

LETTER • OPEN ACCESS

Comparing methods for assessing the effectiveness of subnational REDD+ initiatives

To cite this article: Astrid B Bos *et al* 2017 *Environ. Res. Lett.* **12** 074007

View the [article online](#) for updates and enhancements.

Recent citations

- [Monitoring systems to improve forest conditions](#)
Laura Vang Rasmussen and Martin Rudbeck Jepsen
- [Assessing the effects of ecological restoration approaches in the alpine rangelands of the Qinghai-Tibetan Plateau](#)
Lin Zhen *et al*
- [REDD+ Crossroads Post Paris: Politics, Lessons and Interplays](#)
Esteve Corbera and Heike Schroeder

Environmental Research Letters



LETTER

OPEN ACCESS

RECEIVED

11 January 2017

REVISED

20 March 2017

ACCEPTED FOR PUBLICATION

28 April 2017

PUBLISHED

30 June 2017

Original content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](#).

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.



Comparing methods for assessing the effectiveness of subnational REDD+ initiatives

Astrid B Bos^{1,6}, Amy E Duchelle², Arild Angelsen³, Valerio Avitabile¹, Veronique De Sy¹, Martin Herold¹, Shijo Joseph², Claudio de Sassi², Erin O Sills⁴, William D Sunderlin² and Sven Wunder⁵

¹ Wageningen University & Research, Laboratory of Geo-Information Science and Remote Sensing, Droevendaalsesteeg 3, 6708 PB Wageningen, the Netherlands

² Center for International Forestry Research, 16000 Bogor, Indonesia

³ Norwegian University of Life Sciences, School of Economics and Business, 1432 Aas, Norway

⁴ North Carolina State University, Department of Forestry and Environmental Resources, 27695 Raleigh, NC, United States of America

⁵ Center for International Forestry Research, 1558 Lima, Peru

⁶ Author to whom any correspondence should be addressed.

E-mail: astrid.bos@wur.nl

Keywords: deforestation, climate, forest change, land cover, monitoring, performance assessment, REDD+

Abstract

The central role of forests in climate change mitigation, as recognized in the Paris agreement, makes it increasingly important to develop and test methods for monitoring and evaluating the carbon effectiveness of REDD+. Over the last decade, hundreds of subnational REDD+ initiatives have emerged, presenting an opportunity to pilot and compare different approaches to quantifying impacts on carbon emissions. This study (1) develops a Before-After-Control-Intervention (BACI) method to assess the effectiveness of these REDD+ initiatives; (2) compares the results at the meso (initiative) and micro (village) scales; and (3) compares BACI with the simpler Before-After (BA) results. Our study covers 23 subnational REDD+ initiatives in Brazil, Peru, Cameroon, Tanzania, Indonesia and Vietnam. As a proxy for deforestation, we use annual tree cover loss. We aggregate data into two periods (before and after the start of each initiative). Analysis using control areas ('control-intervention') suggests better REDD+ performance, although the effect is more pronounced at the micro than at the meso level. Yet, BACI requires more data than BA, and is subject to possible bias in the *before* period. Selection of proper control areas is vital, but at either scale is not straightforward. Low absolute deforestation numbers and peak years influence both our BA and BACI results. In principle, BACI is superior, with its potential to effectively control for confounding factors. We conclude that the more local the scale of performance assessment, the more relevant is the use of the BACI approach. For various reasons, we find overall minimal impact of REDD+ in reducing deforestation on the ground thus far. Incorporating results from micro and meso level monitoring into national reporting systems is important, since overall REDD+ impact depends on land use decisions on the ground.

1. Introduction

Reducing emissions from deforestation and forest degradation and enhancing forest carbon stocks (REDD+) has emerged as a key climate change mitigation strategy within the United Nations Framework Convention on Climate Change (UNFCCC). Through the Paris agreement, the necessity for supporting and implementing REDD+ was recon-

firmed and the role of forests as carbon sinks emphasized (UNFCCC 2015). So far, approximately 40 countries⁷ mention either REDD+ or forests as part of the mitigation strategy in their Nationally Determined Contributions (NDCs). This importance makes it critical to monitor and evaluate the carbon effectiveness of REDD+.

⁷ UNFCCC NDC registry www4.unfccc.int/ndcregistry/Pages/All.aspx, 5 December 2016

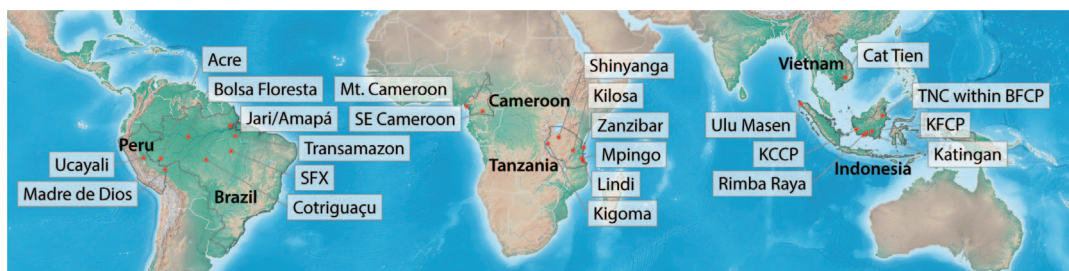


Figure 1. Initiatives included in the Global Comparative Study on REDD+.

The measurement, reporting and verification (MRV) of carbon stocks and emissions is a vital part of national REDD+ schemes (Herold and Skutsch 2009, UNFCCC 2015). Carbon emissions are calculated by multiplying activity data—the area of land use/cover change due to human activity—by its corresponding emission factor (Verchot *et al* 2012). While national forest monitoring systems have progressed, e.g. with PRODES from the Brazilian Institute for Space Research (INPE), capacities in developing and operationalizing these MRV systems vary widely among countries (Romijn *et al* 2015). In the last decade, technical innovations in remote sensing and forest-relevant monitoring techniques resulted in a plethora of national and global datasets with increasing levels of coverage, detail (spatial and temporal) and accuracy. Examples include the Landsat-based Global Forest Change 2000–2014 (Hansen *et al* 2013), global pan-tropical biomass datasets (Baccini *et al* 2012, Saatchi *et al* 2011, Avitabile *et al* 2016), and national carbon maps using LiDAR (Asner *et al* 2013).

Meanwhile, at the subnational level, hundreds of REDD+ projects and programmes are led by a diversity of actors including private non-profit organizations, for-profit companies and government agencies (Simonet *et al* 2015). The implementers of these initiatives are applying a range of REDD+ interventions from enabling measures (such as tenure clarification) to command-and-control measures (disincentives) to direct payments and livelihood improvements (incentives). While data-driven developments facilitate forest and carbon monitoring, it remains unclear how to align information on subnational performance with national level reporting related to NDCs. The implementers of several of these subnational REDD+ initiatives state that ‘vertical integration or nesting of MRV systems is important, but has been elusive’ (Ravikumar *et al* 2015, p 919).

Any effectiveness assessment needs to compare an observed outcome with a hypothetical counterfactual (business-as-usual scenario, baseline or reference level). In the face of dynamic contexts globally (e.g. commodity prices), nationally (e.g. macroeconomic policies), and locally (e.g. newly constructed roads), simple retrospective ‘before-after’ (BA) reference level assessments fail to properly attribute factors of change, and consequently misjudge the impacts of REDD+

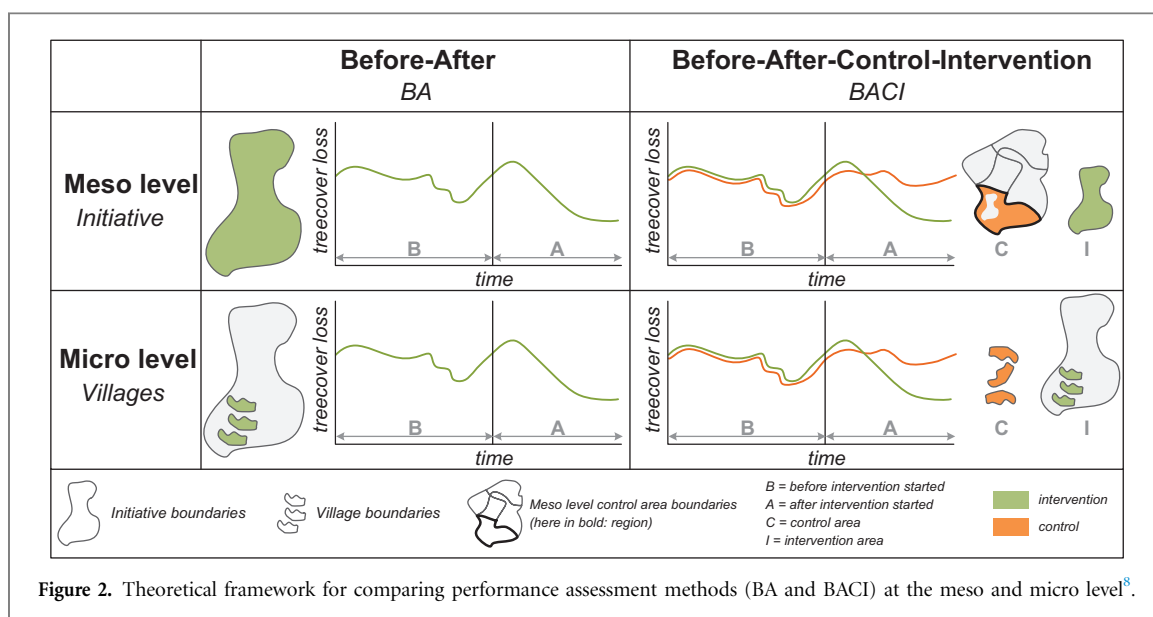
interventions. Establishing a counterfactual that discriminates these confounding effects is the key in assessing true policy impacts. The quasi-experimental Before-After-Control-Intervention (BACI), or differences-in-differences (DID), approach aims to control for these contextual changes. It is applied in ecological studies to assess the effect of a stress or treatment on a given population (Smith 2002) and in econometrics and social sciences for program evaluation (e.g. Imbens and Wooldridge 2009, Jagger *et al* 2010). The unit of interest is measured at (a minimum of) two points in time (before and after the treatment) and in (at least) two different locations, that is, an area subjected to the ‘treatment’ (intervention area) and an area that is not (control area), to identify changes that are additional. The BA approach corresponds to using a conventional reference level, i.e. the average historical deforestation (e.g. past ten years). Hence, unlike BACI, it does not account for changes in drivers during the intervention period. This paper explores the application of both methods to measuring the performance of subnational REDD+ initiatives. The purpose of the comparison is to increase our understanding of conditions under which the more complex and costly BACI approach is essential, and those conditions under which BA might be acceptable.

Here, we (1) develop a BACI method to assess the effectiveness of these REDD+ initiatives; (2) compare the results at the meso (initiative) and micro (village) scales; and (3) compare BACI with BA results. We focus on comparing the results of different methods and scales, rather than on explaining individual performance scores of the REDD+ initiatives.

2. Methodology

2.1. Study area

Our study includes 23 subnational REDD+ initiatives in Brazil, Peru, Cameroon, Tanzania, Indonesia and Vietnam from CIFOR’s Global Comparative Study on REDD+ (GCS) (figure 1). They differ greatly in terms of proponent type (government, NGO, private sector), size (ranging from 28 to approximately 160 000 km²), environmental context (from dense primary rainforest to dry *miombo* woodlands) and interventions applied (Sills *et al* 2014). While specific interventions differ



across sites, most proponents use customized combinations of enabling measures, disincentives and incentives to reduce deforestation and degradation (Duchelle *et al* 2017).

2.2. Tree cover data

We use the Global Forest Change data (version 1.2), which is based on a time series analysis of Landsat satellite imagery, providing tree cover density for 2000 and annual tree cover loss for 2001–2014 (Hansen *et al* 2013). Some have questioned the local accuracy of this global dataset (Bellot *et al* 2014) which may over- or underestimate absolute forest area and forest change in different ways across the globe. Yet, it is currently the only source of annual data on global tree cover loss at medium spatial resolution (Landsat 30 m). Furthermore, for the purpose of comparison among sites and countries, we only present the relative trends of tree cover change and we do not aim to make any claims about deforestation numbers in absolute terms (e.g. ha of forest converted into other land use). That is, in our analysis, we use the data to compare trends within the same region (i.e. comparing villages inside and outside intervention areas, and comparing intervention areas to the surrounding jurisdiction). Thus, we only compare areas that should be subject to the same tendencies towards under- or overestimation of deforestation, thereby removing that bias from the comparison.

Tree cover loss is used as proxy for emissions from deforestation. At this stage, we do not consider carbon emissions (i.e. emission factors). We thus implicitly assume that emissions are mainly driven by activity data. We define forests as areas with >10% tree cover, in line with the FAO (2000) definition.

Accordingly, we generated a forest mask from the *tree cover in 2000* layer from the Hansen data. Forest loss is defined as changes in tree cover from >10% in 2000 to ~0% (see supplementary material of Hansen *et al* 2013) in any subsequent years. Areas of forest loss and, correspondingly, annual forest loss as a percentage of initial forest cover were calculated by using the *area()* function of the Raster package in R (Hijmans 2016).

2.3. Performance assessment framework

For both approaches, we aggregate the time series data on annual tree cover loss into two periods (before and after) (figure 2). To compare assessment approaches, we simultaneously apply BA and BACI approaches. Correspondingly, we calculate relative performance scores to allow for comparison across sites and countries.

REDD+ initiatives' starting years differ, ranging from 2006 to 2013 (Sills *et al* 2014, appendix 6⁹), thus the number of years in the *after* period ranges from two to nine (see table 1). The BA score α is calculated as follows:

$$\text{BA score } \alpha = \bar{x}_{AI} - \bar{x}_{BI} \quad (1)$$

with $\bar{x}_{AI} = \frac{1}{n_a} \sum_{i=1}^{n_a} x_i$ and $\bar{x}_{BI} = \frac{1}{n_b} \sum_{i=1}^{n_b} x_i$

Where \bar{x}_{AI} represents the average annual deforestation rate in the intervention area in the period since the intervention started, as a percentage of the total forest area in 2000; \bar{x}_{BI} represents the average annual deforestation rate in the intervention area in the period from the start year of measurement

⁸ Homogeneous trends in the *before* period like those presented in figure 2 show the ideal situation.

⁹ Start years for Bolsa Floresta, SE Cameroon and KCCP are slightly earlier compared to those reported in Appendix 6 of Sills *et al* (2014) because of activities preceding the official REDD+ initiative start date.

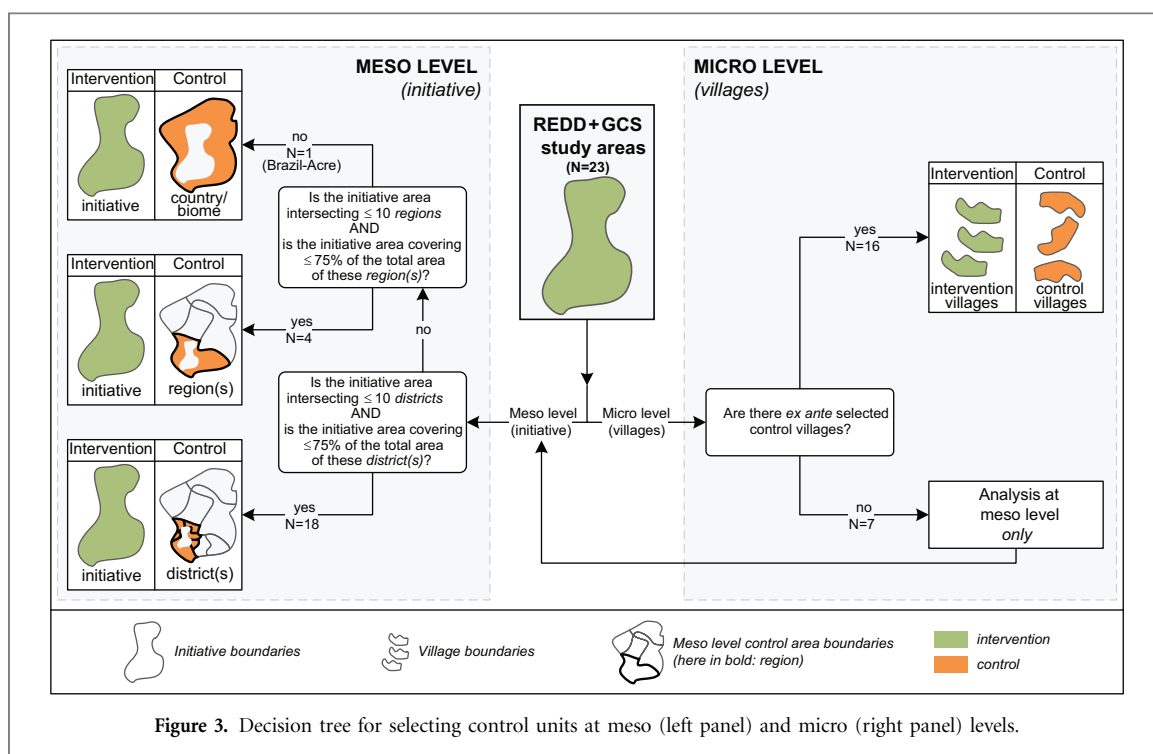


Figure 3. Decision tree for selecting control units at meso (left panel) and micro (right panel) levels.

(here: 2001) up until the intervention started, n_a and n_b the number of years in respectively the *after* and *before* period. A BA score of -2 thus means that the average annual deforestation rate in the intervention area decreased by 2% points when compared to pre-intervention years.

When including control areas in the assessment, the BACI score β is calculated as follows:

$$\text{BA score } \beta = (\bar{x}_{AI} - \bar{x}_{BI}) - (\bar{x}_{AC} - \bar{x}_{BC})$$

$$\text{with } \bar{x}_{AI} = \frac{1}{n_a} \sum_{i=1}^{n_a} x_i, \dots \text{etc.} \quad (2)$$

Here, \bar{x}_{AC} and \bar{x}_{BC} represent the average annual deforestation rates in the control areas in the *after* and *before* period, respectively. β thus scores performance in the intervention area as compared to its control area. A *negative* β indicates a greater reduction or lower rise in deforestation in the intervention area than in the control area, and thus a *positive* REDD+ impact. We calculate the BACI scores β at both meso and micro levels (see next section and figure 3).

2.4. Levels of analysis: initiative and villages

To successfully assess the impacts of REDD+, cross-scale integration is needed (de Sassi *et al* 2015). We use two units of analysis for the intervention area: initiative boundaries (meso level) and intervention village boundaries (micro level), as not all villages within any given initiative area were subject to the same suite of interventions, and thus were not ‘treated’ with the same intensity by implementers. For the meso level analysis, we used the site boundaries of all 23 REDD+ initiatives in the sample. Our control units at

this level differ depending on the size of the initiative. Generally, they consist of the corresponding next higher jurisdictional level (left panel, figure 3), i.e. either districts (18 cases for smaller REDD+ projects), region (four cases for district-level initiatives and larger REDD+ projects) or biome (one state-level jurisdictional program in the Brazilian Amazon)¹⁰.

For the micro level analysis (right panel, figure 3), we focused on 16 of the 23 REDD+ initiatives, known as ‘intensive sites’ in the GCS, where representative control villages were selected based on matched reported percent forest cover, deforestation pressures, market accessibility and socioeconomic factors from an *ex ante* rapid rural appraisal (Sunderlin *et al* 2016). Hence, for the seven sites without matched control villages, we performed the BA and BACI analysis at the meso level only.

Village boundaries were made spatially explicit to reflect the area influenced by villagers. Since the concept of ‘village’ varies by country, and village boundary data were sometimes unavailable, spatial boundaries were compiled to adequately reflect local conditions. These boundaries were either provided by the government; provided by the REDD+ proponents; geo-referenced by field researchers; or obtained by buffering household points (appendix A).

¹⁰ In 17 cases, the intersecting districts were used as the control unit. District is defined as the jurisdictional level below region, which corresponds to the *municipality* in Brazil; *district* in Peru, Tanzania and Vietnam; *department* in Cameroon; and *regency* in Indonesia. In five cases, the region that overlaps with the initiative was used as the control unit. Region is defined as the first subnational jurisdictional level below the country, which is called *state*, *department* and *province* in respectively Brazil, Peru and Indonesia. In the case of Acre’s State System of Incentives for Environmental Services in Brazil, which is the largest initiative in our sample, the area of the Brazilian Amazon biome was used as the control unit.

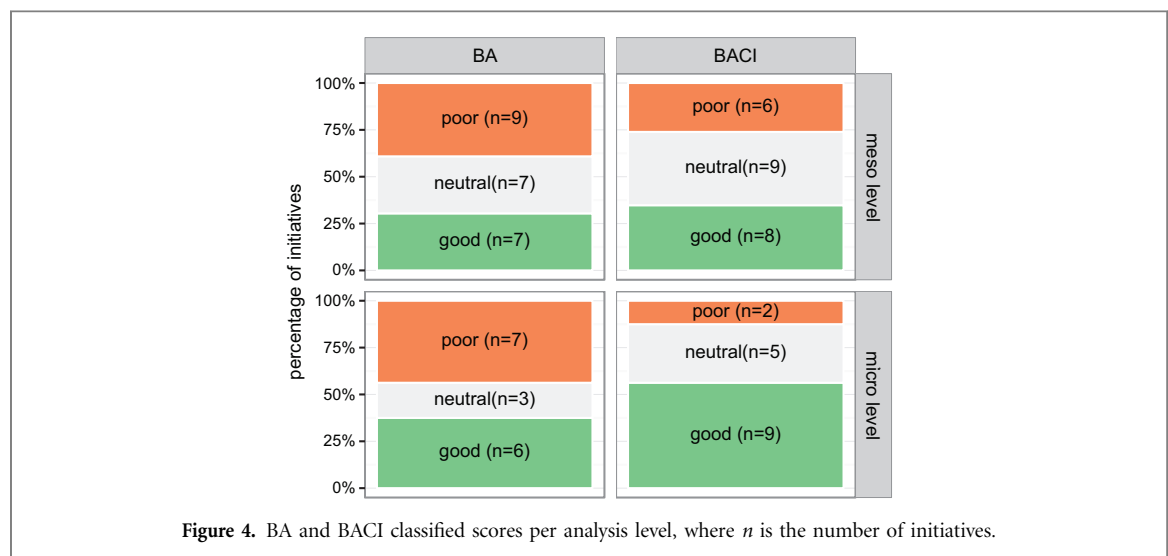


Table 1. Summary statistics.

Level	Variable	Explanation	n	min.	max.	mean	median
both	start year	start year of the initiative	23	2006	2013	2009	2009
both	n_a	years in <i>after</i> period	23	2	9	6	6
both	n_b	years in <i>before</i> period	23	5	12	8	8
meso	α	BA score (in initiative area)	23	-0.903	0.588	0.043	0.083
meso	β	BACI score	23	-1.184	0.315	-0.085	0.002
micro	α	BA score (in intervention villages)	16	-2.139	0.669	-0.271	0.048
micro	β	BACI score	16	-2.277	2.827	-0.449	-0.466

3. Results

3.1. General results

Table 1 shows¹¹ the summary statistics of the main variables introduced in section 2.3.

The results of the BA α and BACI β performance scores were grouped into *good*, *neutral* and *poor*¹², where a good score means a relative reduction in tree cover loss over time (BA, BACI) and/or compared to the control area (BACI) (figure 4).

First, we compare results from the two aggregation levels. At the meso (initiative) level, the median scores for both approaches (BA and BACI) are close to zero (table 1), meaning that there is no substantial change in deforestation rates between the two periods across the sample as a whole. At the micro (village) level, however, the scores are typically *lower* when compared to the results at meso level (i.e. *better* scores in terms of reduced deforestation rates)¹³.

Apparently, the interventions thus had less impact at the more aggregated level. This finding could be due to interventions targeting only a few villages (including the ones studied here) within the site or within-site leakage from treated to untreated villages, which would lower the scores at the meso level.

Second, we compare the two assessment methods. The BA scores (α) range from -2.139 (good performance) to 0.669 (poor) and the BACI scores (β) range from -2.277 (good) to 2.827 (poor). The BACI scores are typically *lower* than the BA scores at both meso and micro levels. Hence, the intervention areas tend to outperform the control areas, regardless of the overall trend in annual deforestation rates over time. Yet, median micro deforestation declines more in intervention than in control areas (median BACI score of -0.466), indicating slightly better REDD+ performance at lower aggregations. In turn, most good BACI scores at meso levels represent cases of increased deforestation trends though these increases were generally lower than in control areas.

3.2. Individual BA and BACI scores

To better understand the methodological differences, in this section we examine specific scenarios. Table 2 shows the occurrences of the prevailing factors that affect the BA and BACI scores, which we explain in more detail below.

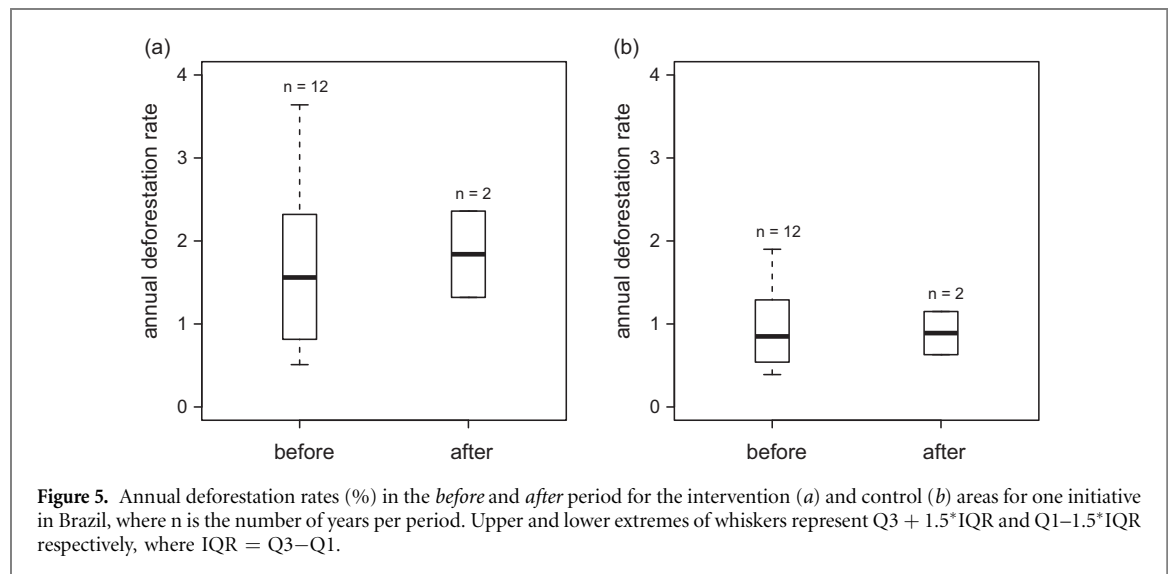
¹¹ See appendix table B1 for an extended version of the summary statistics.

¹² When grouping the scores, the following thresholds were used: good ≤ -0.1 ; $-0.1 > \text{neutral} < 0.1$; and poor ≥ 0.1 . We tested different cut-offs ranging from $(-)-0.05$ to $(-)-0.5$ which all led to similar conclusions, so for illustrative reasons, we decided to use 0.1. Scores close to zero are more likely to be influenced by uncertainties in the data than by a clear direction in performance.

¹³ These results are not influenced by the difference in sample size between the meso and micro level (appendix figure C1).

Table 2. Occurrences per analysis level of factors affecting the BA and BACI scores.

Level	No. cases ^a	Bias in <i>before</i> period ^b	Low absolute deforestation ^a	Peak years ^a	Outperforming control area ^b	Clear comparative performance ^a
Meso	23	7	9	16	1	5
Micro	16	2	8	13	1	1

^a Relevant for both BA & BACI.^b Relevant for BACI only.

3.2.1. Bias in the *before* period

To confidently attribute changes (or lack thereof) to REDD+ activities in the *after* period, tree-cover loss patterns for intervention and control areas should have been similar in the *before* period (figure 2). Yet, two sample t-tests show that in five meso cases, and in two sites at both levels, significant differences in the *before* period influenced the resulting BACI scores (table D1). One such case is shown in figure 5 where meso-level *before* deforestation rates in the initiative area exceeded those in the corresponding control districts.

3.2.2. Low absolute deforestation

For four meso-level cases, three micro cases, and five sites at both levels, median annual deforestation was less than 100 ha in absolute terms. Here, small year-to-year deviations in deforestation can determine the BA and BACI scores. Furthermore, many of these cases correspond to forest change maps where marked tree cover loss speckles may reflect degradation, climatic effects, or input data errors. We should thus be cautious in drawing conclusions from the corresponding scores, which might be driven more by tree cover data uncertainty than factual changes in deforestation dynamics.

3.2.3. Peak year

Single years of exceptionally high tree-cover loss (for intervention or control, *before* or *after*) can heavily

influence our target variable of *mean* annual deforestation for BA and BACI scores alike.

A peak is defined as an observation above the upper quartile. A post-intervention peak might flag failure to target big driver(s) of deforestation, but could also have natural causes. A peak in the control area in the *before* period and a peak in the intervention area in the *after* period (and vice versa) can cancel each other out when having the same magnitude. Only seven meso-level cases and three micro-level cases showed no peaks in the intervention or control areas in the period 2001–2014. We checked the robustness of the BA and BACI scores by recalculating the scores without peak years and recorded the shifts from one category (good or poor) to the opposite (table 3, in bold). The majority of the scores do not shift categories (grey numbers). In one case (meso level, BACI approach), the performance score would change from good to poor if the peak years were excluded from the analysis.

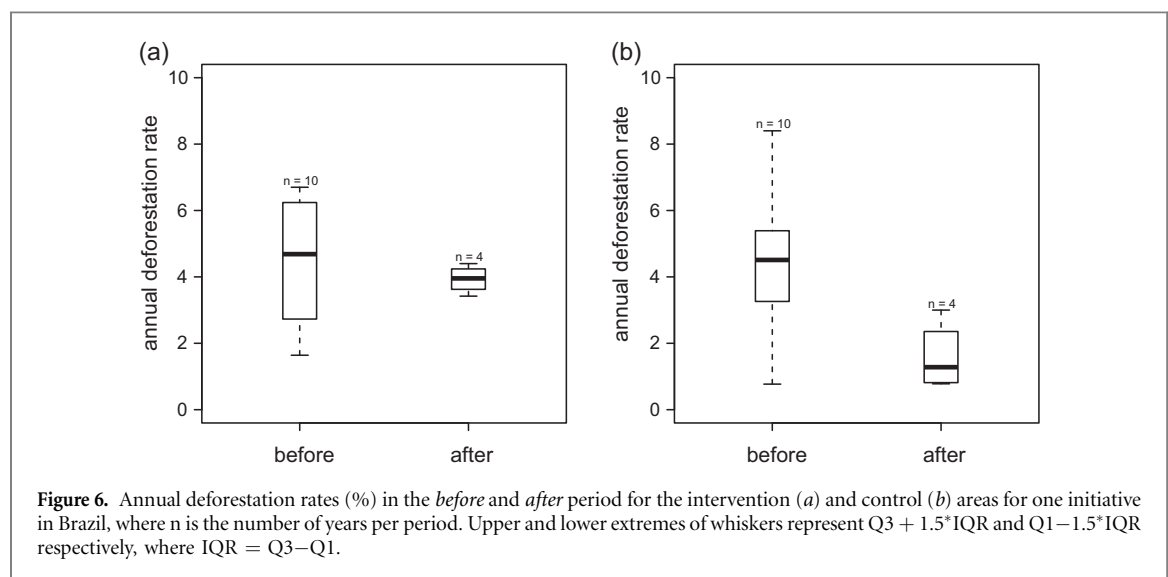
3.2.4. Control area outperforms intervention area

Using the BACI method, good REDD+ performance can only be achieved if deforestation is reduced more in the intervention than in the control area(s). One meso-level (figure 6) and one micro-level case show good BA scores, but poor BACI scores, because control areas improved even more. In those cases, the slowdown in deforestation might have occurred even without the REDD+ intervention (e.g. due to commodity prices or national policies).

Table 3. Evaluating BA and BACI score robustness to peak year influence^a.

BA approach					BACI approach						
Meso level	Original score	excluding peak year			Original score	Original score	excluding peak year				
		good	neutral	poor			good	neutral	poor		
		good	1	3			0	good	3	1	1
		neutral	0	4			1	neutral	1	5	1
		poor	0	1	6	poor	0	2	2		
Micro level	Original score	excluding peak year			Original score	Original score	excluding peak year				
		good	neutral	poor			good	neutral	poor		
		good	5	1			0	good	8	1	0
		neutral	0	1			0	neutral	0	2	1
		poor	0	2	4	poor	0	0	1		

^aBold numbers indicate highly sensitive cases where the particular score shifts from one extreme category (good or poor) to the opposite. Grey numbers indicate robust scores that are not influenced by the peak year.



3.2.5. *Clear comparative performance scores*
Clear comparative performance is defined as a score where we found no bias in the *before* period; no low absolute annual deforestation (median); and where the presence of peak years –if any– did not determine the category of the score. We found three meso level cases, three micro level cases and three sites at both levels with clear comparative performance scores (BA and BACI).

For these clear meso level scores, there were two with good, two with neutral, and two with poor BACI scores. In one site, deforestation increased in its corresponding control area, while deforestation decreased in the intervention area, yielding a good BACI score. One other site had poor BA, but good BACI scores, meaning that deforestation increased during the intervention phase, but less so than in control areas. Yet, arguably, it may be difficult to celebrate this latter case as a victory, since there was still more deforestation in the intervention area in the after period than before the REDD+ initiative started.

For the clear micro level scores, there were four with good, and two with poor BACI scores. At one site,

deforestation decreased in the intervention area, while it increased in the control site, yielding a good BACI score. At another site, deforestation also decreased in the intervention area, while there was a less substantial decrease in the control area, resulting in another good BACI score. The other two good BACI scores represent cases where there was an increase of deforestation in the intervention areas, but less so than in the control areas. The two poor BACI scores represent cases of outperforming control areas similar to those explained in the previous section. That is, one denotes a case where deforestation increased in the intervention areas, while deforestation in the respective control areas increased less. The other is a site where deforestation decreased in the intervention villages (good BA score), but the decrease in the control villages was even stronger.

4. Discussion

We applied BA and BACI approaches at meso and micro levels to assess subnational-level REDD+

Table 4. Main advantages (+) and disadvantages (–) of BA versus BACI assessment approaches, and of using meso versus micro aggregation levels.

Assessment method	
BA approach	BACI approach
+ relatively simple and objective to implement	+ able to discern additionality attributable to the intervention
– susceptible to external factors of influence, i.e. changes in deforestation could wrongfully be attributed to the intervention	– requires careful <i>ex ante</i> control site selection and matching
	– high sensitivity of results to matching method
Aggregation level	
Meso level	Micro level
+ helps understanding trends within context	+ allows more precise comparison between intervention-targeted and non-targeted units
+ may indicate cases of leakage (but further analysis is then still required)	– the notion of village is not universal, and delineating boundaries may be subjective
– defining control areas may be more difficult	– small changes may obscure ‘bigger picture’
	– sensitive to extreme events or single drivers

performance. Both approaches and levels of measurement have advantages and disadvantages for effectiveness assessment (table 4). While the BA approach only considers trend shifts in local deforestation as an indicator for REDD+ performance, the BACI approach adds comparative performance in control areas. In principle, the BACI approach thus enables us to control for changes in deforestation that are unrelated to REDD+ interventions. Where BA measures the direction of change, BACI intends to measure attributive change. This approach, however, requires careful *ex ante* control site matching and selection. The high sensitivity of the results to matching procedures is clear from our results. At seven sites in the meso-level analysis, the jurisdiction used as the control area for the initiative had a significantly different pre-intervention deforestation rate compared with the initiative. Although meso-level assessment puts forest changes observed in the initiative area in a wider context, selecting a suitable control area (i.e. districts, region, or country) is not straightforward, since ideally these control areas should be subject to all of the same time-varying factors as the intervention areas.

Assessing performance at the micro level allows for more precise comparison between targeted and non-targeted villages. Yet, as the notion of village is not universal, delineating village boundaries can turn out to be a subjective process, and small (absolute) forest changes at the village level may wrongfully be interpreted as equivalent to large (absolute) forest changes at higher levels. Moreover, matching intervention and control villages is challenging. At two sites, in our micro-level analysis, baseline deforestation rates in the intervention villages and their control areas were significantly different, which resulted in uninformative BACI scores. For the village matching in GCS, our matched samples of intervention and control villages had statistically similar means across a range of characteristics as later measured in a village

survey (Sills *et al* 2017). Still, the percent forest cover variable used in the matching was based on reported and not observed values, because global comparative satellite data for all sites was not available when the initial matching was performed in 2010. This choice clearly had implications for outcomes subsequently measured through the use of spatial data. Due to recent developments in the remote sensing domain, *ex ante* village matching could now be based on annual tree cover loss data from satellite data instead of reported forest cover loss from cost- and labour-intensive field studies. Although the BACI approach has strong analytical advantages, the sensitivity of results to control selection cannot be overstated.

Independent of approach, we found slightly better performance at the micro level compared to the meso level, possibly reflecting both a higher local treatment intensity, and more occurrence of confounding factors at higher scales, as well as leakage (relocated deforestation activities) from the intervention to control areas. Still, only four sites¹⁴ had both a good BACI score and were not influenced by factors like control area bias, low absolute deforestation and peak years.

The overall underwhelming performance of the studied initiatives could be due to a host of factors. First, performance scores are highly sensitive to cases with a late start year, and one could question how much REDD+ impact is reasonable to expect in the early years of initiative implementation. That is, multiple sites only had a couple of years of *after* observation. Furthermore, funding has been a major constraint for REDD+, meaning that interventions may not have been rolled out in the intensity originally planned (Sunderlin *et al* 2015). Short time spans combined with limited funding would naturally lead

¹⁴ Two sites at micro level, and two sites at both meso and micro level.

to less effective ‘treatment’, which may explain underperformance. Second, we did not consider forest degradation, which contributes to forest-based emissions considerably (Lambin *et al* 2003, Putz *et al* 2008, Nepstad *et al* 1999) and is the focus of REDD+ interventions at many sites (e.g. improved cooking stoves in Tanzania, sustainable forest management in Peru, etc.) (Sills *et al* 2014). While removals due to selective logging, undergrowth fires and fuelwood collection cannot yet be clearly detected by remote sensing based methods (Wertz-Kanounnikoff *et al* 2008), substantial progress has been made in recent years for measuring areas affected by forest degradation (De Sy *et al* 2012, GOF-C-GOLD 2016). The dataset used in this study is unable to identify (reductions in) forest degradation, so any success regarding the second ‘D’ of REDD+ would have been missed here. Third, we only considered change in forest *loss* as proxy for the carbon impact of REDD+ and did not include forest *gain*, i.e. carbon stock enhancements that are integral to REDD+. Indeed, at several sites in the sample, restoration activities are a key part of the overall REDD+ strategy, but would also need more time to become significant and measurable. Finally, possibly the REDD+ proponents did not always effectively target the main driver(s) of deforestation at their sites, which may genuinely affect deforestation outcomes. For instance, most focus their efforts on smallholders, but sometimes these are not the main agents of deforestation, such as in some sites in Brazil and Indonesia (appendix 5 of Sills *et al* 2014, Sunderlin *et al* 2015). This prioritization of interventions targeting smallholders could also explain why we found slightly better results at the village than at the site level. However, as a general caveat, both BA and BACI methods work better with longer timeframes, and with *before* and *after* periods that are approximately equal. Future analysis is thus needed to understand the longer-term impacts of REDD+ at these sites and to better understand why impact varies across initiatives, taking into account the variation in both treatment and context.

5. Conclusion

Much early REDD+ progress has been through the implementation of subnational initiatives, yet we know very little about their carbon effectiveness. In this paper, we compared two approaches for assessing the effectiveness of 23 REDD+ initiatives in six countries through: (1) analysing trend development (BA approach); and (2) including control areas to correct for confounding factors (BACI approach).

We conclude that the more local the scale of performance assessment, the more relevant is the use of the BACI approach. Although BA is a good starting point for assessment, it is not able to distinguish between the REDD+ effect and confounding factors.

BACI allows getting closer to attribution by removing the confounding influence of background dynamics, yet the results are only as good as the choice of control areas. While this remains a key challenge, new global forest datasets allow for improved control area matching and selection.

Nevertheless, there may be local situations where a BA approach, with its focus on the direction of change, is useful. For instance, in cases where BA scores flag *poor* and BACI scores *good* performance, due to increases in deforestation being higher in control areas than in intervention areas, the BA score makes clear that deforestation is still increasing, just less rapidly than would have occurred in the absence of REDD+. The *poor* BA score flags that the goal to reduce deforestation has become more distant (change has overall gone into the wrong direction); the *good* BACI score reflects that under a ‘no intervention’ counterfactual things would have been even worse (positive attribution). Conversely, in situations of generalized positive changes, BA scores alone risk painting a rosier picture than what could reasonably be attributed to the REDD+ intervention.

The BA and BACI assessment approaches used in our research both highlight overall minimal impact of REDD+ in reducing deforestation thus far. This could be due to the slow implementation of REDD+ interventions and low treatment density; proponents focussing primarily on smallholders instead of other important drivers; and/or our analytical focus on deforestation only, without examining degradation or reforestation. Furthermore, we did not examine specific REDD+ intervention mixes and strategies applied at different sites. To better understand what works (or not) in which contexts, linking the performance assessment results to the (types of) interventions would be an important next step.

Results-based payments for REDD+ will use conventional reference level approaches at the national level, yet there is clearly a need to understand the carbon effectiveness of local REDD+ interventions. Indications of which combinations of intervention mixes have shown to be more or less effective under variable contextual circumstances may provide valuable pointers for selective upscaling options to national REDD+ policies. Countries should seek ways to incorporate results from local level monitoring into their national reporting systems, since overall REDD+ impact depends on land use decisions on the ground.

Acknowledgments

This research is part of the Global Comparative Study (GCS) on REDD+ of the Center for International Forestry Research (CIFOR) (www.cifor.org/gcs) with funding support from the German International Climate Initiative, Norwegian Agency for Development Cooperation, Australian Department of Foreign Affairs

and Trade, European Commission (EC), UK Department for International Development, and CGIAR Research Program on Forests, Trees and Agroforestry (FTA) Programme.

We would like to thank all CIFOR researchers and affiliates who helped defining, measuring and compiling village and initiative boundaries. We are grateful to Louis Verchot for helpful discussions throughout the process and thank two anonymous reviewers for their helpful comments.

Appendices

Appendix A. Village boundary delineation

In Tanzania, REDD+ proponents provided official village boundary data. In Indonesia, field researchers

for subsequent digitalization in ArcGIS to delineate village boundaries. In Brazil, village associations are social rather than spatial units, so village boundaries were created through either spatializing social constructs of villages in the field or buffering and merging georeferenced household points. In Vietnam, the lowest official jurisdictional level is commune, which consists of a set of villages, so village boundaries were also estimated using a buffer around household points. In both cases, additional official spatial data (e.g. agrarian reform settlement project boundaries in Brazil, and district limits in Vietnam) were used to inform village extent.

Appendix B. General results extended

Table B1. General results table extended.

Level	Variable	Explanation	<i>n</i>	min.	max.	mean	median
both	start year	start year of the initiative	23	2006	2013	2009	2009
both	n_a	years in <i>after</i> period	23	2	9	6	6
both	n_b	years in <i>before</i> period	23	5	12	8	8
meso	\bar{x}_{AI}	Average annual deforestation rate in intervention area during <i>after</i> period	23	0.037	1.84	0.522	0.43
meso	\bar{x}_{BI}	Average annual deforestation rate in intervention area during <i>before</i> period	23	0.021	1.62	0.479	0.37
meso	\bar{x}_{AC}	Average annual deforestation rate in control area during <i>after</i> period	23	0.065	1.93	0.664	0.605
meso	\bar{x}_{BC}	Average annual deforestation rate in control area during <i>before</i> period	23	0.048	1.62	0.536	0.465
meso	α	Before-After score (in intervention area)	23	−0.903	0.588	0.043	0.083
meso	β	BACI score	23	−1.184	0.315	−0.089	−0.008
micro	\bar{x}_{AI}	Average annual deforestation rate in intervention area during <i>after</i> period	16	0.073	3.933	0.928	0.605
micro	\bar{x}_{BI}	Average annual deforestation rate in intervention area during <i>before</i> period	16	0.068	4.514	1.199	0.489
micro	\bar{x}_{AC}	Average annual deforestation rate in control area during <i>after</i> period	16	0.106	2.479	1.023	0.862
micro	\bar{x}_{BC}	Average annual deforestation rate in control area during <i>before</i> period	16	0.073	4.993	0.845	0.486
micro	α	Before-After score (in intervention area)	16	−2.139	0.669	−0.271	0.048
micro	β	BACI score	16	−2.277	2.827	−0.449	−0.466

used boundaries provided by the government for the study villages as a base for verification with key informants. Village boundaries were later modified through digitalization in ArcGIS/Google Earth based on local knowledge of village limits. In Peru, proponents and other partners provided official spatial data for study villages at the Ucayali site and individual Brazil nut concession boundaries for the Madre de Dios site. Village units in Madre de Dios were constructed by aggregating concessions whose owners were members of the same social association and/or in close spatial proximity to one another. In Cameroon, field researchers geo-referenced a few known borders with the assistance of key informants

Appendix C. BA and BACI classified scores for intensive sites only

Figure C1 reports results at both the meso and micro level for the 16 ‘intensive’ sites only, which as described in section 2.4 include both intervention and matched control villages. These results are mostly consistent with the results presented in figure 4, confirming our finding (presented in section 3.1) that performance generally looks better at the micro than at the meso level (i.e. evaluating REDD+ at the micro level makes it appear more effective in terms of reducing deforestation). Figure C1 confirms that this finding is not due to the difference in sample size for the meso and micro level analysis reported in figure 4.

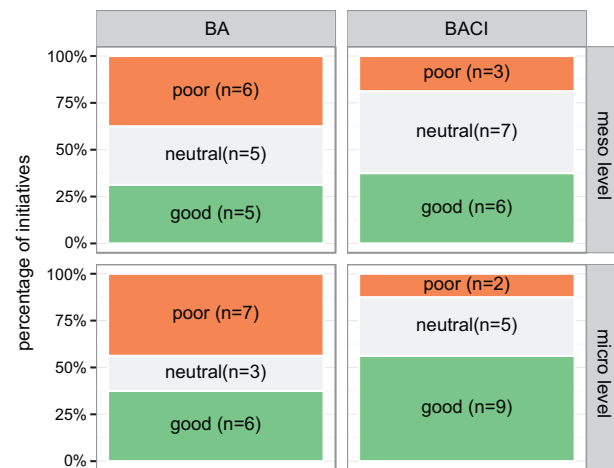


Figure C1. BA and BACI classified scores with equal sample sizes for both levels.

Appendix D. Test results for bias detection

Table D1. Bias test results—Levene's and *t*-tests for discovering significantly differing deforestation trends between the intervention and control area in the *before* period.

	Meso level				Micro level			
	<i>p</i> -value Levene's test	<i>p</i> -value two sample <i>t</i> -test ^a	<i>p</i> -value Welch <i>t</i> -test ^b	Possible bias ^c	<i>p</i> -value Levene's test	<i>p</i> -value two sample <i>t</i> -test ^a	<i>p</i> -value Welch <i>t</i> -test ^b	Possible bias ^c
Brazil-Acre	0.4413	0.8487	N/A	FALSE	0.1062	0.1359	N/A	FALSE
Brazil-Cotriguacu	0.75	0.4233	N/A	FALSE	0.546	0.6723	N/A	FALSE
Brazil-Transamazon	0.0366	N/A	0.0450	TRUE	0.7074	0.5399	N/A	FALSE
Brazil-SFX	0.0268	N/A	0.0001	TRUE	0.0004	N/A	0.0020	TRUE
Brazil-Bolsa Floresta	0.1214	0.0046	N/A	TRUE	N/A	N/A	N/A	N/A
Brazil-Jari Amapa	0.0036	N/A	0.0203	TRUE	N/A	N/A	N/A	N/A
Peru-Madre de Dios	0.01	N/A	0.0001	TRUE	0.2856	0.0267	N/A	TRUE
Peru-Ucayali	0.0001	N/A	0.0004	TRUE	0.432	0.0801	N/A	FALSE
Cameroon-SE Cameroon	0.0611	0.7418	N/A	FALSE	0.1201	0.9229	N/A	FALSE
Cameroon-Mt Cameroon	0.0037	N/A	0.0726	FALSE	0.0129	N/A	0.1361	FALSE
Tanzania-Shinyanga	0.0857	0.1132	N/A	FALSE	0.0081	N/A	0.4008	FALSE
Tanzania-Kilosa	0.2865	0.3505	N/A	FALSE	0.2248	0.5049	N/A	FALSE
Tanzania-Zanzibar	0.8768	0.9332	N/A	FALSE	N/A	N/A	N/A	N/A
Tanzania-Kigoma	0.6068	0.4298	N/A	FALSE	N/A	N/A	N/A	N/A
Tanzania-Mpingo	0.6497	0.2745	N/A	FALSE	N/A	N/A	N/A	N/A
Tanzania-Lindi	0.3748	0.4095	N/A	FALSE	N/A	N/A	N/A	N/A
Indonesia-Ulu Masen	0.0072	N/A	0.0068	TRUE	0.4343	0.7362	N/A	FALSE
Indonesia-KCCP	0.1983	0.6738	N/A	FALSE	0.4354	0.6332	N/A	FALSE
Indonesia-KFCP	0.4693	0.9611	N/A	FALSE	0.2778	0.5318	N/A	FALSE
Indonesia-Rimba Raya	0.9571	0.2019	N/A	FALSE	N/A	N/A	N/A	N/A
Indonesia-Katingan	0.4841	0.0716	N/A	FALSE	0.0744	0.4623	N/A	FALSE
Indonesia-TNC within BFCP	0.2803	0.663	N/A	FALSE	0.539	0.5952	N/A	FALSE
Vietnam-Cat Tien	0.8567	0.8992	N/A	FALSE	0.074	0.2737	N/A	FALSE

^a Equal variances assumed.

^b Unequal variances assumed.

^c Using confidence level of 0.95.

References

- Asner G P *et al* 2013 High-fidelity national carbon mapping for resource management and REDD+ *Carbon Balance Manage.* **8** 7
- Avitabile V *et al* 2016 An integrated pan-tropical biomass map using multiple reference datasets *Glob. Change Biol.* **22** 1406–20
- Baccini A *et al* 2012 Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps *Nat. Clim. Change* **2** 182–5
- Bellot F-E, Bertram M, Navratil P, Siegert F and Dotzauer H 2014 *The High-resolution Global Map of 21st-century Forest Cover Change from the University of Maryland ('Hansen Map') is Hugely Overestimating Deforestation in Indonesia* (Jakarta: FORCLIME Forests and Climate Change Programme)
- Duchelle A E, de Sassi C, Jagger P, Cromberg M, Larson A M, Sunderlin W D, Atmadja S S, Resosudarmo I A P and Pratama C D 2017 Balancing carrots and sticks in REDD+: Implications for social safeguards *Ecol. Soc.* submitted
- De Sy V, Herold M, Achard F, Asner G P, Held A, Kellndorfer J and Verbesselt J 2012 Synergies of multiple remote sensing data sources for REDD+ monitoring *Curr. Opin. Environ. Sustain.* **4** 696–706
- de Sassi C, Joseph S, Bos A B, Duchelle A E, Ravikumar A and Herold M 2015 Towards integrated monitoring of REDD+ *Curr. Opin. Environ. Sustain.* **14** 93–100
- FAO 2000 *FRA 2000 on definitions of forest and forest change* (Rome)
- GOFC-GOLD 2016 A sourcebook of methods and procedures for monitoring and reporting anthropogenic greenhouse gas emissions and removals associated with deforestation, gains and losses of carbon stocks in forests remaining forests, and forestation *GOFC-GOLD Report version COP22-1* (Wageningen, the Netherlands)
- Hansen M C *et al* 2013 High-resolution global maps of 21st-century forest cover change *Science* **342** 850–3
- Herold M and Skutsch M M 2009 Measurement, reporting and verification for REDD+: objectives, capacities and institutions *Realising REDD+: National Strategy and Policy Options* ed A Angelsen (Bogor, Indonesia: Center for International Forestry Research (CIFOR)) pp 84–100 (<https://doi.org/10.17528/cifor/002871>)
- Hijmans R J 2016 *raster: Geographic Data Analysis and Modeling*. R package version 2.5-8. (<https://cran.r-project.org/package=raster>)
- Imbens G W and Wooldridge J M 2009 Recent developments in the econometrics of program evaluation *J. Econ. Lit.* **47** 5–86
- Jagger P, Sills E, Lawlor K and Sunderlin W D 2010 *A Guide to Learning about Livelihood Impacts of REDD+* (Bogor, Indonesia: Center for International Forestry Research (CIFOR)) (<https://doi.org/10.17528/cifor/003283>)
- Lambin E F, Geist H J and Lepers E 2003 Dynamics of land-use and land-cover change in tropical regions *Annu. Rev. Environ. Resour.* **28** 205–41
- Nepstad D C *et al* 1999 Large-scale impoverishment of Amazonian forests by logging and fire *Nature* **398** 505–8
- Putz F E, Zuidema P A, Pinard M A, Boot R G A, Sayer J A, Sheil D, Sist P, Elias and Vanclay J K 2008 Improved tropical forest management for carbon retention *PLoS Biol.* **6** 1368–9
- Ravikumar A, Larson A M, Duchelle A E, Myers R and Gonzales Tovar J 2015 Multilevel governance challenges in transitioning towards a national approach for REDD+: evidence from 23 subnational REDD+ initiatives *Int. J. Commons* **9** 909
- Romijn E, Lantican C B, Herold M, Lindquist E, Ochieng R, Wijaya A, Murdiyarso D and Verchot L 2015 Assessing change in national forest monitoring capacities of 99 tropical countries *Forest Ecol. Manage.* **352** 109–23
- Saatchi S S *et al* 2011 Benchmark map of forest carbon stocks in tropical regions across three continents *Proc. Natl Acad. Sci. USA* **108** 9899–904
- Sills E O, Atmadja S, de Sassi C, Duchelle A E, Kweka D, Resosudarmo I A P and Sunderlin W D ed 2014 REDD+ on the ground: A case book of subnational initiatives across the globe (Bogor, Indonesia: Center for International Forestry Research (CIFOR)) (<https://doi.org/10.17528/cifor/005202>)
- Sills E O, de Sassi C, Jagger P, Lawlor K, Miteva D A, Pattanayak S K and Sunderlin W D 2017 Building the evidence base for REDD+: study design and methods for evaluating the impacts of conservation interventions on local well-being *Glob. Environ. Change* **43** 148–60
- Simonet G, Karsenty A, de Perthuis C, Newton P and Schaap B 2015 *REDD+ Projects in 2014: An Overview Based on a New Database and Typology* (Paris: Chaire Economie du Climat)
- Smith E P 2002 BACI design *Encyclopedia of Environmetrics* ed A H El-Shaarawi and Piegorisch (Chichester: Wiley) pp 141–8
- Sunderlin W D *et al* 2015 REDD+ at a critical juncture: assessing the limits of polycentric governance for achieving climate change mitigation *Int. Forest Rev.* **17** 400–13
- Sunderlin W D *et al* 2016 *Technical Guidelines for Research on REDD+ Subnational Initiatives* (Bogor, Indonesia: Center for International Forestry Research (CIFOR)) (<https://doi.org/10.17528/cifor/006009>)
- UNFCCC 2015 *Adoption of the Paris Agreement* (Paris: United Nations)
- Verchot L V V, Anitha K, Romijn E, Herold M and Hergoualc'h K 2012 Emissions factors: converting land use change to CO₂ estimates *Analysing REDD+: Challenges and Choices* ed Angelsen (Bogor, Indonesia: Center for International Forestry Research (CIFOR)) pp 261–78 (<https://doi.org/10.17528/cifor/003805>)
- Wertz-Kanounnikoff S, Verchot L V, Kanninen M, Murdiyarso D and Angelsen A 2008 How can we monitor, report and verify carbon emissions from forests? *Moving Ahead with REDD: Issues, Options and Implications* ed Angelsen (Bogor, Indonesia: Center for International Forestry Research (CIFOR)) pp 87–98 (<https://doi.org/10.17528/cifor/002601>)