biases. Specifically, we address the following questions: (1) How accurate and precise is each method for estimating counterfactual outcomes? (2) How does the performance of each method change through time, and how does it inform us on the strengths and weaknesses of these methods? Finally, we discuss how the placebo approach can contribute towards an operational benchmarking tool for counterfactual-estimating methods.

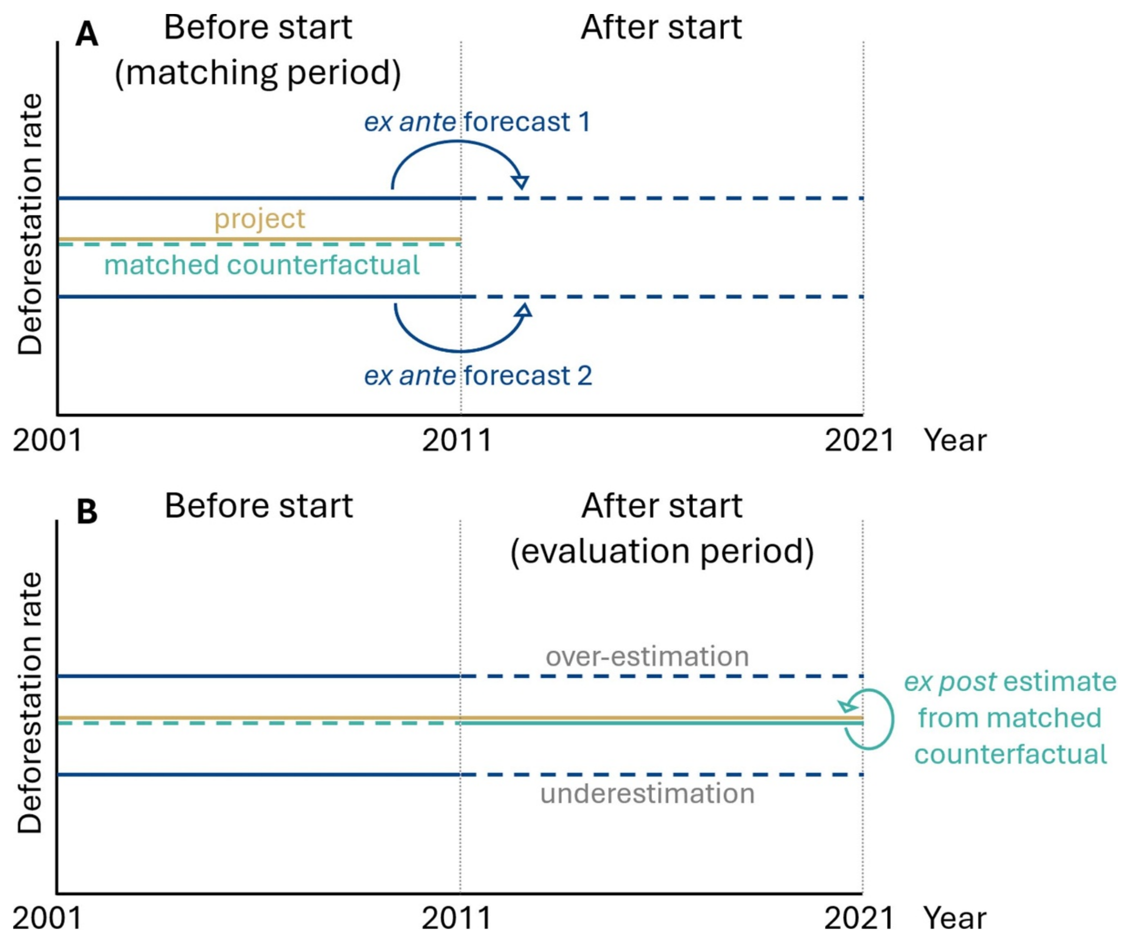

Figure 1. Illustration of deforestation rate estimates for a placebo project with a putative start year of 2011, where the 'project' and counterfactual deforestation rates are expected to follow the same trends due to the absence of interventions. (A) During the decade before project start (2001–2011), different methods are used to produce *ex-ante* forecasts (dashed blue lines) based on historical deforestation rates (solid blue lines), and pixel matching was used to find matched counterfactual pixels (dashed green lines) for the project. (B) During the decade after project start (2011–2021), the *ex-post* estimate of counterfactual deforestation rate is calculated from matched pixels (solid green line), and both *ex-ante* forecasts and *ex-post* estimates are compared against observed project deforestation rates (solid yellowline). Overestimation or underestimation occurs when the predicted deforestation rate is higher or lower than the observed deforestation rate, respectively. Note that this figure depicts one of the many possible scenarios, where project interventions are the only factor influencing deforestation, and that in the absence of intervention, deforestation rate does not change over time. In practice, this assumption is likely to not hold true, and time-varying deforestation drivers would result in deforestation rates of project or matched counterfactuals not perfectly lining up between the two period. Ultimately, under the parallel trends assumption, the best method should still be the one that produces a counterfactual that is closest to the post-start observed deforestation rate in the placebo projects.

2. Methods

2.1. Overview

We adopted a pixel-based approach to identify counterfactuals and track deforestation, based on the PACT 2.0 methodology (Swinfield and Balmford 2023), detailed in Balmford *et al* (2023) and implemented in Python (Dales *et al* 2024). Data analyses and visualisation were conducted in Google Earth Engine and Python 3.12.2 with the libraries *ee, matplotlib, numpy, pandas, random*, and *scipy* (Hunter 2007, Van Rossum and Drake 2009, McKinney 2010, Gorelick *et al* 2017, Harris *et al* 2020, Reback *et al* 2020, Van Rossum 2020, Virtanen *et al* 2020, Google Earth Engine Developers 2024), and the code is available at a GitHub repository (Rau and Holland 2025).

The PACT methodology uses the TMF data product (Vancutsem *et al* 2021) to track deforestation within project areas and a set of counterfactual pixels, identified via pixel matching as having similar environmental variables to project pixels. The TMF data is a Landsat-derived dataset that classifies the land class of every 30-m pixel in the TMF biome annually, providing a time series of forest change since 1990. We focused on the four land classes related to forest cover (undisturbed forest, degraded forest, deforested land, and forest regrowth), and derived compound annual deforestation rates (r, %) over a given period from the geometric mean of the fraction of remaining forest over said period:

$$r = \left[1 - \left(\frac{N_t}{N_0}\right)^{\frac{1}{t}}\right] \times 100 \qquad (1)$$

where N_0 and N_t represent the number of pixels with undisturbed forest class at the start and end of the period, respectively, and t represents the number of years within the period.

The use of pixel matching allows us to explicitly control for environmental variables that influence deforestation rates (Busch and Ferretti-Gallon 2023), and the use of global geospatial data, as is common practice with most credit issuing standards and rating agencies (Wawrzynowicz *et al* 2023), ensures consistency, transparency and replicability of our deforestation estimates.

2.2. Estimating counterfactual deforestation with ex-ante and ex-post methods

We designed 27 placebo projects as areas with comparable size and remoteness to existing REDD+ sites but independent of ongoing REDD+ activities, across the wet tropics of South America (10), Africa (8) and Southeast Asia (10) (figure 2). Detailed steps of placebo project selection can be found in Supplementary Information 1.

We defined **project pixels** as randomly sampled pixels within placebo project areas, at a density of 0.25 points/ha for projects <250 000 ha and 0.05 points/ha for larger projects (figure 4, Step 2). We defined the regional **candidate pixels**, to be used for matching and for *ex-ante* forecasts, as pixels sharing the same countries and ecoregions as the project pixels and were unaffected by REDD interventions: detailed procedures for identifying candidate pixels can be found in Supplementary Information 2.

We matched pixels using the following environmental variables that are known proxies of deforestation drivers (Busch and Ferretti-Gallon 2023): (1) historical TMF land classes (undisturbed forest, degraded forest, deforested land, and forest regrowth), (2) historical forest cover, (3) slope and elevation, and (4) remoteness. Historical forest cover was calculated as the proportional cover of the 'undisturbed forest' or 'deforested land' TMF land classes within a 1-km radius of each TMF pixel. Slope and elevation were derived from SRTM data (Jarvis *et al* 2008).

We applied three *ex-ante* forecasting methods (figure 3), setting the putative year of start for placebo projects to be 2011: this is the year with historically the highest number of new REDD+ projects (Atmadja *et al* 2022, figure 5), and the latest year allowing for the availability of at least a decade of post-start TMF data (2011–2021).

1. **Regional method**: we selected all candidate pixels with the same land classes during the matching period (at 2001, 2006, and 2011), and similar values for continuous environmental variables to project pixels: ±0.1 for historical forest cover, ±200 m for elevation, ±2.5° for slope, and ±10 min for remoteness (measured in walking distances to nearest healthcare facility, proxy for distance to settlements). This method aims to capture deforestation pressures from the broader landscape that the project area had not yet experienced before the project start, but might be exposed to afterward. To prevent out-of-memory errors during computing, we limited the size of the pixel set to 250 000 pixels through random downsampling.
2. **Project method**: we selected sampled project pixels without matching. This method forecasts the counterfactual deforestation rate after project start to follow the historical project deforestation rate.
3. **Time-shifted matching method**: we selected candidate pixels where conditions 20–10 years before the project start (at 1991, 1996 and 2001) matched those of project pixels 10–0 years before start (at 2001, 2006, and 2011), matching pixels with the same land classes and the lowest Mahalanobis distance for continuous environmental variables during the matching period (Guizar-Coutiño *et al* 2022, Balmford *et al* 2023). The rationale for this method is that it produces a 10-year dynamic forecast, using deforestation trends in the past to update the forecast for the future over time, which could help account for time-varying deforestation drivers.

For all three *ex-ante* methods, forecasted counterfactual deforestation rates were calculated as the annual deforestation rates in the selected pixels over the pre-start period (2001–2011). We also estimated counterfactual deforestation rates using the ***ex-post*** **matching method**, identifying *matched counterfactual pixels* as candidate pixels whose conditions matched those of project pixels 10–0 years before start (at 2001, 2006, and 2011), with the same land classes and the lowest Mahalanobis distance for continuous environmental variables. We then calculated the *ex-post* counterfactual deforestation rate in the matched counterfactual pixels over the decade after project start (2011–2021). A detailed mathematical representation of the methods used is provided in Supplementary Information 3.

2.3. Evaluating predictive performance

We compared and plotted both *ex-ante* forecasts and *ex-post* estimates with the observed deforestation rates in all placebo projects after project start (2011–2021) (figure 4). We assessed predictive performance across all projects with the following three metrics: (1) mean absolute error (MAE) between observed and estimated deforestation rates (lower values indicating higher precision); (2) mean bias, calculated as the mean difference of predicted minus observed values (lower absolute values

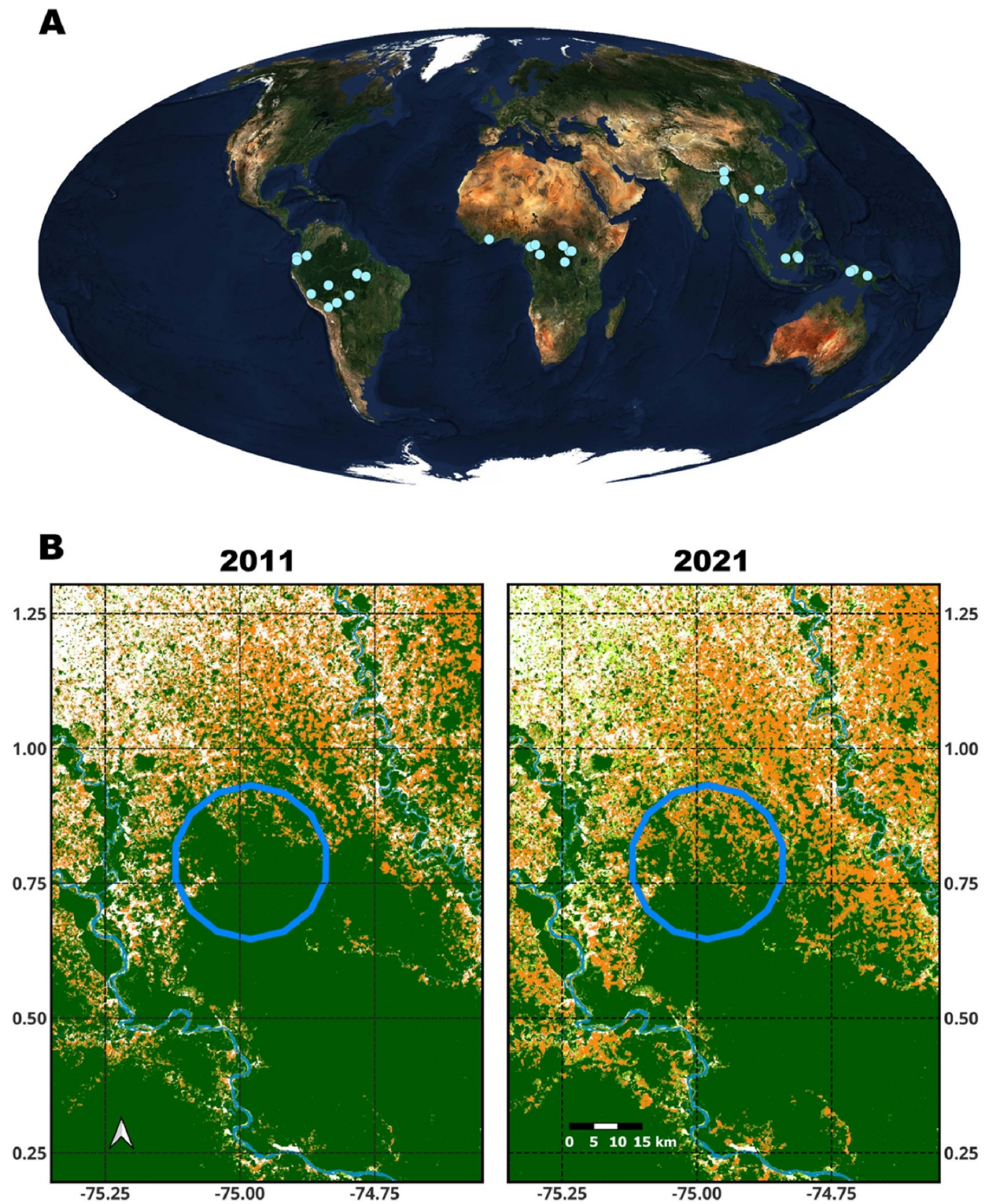

Figure 2. (A) Map of the locations of the 27 placebo projects (blue dots). Basemap: Sentinel-2 cloudless mosaic in 2024 in Molleweide projection (EOX IT Services GmbH 2024). (B) Example of one placebo project (blue circle) in South America overlaid on TMF Annual Change Collection maps from 2011 (left) and 2021 (right). Dark green: undisturbed forest; light green: degraded forest; orange: deforested land; yellow-green: forest regrowth; white: non-forest.

indicating higher accuracy); (3) goodness-of-fit, calculated as the coefficient of determination between predicted and observed values (R^2 over the identity line) (higher values indicating higher precision and accuracy). We also evaluated how well different methods perform when used to predict deforestation rates over periods of varying lengths after the start of the project (figure 5): the detailed method of this analysis can be found in Supplementary Information 4.

3. Results

Among the three *ex-ante* forecasting methods, the regional method (MAE = 0.876, goodness-of-fit = 0.50) and the time-shifted matching method (MAE = 0.772, goodness-of-fit = 0.51) produced forecasts with higher robustness for observed deforestation rates than the project method (MAE = 0.904, goodness-of-fit = 0.13) (figure 4(A)). *Ex-post* estimates showed considerably higher robustness than all

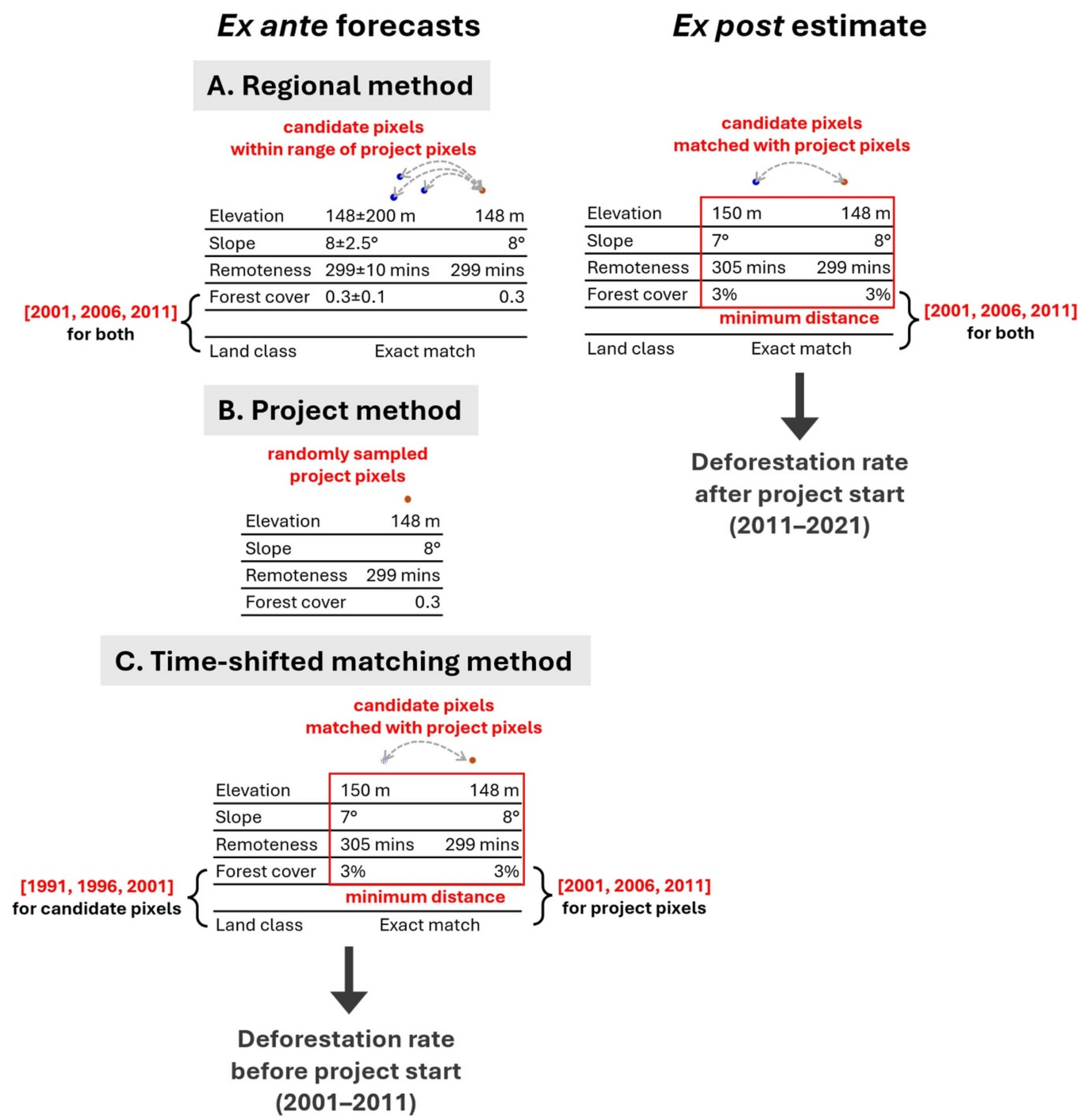

Figure 3. Illustration of ***ex-ante* forecasts** and ***ex-post* estimation** of counterfactual deforestation rates for a placebo project, setting 2011 as the putative start. After randomly sampling a set of project pixels and a set of candidate pixels, *ex-ante* forecasts are calculated via one of the three methods: (A) deforestation rates in the regional candidate pixels, falling within range of project pixels, in 2001–2011; (B) deforestation rates in the project pixels in 2001–2011; (C) deforestation rates in a set of matched candidate pixels in 2001–2011, whose characteristics in 1991–2001 are matched to those of the project pixels in 2001–2011. *Ex-post* estimates are calculated as deforestation rates in a set of matched candidate pixels in 2011–2021, whose characteristics in 2001–2011 are matched to those of the project pixels in the same period.

three *ex-ante* forecasts (MAE = 0.480, goodness-of-fit = 0.85) (figure 4(B)). All *ex-ante* methods showed negative biases (from −0.213 to −0.847), indicating underestimation of the observed deforestation rates, while the *ex-post* method showed positive bias (0.297), indicating overestimation (figure 4).

All three *ex-ante* methods had higher MAEs than the *ex-post* method for all predicted periods (figure 5(A)). Both regional and time-shifted methods slightly over-estimated deforestation rates for periods in the nearer future, but shifted to slight underestimation for longer periods, whereas the project method consistently underestimated deforestation rate, with increasingly severe underestimation; in contrast, the *ex-post* method consistently exhibited slight overestimation (figure 5(B)). The goodness-of-fit of different methods revealed different temporal trends: initially, the regional method performed poorly with a low goodness-of-fit for shorter time periods in the nearer future, but it improved over time for longer periods (explaining up to 50% of the variation in the observed deforestation rate in the period of 2011–2021). Whereas the project method performed well initially (explaining up to 70% of the variation in the observed deforestation rate in the period of 2011–2012) but deteriorated over time for longer periods. The *ex-post* method consistently showed the highest goodness-of-fit across all predicted periods (figure 5(C)).

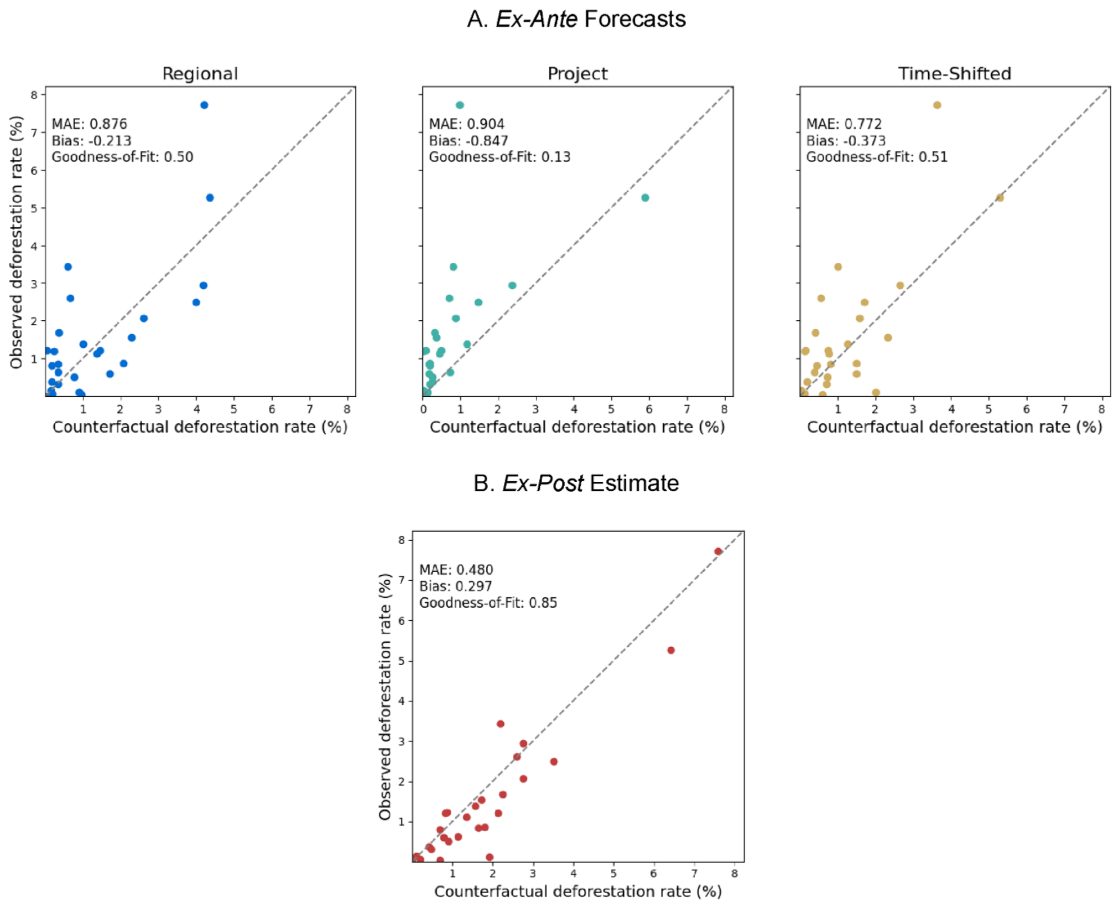

Figure 4. Observed compound annual deforestation rates (%) in the placebo project area versus counterfactual deforestation rates that are (A) forecasted via each of the three *ex-ante* methods and (B) estimated from the *ex-post* matching method. Three statistics are reported: (1) mean absolute error (MAE) between observed and counterfactual deforestation rates; (2) mean bias, calculated as the mean difference between predicted and observed values: positive bias indicates over-estimation and negative bias indicates underestimation; (3) goodness-of-fit, calculated as the R^2 over the identity line (coefficient of determination between predicted and observed values). Dashed lines indicate the 1:1 relationship between predicted and observed values.

4. Discussion

This study introduces a reliable way to evaluate the accuracy and robustness of methods that estimate counterfactual deforestation rates for project-based REDD+ initiatives. By using carefully designated *placebo projects* that are unaffected by REDD+ activities, the unobservable counterfactual deforestation rates become observable as the deforestation rate within the placebo projects. By controlling for properties such as size, remoteness and initial forest cover, we ensure that the placebo projects are representative of the environmental characteristics in existing REDD+ projects and comparable in terms of suitability for project development.

The *ex-post* matching method produced estimates with a remarkably high goodness-of-fit (85%) between observed and counterfactual deforestation rates over a 10 year period. In contrast, the *ex-ante* project method exhibited much lower goodness-of-fit (13%). Because this method uses historical deforestation rates in the placebo projects to directly forecast future deforestation rates in the same areas, the only factor differing between forecasts and observed values is the timing of the measurement: this strongly suggests that its low performance was caused by time-varying deforestation drivers, whose strength and direction may vary across placebo projects and their surrounding landscapes (Busch and Ferretti-Gallon 2023, Hänggli *et al* 2023, Wells *et al* 2023). The regional and time-shifted methods exhibited higher goodness-of-fit (50%–51%), suggesting that they partially correct for the time variance effect, by using either deforestation trends in the wider surrounding landscape that are expected to progress into project areas over time (regional method), or deforestation trends from an earlier time period that serve as a proxy for future changes (time-shifted method). Nevertheless, the comparison between the *ex-ante* time-shifted method and the *ex-post* method, both using the same pixel-matching criteria and differing only in the matching period, suggests that the time variance effect remains even after these corrections. The negative biases (underestimation of post-start counterfactual deforestation rates) exhibited by all three *ex-ante* methods may suggest overall

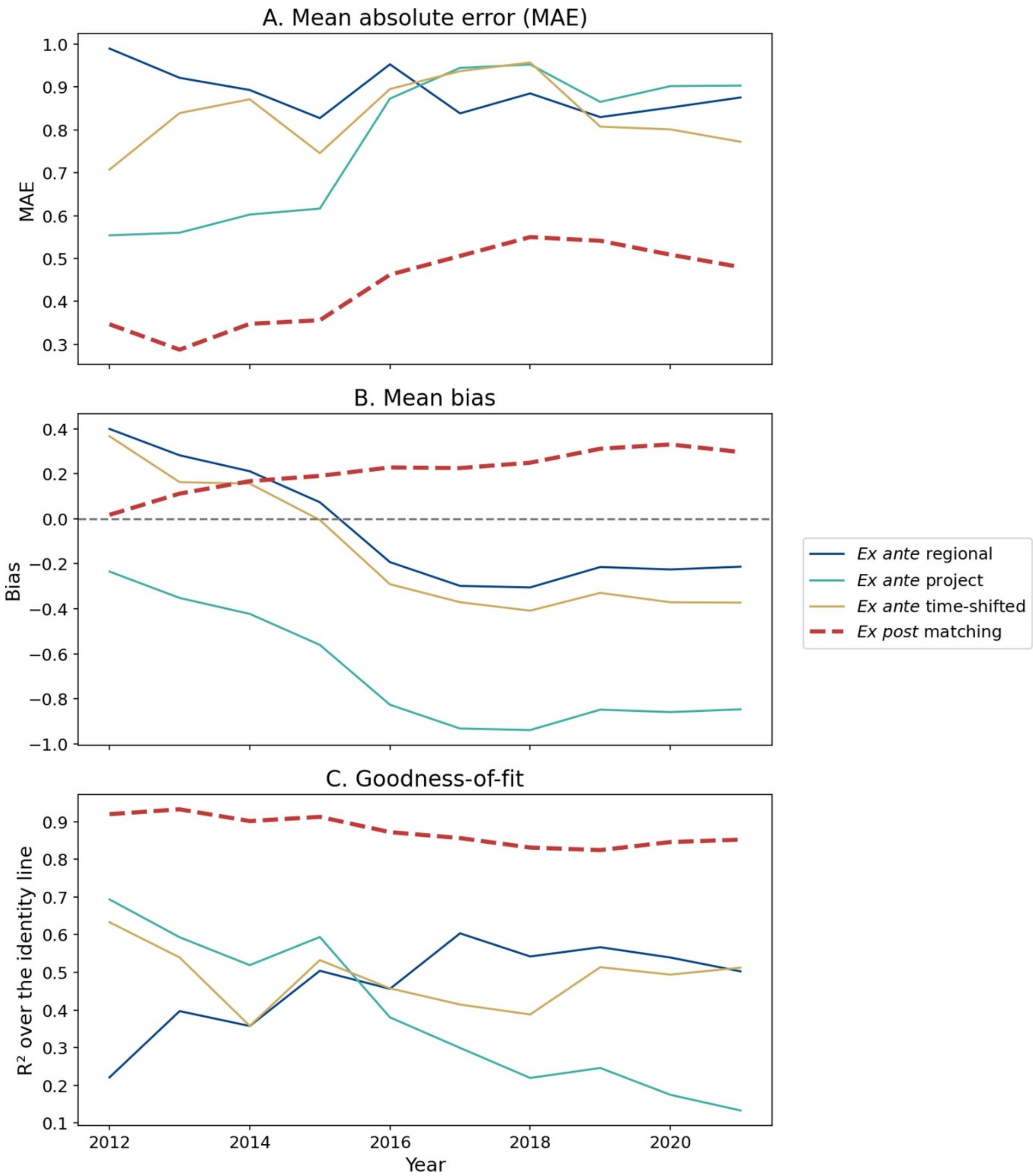

Figure 5. Predictive performance of the three *ex-ante* forecasting methods (thin solid lines; blue: regional; green: project, brown: time-shifted) and the *ex-post* matching method (thick dashed red line) when used to predict the observed counterfactual deforestation rates (%) over the period starting from 2011 and ending in different years, indicated on the *x*-axis. Predictive performance is measured in (A) mean absolute error, (B) mean bias and (C) goodness-of-fit (R^2 over the identity line, or coefficient of determination between predicted and observed values). The horizontal dashed line in panel B indicates zero bias: values above zero indicate over-estimation, whereas values below zero indicate underestimation.

accelerating deforestation in the areas examined during 2011–2021 compared to the previous decade, although only partial evidence supports this hypothesis in specific regions (e.g. West and Central Africa and Central America) (Vancutsem *et al* 2021), and the underlying causes of these biases remain uncertain and warrant further investigation. Similarly, the positive bias (overestimation) of the *ex-post* method suggests that the currently used matching variables are not perfect proxies of deforestation trends, and that there is room for improvement. When predicting counterfactual deforestation rates over longer periods in the future, the *ex-ante* project method showed deteriorating performance, suggesting that time variance of deforestation drivers increase over time; in contrast, the regional method showed slightly improving performance, suggesting that the deforestation trends in the project area takes time to align with the deforestation trends in the wider landscape.

The comparison between *ex-ante* and *ex-post* methods has concrete implications for the REDD+ crediting mechanism. Responding to multiple criticisms towards project-based methodologies

(West *et al* 2023, Swinfield *et al* 2024), jurisdictional approaches have been increasingly advocated in recent years. Jurisdictional approaches evaluate emissions reductions at the level of jurisdictions (national or subnational administrative entities) (von Essen and Lambin 2021). The use of jurisdiction-wide counterfactual scenarios relieves the burden and reduces adverse incentives for local project developers, and the involvement of governments, which have stronger capabilities of enacting policies and land use monitoring, enables the efforts of multiple stakeholders to be aggregated, ensuring a more effective governance (DeFries *et al* 2022, Zhao *et al* 2025). However, newly developed methodologies (e.g. VM0048, ART TREEs) still rely on *ex-ante* methods to provide counterfactual scenarios based on historic averages (ART Secretariat 2021, Bird 2022, Verra 2023). Our results suggest that the performance of *ex-ante* methods could be limited by biases arising from complex and time-variant deforestation drivers (Teo *et al* 2023, Schneider *et al* 2024, Swinfield *et al* in review), while the *ex-post* method adequately controls for such biases because project and counterfactual outcomes are matched and compared over the same period. While we do not claim that our results should be generalised as definitive proof of the superiority of all *ex-post* counterfactuals over all *ex-ante* counterfactuals, they provide support to the strengths of *ex-post* methods and strongly encourage further exploring how they could be adopted in the crediting mechanism. Such a change would not necessarily delay the timing of credit issuance, since *ex-post* evaluation of project outcomes is already required in current methodologies, and it is also in line with the Core Carbon Principles, which recommends that emission reductions be verified *ex post* to ensure robustness (ICVCM 2024).

Nevertheless, we acknowledge that *ex-ante* forecasts can still be valuable for project planning, such as by aiding the selection of locations with high deforestation risks (Olander *et al* 2008, Wendland *et al* 2010, Lin *et al* 2014), or by demonstrating the potential impacts of projects, which is needed to attract investment (Norman and Nakhooda 2015). This is especially important since REDD+ projects frequently experience delays in credit revenue (∼5 years) (West *et al* 2020, Atmadja *et al* 2022), and they often require significant upfront investment for readiness activities (Zhu *et al* 2010, Bernard *et al* 2014). Overcoming this obstacle will require carefully designing governance structures to incorporate mechanisms to allow credit issuance to be adjustable over time as *ex-post* measurements are gathered to complement *ex-ante* forecasts, balancing the need for upfront financing with robust quantification of additionality. One relevant example is the UK Woodland Carbon Code, which offers landowners the option to receive credits based on *ex-ante* forecasts or to wait for *ex-post* verification in exchange for higher credit valuation (Woodland Carbon Code 2025).

Although both *ex-ante* and *ex-post* counterfactual-estimating methods have their role to play within REDD+, the lack of systematic evaluation of the robustness of different methods impedes the development and adoption of best-practice crediting methods (Delacote *et al* 2024), and partially contributes to the low market confidence and price of carbon credits. The placebo approach represents a practical way to fill this technical gap and has the potential to help elevate levels of trust in REDD+ credits: this is needed to increase credit demand and in turn price. We propose that an operational benchmark procedure would involve a publicly available set of placebo polygons that are separated into a training set and a test set. This dataset can be used by developers of counterfactual-estimating methods to evaluate the performance of their methods, akin to how benchmarking datasets are used in machine learning to evaluate algorithm performance and to aid the development of more accurate algorithms (e.g. ImageNet or FOR-instance) (Deng *et al* 2009, Puliti *et al* 2023). They can also be adopted by credit rating agencies as part of their toolkit to evaluate crediting methodologies. The placebo approach could be applied to both voluntary and regulatory markets, and could be adapted to work under either the project-based or jurisdictional framework. Since jurisdictional approaches usually incorporate projects nested under a jurisdiction, placebo projects could still be useful to benchmark the robustness of the evaluation of the nested projects. More generally, it is also conceivable to directly identify placebos for jurisdictions, as is done in social sciences (Abadie *et al* 2010), to assess methods for producing reference emission levels, provided that the challenge of finding adequate candidate regions can be overcome. Moreover, the placebo approach could be extended to other types of nature credits, such as carbon credits generated by Afforestation, Reforestation, & Revegetation (ARR) projects (Agarwal *et al* 2025), or biodiversity credits (Wunder *et al* 2024, zu Ermgassen *et al* 2025), which all follow the principle of additionality: although *ex-post* methodologies for ARR credits have been developed (e.g. VM0047 by Verra), the emerging biodiversity markets so far largely rely on *ex-ante* counterfactuals (Wauchope *et al* 2024, Wunder *et al* 2024), and all of these counterfactuals remain in need of formal assessment.

One important caveat of our study is that the outcome of the placebo evaluation is affected by the data and the criteria used in placebo selection and in the matching process. For example, replacing the static remoteness data with time series data capturing changes in remoteness over time due to road expansion or creation of new settlements, such as the annual global Human Footprint dataset (Mu *et al* 2022), could potentially improve matching quality. In addition, when designating placebo projects, we explicitly excluded areas with existing REDD+ projects but not other types of protected areas. Although activities occurring in these protected areas are highly diverse and not always comprehensively documented, formally quantifying their effects, for example by examining overlap between placebo projects with protected areas from a global dataset (e.g. WDPA) (Bingham *et al* 2019), is an important future direction. It is also important to assess the extent to which benchmark metrics are influenced by hidden covariates such as shape, spatial continuity, jurisdiction, putative project start and geographical location, and whether accounting for those covariates could improve performance of counterfactuals. Uncertainties in remotely-sensed estimates of deforestation and carbon loss, resulting from factors such as uneven geographic and temporal coverage, geolocation imprecision, and uncertainties in biomass estimation, must also be adequately quantified and reported. Uncertainty estimates provided in available data products should be reported and integrated in the analysis: for example, the GEDI L4A data set contains standard error of the predictions of aboveground biomass density (Dubayah *et al* 2022). Regional geospatial datasets, which may be available in higher resolutions or be validated with more sufficient ground truthing, can also complement global datasets. Finally, systematic propagation of all sources of error should be a crucial part of all carbon reporting, via tools such as Bayesian inference or Monte Carlo simulation, and increasing the expertise in these techniques should be a part of the more concerted governance and implementation (Butler *et al* 2024). Detailed quantification and propagation of uncertainty is recommended by both the Core Carbon Principles (ICVCM 2023) and the UNFCCC COP decisions (UNFCCC 2010), and will also allow us to integrate advances in remote sensing and higher-quality data to improve estimation accuracy (Jucker *et al* 2018, Besnard *et al* 2025). Finally, increasing the number of placebo projects would enhance statistical power and improve the quantification of uncertainties and biases in the estimates of deforestation rates, counterfactual estimates and additionality claims (Golmant *et al* 2023).

In conclusion, this study presents a practical way to evaluate and potentially improve counterfactual-estimating methods in REDD+ programmes using placebos, showing that the *ex-post* matching method significantly outperforms all three of the *ex-ante* forecasting methods we trialled. We argue a standardised placebo dataset should be used for systematic assessment of the suitability of existing and proposed REDD+ methodologies for carbon credit issuance. By facilitating methodological advancement, this practical tool could ultimately enhance market trust and viability and reinforce the role of REDD+ in global climate mitigation efforts.

Data availability statement

The Python code used for data processing, analysis and visualisation are available at https://github.com/epingchris/placebo_evaluation (commit: 53e0eb0).

The data that support the findings of this study are openly available at the following URL/DOI: https://doi.org/10.5281/zenodo.15183844.

Supplementary Information available at https://doi.org/10.1088/1748-9326/ae0f44/data1.

Acknowledgments

We would like to thank Charlotte Wheeler and Elliot Kinsey for discussing and exchanging thoughts on the research topic over the course of the manuscript, Michael Dales, Patrick Ferris, and Anil Madhavapeddy for providing maintenance and support for the computing infrastructure of the Department of Computer Science and Technology at the University of Cambridge, as well as all authors of the PACT Methodology and its implementation code. We are grateful for the administrative, informatics and financial support provided by the Cambridge Centre for Carbon Credits (4C). This research was partly funded by a donation from the Tezos Foundation (NRAG/719). D C and E-P R is supported by Centre for Landscape Regeneration (CLR) (Grant No. NE/W00495X/1).

Conflict of interest

T S is the CSO of Canopy PACT, a registered charity that supports the development of carbon accrediting methodology for REDD+ projects. S K and D A C are trustees of Canopy PACT. The Cambridge Centre for Carbon Credits (4C) has no commercial interest in carbon credits.

Author contributions

E-Ping Rau 0000-0001-6344-9655
Conceptualization (equal), Data curation (equal), Formal analysis (equal), Investigation (equal), Methodology (equal), Visualization (equal), Writing – original draft (equal), Writing – review & editing (equal)

Jody Holland 0009-0007-8416-0077
Data curation (equal), Formal analysis (equal), Investigation (equal), Methodology (equal),

Visualization (equal), Writing – original draft (equal), Writing – review & editing (equal)

Thomas Swinfield 0000-0001-9354-5090
Conceptualization (equal), Data curation (equal), Formal analysis (equal), Investigation (equal), Methodology (equal), Writing – original draft (equal), Writing – review & editing (equal)

Abby Williams 0000-0002-9119-5880
Data curation (equal), Formal analysis (equal), Investigation (equal), Methodology (equal), Writing – review & editing (equal)

Srinivasan Keshav 0000-0002-6549-0464
Conceptualization (equal), Investigation (equal), Project administration (equal), Resources (equal), Supervision (equal), Writing – original draft (equal), Writing – review & editing (equal)

David A Coomes 0000-0002-8261-2582
Conceptualization (equal), Data curation (equal), Funding acquisition (equal), Investigation (equal), Project administration (equal), Supervision (equal), Writing – original draft (equal), Writing – review & editing (equal)

References

Abadie A, Diamond A and Hainmueller J 2010 Synthetic control methods for comparative case studies: estimating the effect of california's tobacco control program *J. Am. Stat. Assoc.* **105** 493–505

Abadie A and Gardeazabal J 2003 The economic costs of conflict: a case study of the basque country *Am. Econ. Rev.* **93** 113–32

Agarwal S, Clark B, Vaddi U, Goveas N, Nagendra H and DeFries R 2025 Achieving high-integrity tree-planting projects in the voluntary carbon market *Environ. Res. Lett.* **20** 044052

ART Secretariat 2021 TREES: the REDD+ environmental excellence standard (available at: www.artredd.org/trees/)

Atmadja S S, Duchelle A E, De Sy V, Selviana V, Komalasari M, Sills E O and Angelsen A 2022 How do REDD+ projects contribute to the goals of the paris agreement? *Environ. Res. Lett.* **17** 044038

Baccini A, Walker W, Carvalho L, Farina M, Sulla-Menashe D and Houghton R A 2017 Tropical forests are a net carbon source based on aboveground measurements of gain and loss *Science* **358** 230–4

Balmford A, Coomes D A, Hartup J, Jaffer S, Keshav S, Lam M, Madhavapeddy A, Rau E-P, Swinfield T and Wheeler C 2023 PACT tropical moist forest accreditation methodology *Cambridge Open Engage* (https://doi.org/10.33774/coe-2023-g584d-v2)

Barlow J *et al* 2016 Anthropogenic disturbance in tropical forests can double biodiversity loss from deforestation *Nature* **535** 144–7

Bernard F, Minang P A, Adkins B and Freund J T 2014 REDD+ projects and national-level readiness processes: a case study from Kenya *Clim. Policy* **14** 788–800

Besnard S, Heinrich V H A, Carvalhais N, Ciais P, Herold M, Luijkx I, Peters W, Suarez D R, Santoro M and Yang H 2025 Global covariation of forest age transitions with the net carbon balance *Nat. Ecol. Evol.* 1–13

Betts M G, Wolf C, Ripple W J, Phalan B, Millers K A, Duarte A, Butchart S H M and Levi T 2017 Global forest loss disproportionately erodes biodiversity in intact landscapes *Nature* **547** 441–4

Bingham H C *et al* 2019 Sixty years of tracking conservation progress using the world database on protected areas *Nat. Ecol. Evol.* **3** 737–43

Bird S C and (Wildlife Works Carbon) 2022 Jurisdictional & nested redd+ framework (Verra (available at: https://verra.org/programs/jurisdictional-nested-redd-framework/) (Accessed 7 October 2022)

Busch J and Ferretti-Gallon K 2023 What drives and stops deforestation, reforestation, and forest degradation? An updated meta-analysis *Rev. Environ. Econ. Policy* **17** 217–50

Butler B J, Sass E M, Gamarra J G, Campbell J L, Wayson C, Olguín M, Carrillo O and Yanai R D 2024 Uncertainty in REDD+ carbon accounting: a survey of experts involved in REDD+ reporting *Carbon Balance Manag.* **19**

Dales M, Ferris P, Message R, Holland J and Williams A 2024 GitHub repository: tropical moist forest accreditation methodology implementation (available at: https://github.com/quantifyearth/tmf-implementation.commit:7f15246)

DeFries R, Ahuja R, Friedman J, Gordon D R, Hamburg S P, Kerr S, Mwangi J, Nouwen C and Pandit N 2022 Land management can contribute to net zero *Science* **376** 1163–5

Delacote P *et al* 2025 Restoring credibility in carbon offsets through systematic ex post evaluation *Nat. Sustain.* **8** 733–40

Delacote P, Le Velly G and Simonet G 2024 Distinguishing potential and effective additionality of forest conservation interventions *Environ. Dev. Econ.* **29** 518–38

Deng J, Dong W, Socher R, Li-Jia L, Li K and Fei-Fei L 2009 ImageNet: a large-scale hierarchical image database *2009 IEEE Conf. on Computer Vision and Pattern Recognition* pp 248–55

EOX IT Services GmbH 2024 Sentinel-2 cloudless 2024 available at: https://s2maps.eu)

Girardin C A J, Jenkins S, Seddon N, Allen M, Lewis S L, Wheeler C E, Griscom B W and Malhi Y 2021 Nature-based solutions can help cool the planet—if we act now *Nature* **593** 191–4

Golmant N, Morrissey M, Silva C, Dorrek F and Engstrand R 2023 Assessing forest carbon offset additionality with dynamic baselines and uncertainty quantification *KDD 2023 Workshop Fragile Earth* (available at: https://openreview.net/forum?id=tWVNnTLDvf)

Google Earth Engine Developers 2024 Earth engine python API, version 1.5.24 (available at: https://developers.google.com/earth-engine)

Gorelick N, Hancher M, Dixon M, Ilyushchenko S, Thau D and Moore R 2017 Google earth engine: planetary-scale geospatial analysis for everyone *Remote Sens. Environ.* **202** 18–27

Greenfield P 2024 Market value of carbon offsets drops 61%, report finds (The Guardian) (available at: www.theguardian.com/environment/article/2024/may/31/market-value-of-carbon-offsets-drops-61-aoe) (Accessed 31 May 2024)

Guizar-Coutiño A, Jones J P G, Balmford A, Carmenta R and Coomes D A 2022 A global evaluation of the effectiveness of voluntary REDD+ projects at reducing deforestation and degradation in the moist tropics *Conserv. Biol.* **36** e13970

Hänggli A, Levy S A, Armenteras D, Bovolo C I, Brandão J, Rueda X and Garrett R D 2023 A systematic comparison of deforestation drivers and policy effectiveness across the Amazon biome *Environ. Res. Lett.* **18** 073001

Harris C R *et al* 2020 Array programming with NumPy *Nature* **585** 357–62

Hunter J D 2007 Matplotlib: a 2D graphics environment *Comput. Sci. Eng.* **9** 90–95

ICVCM 2023 The core carbon principles (available at: https://icvcm.org/core-carbon-principles/) (Accessed 22 December 2023)
ICVCM 2024 Core carbon principles assessment framework and procedure. Part 3: summary for decision makers (available at: https://icvcm.org/wp-content/uploads/2024/02/CCP-Section-3-V1.1-FINAL-15May24.pdf) (Accessed January 2024)
Jarvis A, Reuter H I, Nelson A and Guevara E 2008 Hole-filled SRTM for the globe version 4, available from the CGIAR-CSI SRTM 90m database (available at: http://srtm.csi.cgiar.org)
Jucker T *et al* 2018 Estimating aboveground carbon density and its uncertainty in Borneo's structurally complex tropical forests using airborne laser scanning *Biogeosciences* **15** 3811–30
Lin L, Sills E and Cheshire H 2014 Targeting areas for reducing emissions from deforestation and forest degradation (REDD+) projects in Tanzania *Glob. Environ. Change* **24** 277–86
McKinney W 2010 Data structures for statistical computing in Python *Proc. 9th Python in Science Conf.* ed S van der Walt and J Millman pp 56–61
Morita K and Matsumoto K 2023 Challenges and lessons learned for REDD+ finance and its governance *Carbon Balance Manage.* **18** 1–18
Mu H, Li X, Wen Y, Huang J, Du P, Su W, Miao S and Geng M 2022 A global record of annual terrestrial human footprint dataset from 2000 to 2018 *Sci. Data* **9** 176
Norman M and Nakhooda S 2015 The state of REDD+ finance Center for Global Development Working Paper **378**
Olander L P, Gibbs H K, Steininger M, Swenson J J and Murray B C 2008 Reference scenarios for deforestation and forest degradation in support of REDD: a review of data and methods *Environ. Res. Lett.* **3** 025011
Pauly M, Crosse W, Moore C, Brown K, Griffin O, Brade T, Mitchard E T A, Freund J, Appleby S and Tosteson J 2024 A holistic approach to assessing REDD+ forest loss baselines through ex post analysis *Environ. Res. Lett.* **19** 124096
Probst B S *et al* 2024 Systematic assessment of the achieved emission reductions of carbon crediting projects *Nat. Commun.* **15** 9562
Procton A 2024 State of the voluntary carbon market 2024: On the path to maturity Forest Trends' Ecosystem Marketplace
Puliti S, Pearse G, Surový P, Wallace L, Hollaus M, Wielgosz M and Astrup R 2023 FOR-instance: a UAV laser scanning benchmark dataset for semantic and instance segmentation of individual trees (arXiv: 2309.01279)
Rau E-P and Holland J 2025 (commit: 53e0eb0) GitHub Repository: Evaluating methods of counterfactual estimation with placebos (available at: https://github.com/epingchris/placebo_evaluation)
Reback J *et al* 2020 pandas-dev/pandas: Pandas 1.0.5 (https://doi.org/10.5281/zenodo.3898987)
Roe S *et al* 2019 Contribution of the land sector to a 1.5 °C world *Nat. Clim. Change* **9** 817–28
Schneider L, Fuessler J, Oberpriller Q and Spalding-Fecher R 2024 The ICVCM approval of three REDD methodologies presents risks to the integrity of the initiative (Oeko.de) (available at: www.oeko.de/en/blog/the-icvcm-approval-of-three-redd-methodologies-presents-risks-to-the-integrity-of-the-initiative/) (12 October 2024)
Seddon N, Smith A, Smith P, Key I, Chausson A, Girardin C, House J, Srivastava S and Turner B 2021 Getting the message right on nature-based solutions to climate change *Glob. Change Biol.* **27** 1518–46
Seymour F and Langer P 2021 Consideration of nature-based solutions as offsets in corporate climate change mitigation strategies (World Resources Institute)
Swinfield T and Balmford A 2023 Cambridge carbon impact: evaluating carbon credit claims and co-benefits *Cambridge Open Engage* (https://doi.org/10.33774/coe-2023-bl26j)
Swinfield T *et al* in review Understanding the reasons for over-crediting in REDD+ *Nat. Commun.*
Swinfield T, Shrikanth S, Bull J W, Madhavapeddy A and Zu Ermgassen S O S E 2024 Nature-based credit markets at a crossroads *Nat. Sustain.* **7** 1217–20
Teo H C *et al* 2023 Uncertainties in deforestation emission baseline methodologies and implications for carbon markets *Nat. Commun.* **14** 8277
Van Rossum G 2020 *The Python Library Reference, Release 3.8.2.* Python Software Foundation
Van Rossum G and Drake F L 2009 (CreateSpace) Python 3 reference manual
Vancutsem C, Achard F, Pekel J-F, Vieilledent G, Carboni S, Simonetti D, Gallego J, Aragão L E O C and Nasi R 2021 Long-term (1990–2019) monitoring of forest cover changes in the humid tropics *Sci. Adv.* **7** eabe1603
Verra 2023 VM0048 reducing emissions from deforestation and forest degradation, v1.0 (Verra) (available at: https://verra.org/methodologies/vm0048-reducing-emissions-from-deforestation-and-forest-degradation-v1-0/) (Accessed 27 November 2023)
Virtanen P *et al* 2020 SciPy 1.0: fundamental algorithms for scientific computing in Python *Nat. Methods* **17** 261–72
von Essen M and Lambin E F 2021 Jurisdictional approaches to sustainable resource use *Front. Ecol. Environ.* **19** 159–67
Wauchope H S, Ermgassen S O S E Z, Jones J P G, Carter H, Bühne H S T and Milner-Gulland E J 2024 What is a unit of nature? Measurement challenges in the emerging biodiversity credit market *Proc. Biol. Sci.* **291** 20242353
Wawrzynowicz I, Krey M and Samaniego X 2023 Assessing and comparing carbon credit rating agencies (Perspectives Climate Group) (available at: https://carbonmarketwatch.org/wp-content/uploads/2023/09/PCG_CMW_rating_agencies_final_report_.pdf)
Wells G, Pascual U, Stephenson C and Ryan C M 2023 Confronting deep uncertainty in the forest carbon industry *Science* **382** 41–43
Wendland K J, Honzák M, Portela R, Vitale B, Rubinoff S and Randrianarisoa J 2010 Targeting and implementing payments for ecosystem services: opportunities for bundling biodiversity conservation with carbon and water services in madagascar *Ecol. Econ.* **69** 2093–107
West T A P, Bomfim B and Haya B K 2024 Methodological issues with deforestation baselines compromise the integrity of carbon offsets from REDD+ *Glob. Environ. Change* **87** 102863
West T A P, Börner J, Sills E O and Kontoleon A 2020 Overstated carbon emission reductions from voluntary REDD+ projects in the Brazilian Amazon *Proc. Natl Acad. Sci. USA* **117** 24188–94
West T A P, Wunder S, Sills E O, Börner J, Rifai S W, Neidermeier A N, Frey G P and Kontoleon A 2023 Action needed to make carbon offsets from forest conservation work for climate change mitigation *Science* **381** 873–7
Woodland Carbon Code 2025 Requirements and guidance for woodland carbon code projects (available at: www.woodlandcarboncode.org.uk/view-the-code) (Accessed 1 August 2025)
World Resources Institute 2025 Forest loss indicator (Global Forest Review) (available at: https://research.wri.org/gfr/forest-extent-indicators/forest-loss) (Accessed 21 May 2025)
Wunder S *et al* 2024 Biodiversity credits: learning lessons from other approaches to incentivize conservation *OSF Preprints* (https://doi.org/10.31219/osf.io/qgwfc_v1)
Zhao F, Shen Y and Yin R 2025 Reframing forest-based climate actions *Environ. Sci.: Adv.* **4** 705–12
Zhu X, Møller L R, De Lopez T T and Zaballa M E 2010 *Pathways for Implementing REDD+: Experiences from Carbon Markets and Communities* (Danmarks Tekniske Universitet, Risø Nationallaboratoriet for Bæredygtig Energi)
zu Ermgassen S Z, Swinfield T, Bull J W, Duffus N, Macintosh A, Maron M, Theis S, White T and Evans M C 2025 Five golden rules for scientifically-credible nature markets (https://doi.org/10.31235/osf.io/37w8p_v1)