Towards performance assessment of subnational forest-based climate change mitigation initiatives

Astrid Brigitte Bos

Propositions

1. *Ex post* effectiveness assessment requires *ex ante* study design.
   (this thesis)
2. A *doing the right things* verdict necessitates a *doing things right* evaluation.
   (this thesis)
3. Medical research on blood donors is like environmental research on nature conservation communities; both are inevitably affected by selection bias.
4. Prioritising writing over visualisation skills in academic curricula hampers effective science communication.
5. Twitter thwarts truthfulness at scientific conferences.
6. Using football pitches as unit of area bolsters people's spatial ignorance.
7. Women groups in isolation sustain gender inequality.

Propositions belonging to the thesis, entitled

Towards performance assessment of subnational forest-based climate change mitigation initiatives

Astrid Brigitte Bos

Wageningen, 17 January 2020

Towards performance assessment of subnational forest-based climate change mitigation initiatives

Astrid Brigitte Bos

Thesis committee

Promotor:
Prof. Dr M. Herold
Professor of Geo-information Science and Remote Sensing
Wageningen University & Research

Co-promotors:
Dr V. De Sy
Researcher, Laboratory of Geo-information Science and Remote Sensing
Wageningen University & Research

Dr A.E. Duchelle
Senior Scientist, Climate Change, Energy & Low Carbon Development
Center for International Forestry Research (CIFOR), Bogor, Indonesia

Other members:
Prof. Dr L.G. Hein, Wageningen University & Research
Prof. Dr I.J. Visseren-Hamakers, Radboud University Nijmegen
Dr A. Lotsch, World Bank Forest Carbon Partnership Facility, Hà Nội, Vietnam
Dr B. Mora, CS Communication & Systèmes, Toulouse, France

This research was conducted under the auspices of the C.T. de Wit Graduate School
of Production Ecology & Resource Conservation (PE&RC)

Towards performance assessment of subnational forest-based climate change mitigation initiatives

Astrid Brigitte Bos

Thesis
submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Friday 17 January 2020
at 4 p.m. in the Aula.

Astrid Brigitte Bos
Towards performance assessment of subnational forest-based climate change mitigation initiatives
180 pages.

PhD thesis, Wageningen University, Wageningen, NL (2020)
With references, with summaries in English and Dutch

ISBN 978-94-6395-214-9
DOI https://doi.org/10.18174/507034

Summary

The central role of forests in climate change mitigation, as recognized in the Paris agreement, makes it important to develop and test methods for monitoring and evaluating the carbon effectiveness of efforts that aim to reduce emissions from deforestation and forest degradation and enhance carbon stocks (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 and to assess social synergies or trade-offs in parallel.

The study discussed in chapter 2 develops (1) a Before-After-Control-Intervention (BACI) method to assess the effectiveness of these REDD+ initiatives; compares (2) the results at the meso (initiative) and micro (village) scales; and compares (3) BACI with the simpler Before-After (BA) results. Our study covers 23 subnational REDD+ initiatives in Brazil, Peru, Cameroon, Tanzania, Indonesia and Vietnam. Annual tree cover loss was used as proxy for deforestation. Data was aggregated 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.

Chapter 3 delves deeper into the forest cover change data required for performance assessment of REDD+. Innovations in remote sensing and forest monitoring provide ever-increasing levels of coverage, spatial and temporal detail, and accuracy. More global products and advanced open-source algorithms are becoming available. Still, these datasets and tools are not always consistent or complementary, and their suitability for local REDD+ performance assessments remains unclear. These assessments should, ideally, be free of any confounding factors, but performance estimates are affected by data uncertainties in unknown ways. In this chapter we analyse (1) differences in accuracy between datasets of forest cover change; (2) if and how combinations of datasets can increase accuracy; and we demonstrate (3) the effect of (not) doing accuracy assessments for REDD+ performance measurements. In this chapter we cover five local REDD+ initiatives in four countries across the tropics. We compared accuracies of a readily available global forest cover change dataset and a locally modifiable open-source break detection algorithm. We applied human interpretation validation tools using Landsat Time Series data and high-resolution optical imagery. Next, we assessed whether and how combining different datasets can increase accuracies using several combination strategies. Finally, we demonstrated the consequences of using the input datasets for REDD+ performance assessments with and without considering their accuracies and uncertainties. Estimating the amount of deforestation using validation samples could substantially reduce uncertainty in REDD+ performance assessments. We found that the accuracies of the various data sources differ at site level, although on average neither one of the input products consistently excelled in accuracy. Using a combination of both products as stratification for area estimation and validated with a sample of high-resolution data seems promising. In these combined products, the expected trade-offs in accuracies across change classes (before, after, no change) and across accuracy types (user's and producer's accuracy) were negligible, so their use is advantageous over single-source datasets. More locally calibrated wall-to-wall products should be developed to make them more useful and applicable for REDD+ purposes. The direction and degree of REDD+ performance remained statistically uncertain, as confidence intervals were overlapping in most cases for the deforestation estimates before and after the start of the REDD+ interventions. Given these uncertainties and inaccuracies and to increase the credibility of REDD+ it is advised to (1) be conservative in REDD+ accounting, and (2) not to rely on results from single currently available global data sources or tools without sample-based validation if results-based payments are intended to be made on this basis.

In chapter 4 drivers of deforestation and forest degradation (DD) and REDD+ interventions were assessed at five sites in Brazil, Peru, Indonesia and Vietnam. Early REDD+ programs and local/subnational projects used various interventions (i.e. enabling measures, disincentives and incentives), implemented by government, the commercial and non-commercial private sector, but are currently understudied vis-à-vis their effectiveness to address site-specific DD drivers. The study design included an integrated assessment of remotely sensed, spatially modelled, and locally reported drivers. First, follow-up land use from high resolution imagery is assessed as proxy for direct deforestation drivers. Second, spatial Random Forest modelling of DD drivers allowed for influence quantification of topographic, climatic and proximity variables at each site. Third, direct and indirect DD drivers were identified from pre-intervention surveys and semi-structured interviews with five REDD+ implementers, 40 villages and 1200 households.

Data gathered included perceived changes in forest cover and quality, and their causes. We found general agreement between observed, modelled and reported local DD drivers, yet some were inadequately addressed by interventions. Intra-site differences in drivers underscores the importance of analysing micro-level DD drivers. Our interdisciplinary approach reveals the complexities of local direct and indirect DD drivers, and the complementarity of remotely sensed, spatially modelled and locally reported methods for driver identification. A better understanding of the alignment between DD drivers and REDD+ interventions is vital for practitioners and policy makers to enhance the effectiveness, efficiency, equity and co-benefits of REDD+ at the local level.

Chapter 5 consists of an exploratory analysis to examine the relationships between the treatment intensity of different types of forest interventions, changes in forest cover (loss) and changes in income and well-being in seventeen subnational sites across the tropics. Although the aim of forest-based climate change mitigation interventions, such as REDD+, is to protect and enhance forests, there are legal, moral and practical reasons for making sure that this objective is achieved while at a minimum not harming, and ideally ensuring benefits for, local people. Information on interventions, household income and perceived well-being was gathered from village and household level interviews from nearly 130 villages and 4,000 households. Global Forest Change data (2000-2018) was used to derive information on forest cover and forest cover change at the village level. Clusters of villages were defined based on similar levels of intervention treatment intensities and deforestation trends to compare pre- and post-intervention characteristics. Villages in the cluster with high treatment intensities and reduced deforestation rates consisted mostly of Brazilian villages. These villages had higher income levels and deforestation rates in the pre-intervention period. In the post-intervention period, these villages were generally associated with an increase in income and its households reported a slightly better level of perceived well-being. No clear differences in outcomes among different intervention types were found, nor were there any indications of pronounced trade-offs between forest conservation and well-being outcomes at all villages and households as a whole, although trade-offs at specific villages and households could not be ruled out. This analysis provides one way of looking at the forest change and well-being outcomes of different forest interventions, and their possible trade-offs. Such information can provide valuable insights for policy makers and practitioners interested in developing REDD+ strategies that can provide both conservation and livelihood benefits.

This thesis contributes to the understanding of the complexities involved in REDD+ performance assessments at the subnational level. In chapter 2, 3 and 4, internal, methodological factors influencing the performance assessment are being addressed, whilst chapter 4 and 5 address the performance assessment outcomes. Hereby, this study contributes to the development of performance assessments themselves as well as to the understanding of the effectiveness of the interventions they intend to evaluate.

Samenvatting

Bossen spelen een centrale rol in het beperken van klimaatverandering, wat erkend is in het akkoord van Parijs. In het afgelopen decennium zijn op subnationaal niveau honderden initiatieven geïmplementeerd die zich ten doel hebben gesteld om koolstofemissies van ontbossing en bosdegradatie te verminderen, en koolstofvoorraden in bossen te vergroten (*Reducing Emissions from Deforestation and Forest Degradation and Enhancing Carbon Stocks* in het Engels, afgekort REDD+). Om te weten of de initiatieven dit doel behalen is het belangrijk om de koolstofemissies te monitoren en te evalueren. Aan de hand van deze praktijkvoorbeelden kunnen we verschillende kwantificatiemethodes testen en vergelijken, en tegelijkertijd onderzoeken hoe (veranderingen in) koolstofemissies samenhangen met lokale sociaal-economische effecten.

Hoofdstuk 2 behandelt een studie waarin (1) een Voor-Na-Controle-Interventie (*Before-After-Control-Intervention* in het Engels, afgekort BACI) methode wordt ontwikkeld; (2) de resultaten op meso (initiatief-) en micro (dorps-) schaal worden vergeleken; en (3) BACI-resultaten worden vergeleken met die van de eenvoudigere Voor-Na (*Before-After* in het Engels, afgekort BA) methode. Deze studie omvat 23 subnationale REDD+ initiatieven in Brazilië, Peru, Kameroen, Tanzania, Indonesië en Vietnam. De jaarlijkse vermindering van de kroonbedekking (*tree cover loss* in het Engels) werd gebruikt als proxy voor ontbossing. Voor het meten van ontbossing werden twee periodes onderscheiden, dat wil zeggen, voor en na de start van elk REDD+ initiatief. Analyse met behulp van controlegebieden ('controle-interventie') suggereert betere REDD+ prestaties dan wanneer enkel een vergelijking tussen ontbossing voor en na de start van het REDD+ initiatief wordt gemaakt. Het REDD+ effect is meer uitgesproken op micro- dan op mesoniveau. BACI vereist echter meer data dan BA en kan mogelijk beïnvloed zijn door verschillen in ontbossingstrends tussen de controle- en interventiegebieden in de pre-interventieperiode. Selectie van de juiste, representatieve controlegebieden is van groot belang, maar dit is op beide schaalniveaus niet vanzelfsprekend. Lage absolute ontbossingsaantallen en piekjaren beïnvloedden zowel onze BA- als BACI-resultaten. In principe is BACI superieur, omdat deze methode rekening kan houden met verstorende factoren. We concluderen dat de

relevantie van een BACI-benadering toeneemt naargelang de REDD+ evaluatie op een meer lokaal niveau uitgevoerd wordt. Over het algemeen vinden we een minimale invloed van REDD+ op het verminderen van ontbossing tot nu toe. Het is belangrijk om dit soort micro- en mesoniveau resultaten op te nemen in nationale rapportages, aangezien de algehele invloed van REDD+ afhankelijk is van beslissingen over landgebruik op lokaal niveau.

Hoofdstuk 3 gaat dieper in op de ruimtelijke kroonbedekkingsdata die nodig zijn voor REDD+ evaluaties. Innovaties op het gebied van aardobservatie en bosmonitoring zorgen voor een steeds groter ruimtelijk bereik, voor meer detail in tijd en ruimte, en voor een grotere nauwkeurigheid. Ook worden er meer producten met wereldwijde dekking aangeboden, alsook geavanceerde open-source algoritmen voor het detecteren en in kaart brengen van ontbossing. Toch zijn deze datasets en hulpmiddelen niet altijd consistent of complementair en blijft het onduidelijk in hoeverre ze geschikt zijn voor lokale REDD+ evaluaties. Deze evaluaties zouden idealiter geen verstorende factoren mogen bevatten, maar onderzoek is nodig om vast te stellen of en hoe resultaten worden beïnvloed door onzekerheden in de data die gebruikt worden als input voor deze evaluaties. Daarom behandelt dit hoofdstuk (1) verschillen in nauwkeurigheid tussen datasets die gebruikt worden om veranderingen in kroonbedekking te meten; (2) of en hoe combinaties van datasets de nauwkeurigheid zouden kunnen verhogen; en (3) laat het zien wat het effect is van het (niet) uitvoeren van nauwkeurigheidstests voor REDD+ evaluaties. Dit hoofdstuk omvat vijf lokale REDD+ initiatieven in vier verschillende tropische landen. We vergeleken de nauwkeurigheid van een gebruiksvriendelijke ontbossingskaart en een lokaal aanpasbaar open-source algoritme voor ontbossingsdetectie. We hebben een visuele validatie toegepast met behulp van Landsat Time Series satellietdata en hoge resolutie beelden. Het schatten van de hoeveelheid ontbossing aan de hand van validatiesteekproeven kan de onzekerheid in REDD+ evaluaties aanzienlijk verminderen. We hebben geconstateerd dat de nauwkeurigheid van de verschillende gegevensbronnen op initiatiefniveau verschilt, hoewel over het algemeen geen van de twee producten de ander in nauwkeurigheid overtrof. Op basis van onze resultaten constateren we dat het combineren van beide producten als stratificatie voor het ramen van ontbossing veelbelovend is wanneer dit samengaat met een visuele validatie aan de hand van een steekproef en met beelden met een hoge resolutie. In deze gecombineerde producten waren de verwachte compromissen in nauwkeurigheid tussen veranderingsklassen (vóór REDD+, na REDD+, geen verandering) en tussen nauwkeurigheidstypes (gerelateerd aan de fouttypen vals-positief en vals-negatief) te verwaarlozen, dus het gebruik van een gecombineerd product biedt voordelen ten opzichte van het gebruik van een enkele dataset. Voor REDD+ gerelateerde analyses moeten meer lokaal gekalibreerde, adequate en toepasbare “wall-to-wall” producten worden ontwikkeld. De richting en de mate van REDD+ prestaties bleven statistisch onzeker, omdat de betrouwbaarheidsintervallen van de ontbossingsramingen van vóór en na de start van REDD+ in de meeste gevallen elkaar overlapten. Gezien deze onzekerheden en onnauwkeurigheden en om de geloofwaardigheid van REDD+ te vergroten, adviseren we om (1) conservatief te zijn in REDD+ resultatenramingen, en (2) niet af te gaan op resultaten van een enkele ontbossingskaart of hulpmiddel zonder steekproefsgewijze validatie, met name wanneer conditionele REDD+ vergoedingen (*results-based payments* in het Engels) gebaseerd zullen worden op dit soort metingen.

Hoofdstuk 4 beschrijft de oorzaken van ontbossing en bosdegradatie (respectievelijk *deforestation* en *forest degradatio*n in het Engels, afgekort DD) en REDD+ interventies op vijf locaties in Brazilië, Peru, Indonesië en Vietnam. De eerste ingevoerde REDD+ programma's en lokale en subnationale projecten omvatten verschillende typen interventies (dat wil zeggen faciliterende maatregelen, belemmerende maatregelen, en positieve prikkels), die geïmplementeerd worden door de overheid en door de commerciële en niet-commerciële particuliere sector. De effectiviteit van deze initiatieven om locatie-specifieke oorzaken van DD aan te pakken is tot nu toe onvoldoende onderzocht. In ons onderzoek integreerden we (1) met satellietdata waargenomen, (2) ruimtelijk gemodelleerde en (3) lokaal gerapporteerde DD oorzaken. Eerst werd, op basis van hoge resolutie beelden, landgebruik na ontbossing gebruikt als proxy voor de directe ontbossingsoorzaak. In het tweede deel werd een ruimtelijk computermodel van DD oorzaken gemaakt, een zogeheten Random Forest model. Hiermee werd voor ieder studiegebied de invloed van verschillende onderliggende topografische, klimatologische en nabijheidsvariabelen gekwantificeerd en gerangschikt. In het derde deel werden directe en onderliggende DD oorzaken geïdentificeerd op basis van informatie uit verschillende enquêtes met vijf REDD+ initiatiefnemers, in 40 groepsgesprekken op dorpsniveau en met 1200 individuele huishoudens. Deze enquêtes gingen onder andere over waargenomen veranderingen in bosareaal en -kwaliteit, en de oorzaken daarvan. Over het algemeen kwamen de waargenomen, gemodelleerde en gerapporteerde lokale DD oorzaken overeen, maar sommige van deze oorzaken werden onvoldoende aangepakt door de REDD+ interventies. Omdat er niet alleen *tussen*, maar ook *binnen* studiegebieden (ruimtelijke) verschillen in DD oorzaken werden gevonden, onderstreept dit het belang van onderzoek naar DD oorzaken op lokaal niveau. Met onze interdisciplinaire aanpak hebben we aangetoond dat het vaststellen van lokale directe en indirecte DD oorzaken een complex proces is. Bovendien hebben we laten zien hoe de aardobservatie met hoge resolutie beelden, ruimtelijke modellen en sociaal-economische enquêtes elkaar kunnen aanvullen om tot een completer beeld van de lokale oorzaken van DD te komen. Een beter begrip van de afstemming tussen DD oorzaken en REDD+ interventies is van vitaal belang voor REDD+ initiatiefnemers en beleidsmakers om de effectiviteit, efficiëntie, billijkheid en nevenvoordelen van REDD+ op lokaal niveau te verbeteren.

Hoofdstuk 5 bestaat uit een verkennende studie waarin de relaties worden onderzocht tussen de behandelingsintensiteit van verschillende soorten bos-gerelateerde interventies, veranderingen in (de reductie van) bosareaal en veranderingen in inkomen en welzijn in zeventien subnationale initiatieven in de tropen. Het doel van bos-gerelateerde maatregelen ter beperking van de klimaatverandering, waaronder REDD+, is om bossen te beschermen en hun areaal te vergroten. Echter, er zijn wettelijke, morele en praktische redenen om ervoor te pleiten dat deze doelstelling wordt bereikt zonder de lokale bevolking te benadelen, en er idealiter voor te zorgen dat het hen ook voordelen biedt. Aan de hand van enquêtes op dorps- en huishoudenniveau in bijna 130 dorpen en 4000 huishoudens werd informatie verzameld over interventies, gezinsinkomen en subjectief welzijn. De Global Forest Change-dataset (2000-2018) werd gebruikt voor het bepalen van de kroonbedekking en veranderingen daarin. Dorpen werden gegroepeerd in clusters op basis van vergelijkbare niveaus van behandelingsintensiteit en ontbossingstrends. Het cluster met hoge behandelingsintensiteit en vermindering van ontbossingsgraad bestond voornamelijk uit Braziliaanse dorpen. De dorpen in deze cluster hadden in de pre-interventieperiode hogere inkomensniveaus en hogere ontbossingspercentages. In de post-interventieperiode werden deze dorpen over het algemeen getypeerd door een toename van het inkomen en de huishoudens rapporteerden een iets beter niveau van subjectief welzijn. Er werden geen duidelijke verschillen in uitkomsten tussen verschillende typen interventies gevonden, noch waren er aanwijzingen voor uitgesproken compromissen tussen bosbehoud en welzijnsresultaten in alle dorpen en huishoudens in hun geheel. Het kan echter niet worden uitgesloten dat deze compromissen voorkomen in specifieke dorpen en huishoudens. Dit is een nieuwe manier om naar de invloed van verschillende bos-gerelateerde interventies op ontbossing en welzijn en de daarmee gepaard gaande compromissen te kijken. Dergelijke informatie kan waardevol zijn voor beleidsmakers en REDD+ initiatiefnemers die REDD+ strategieën ontwikkelen die zowel voordelen op het gebied van bosbehoud als welzijn kunnen bieden.

De resultaten uit dit proefschrift dragen bij aan een beter begrip van de complexiteit van REDD+ evaluaties op subnationaal niveau. In hoofdstuk 2, 3 en 4 worden interne, methodologische factoren die de evaluatie beïnvloeden behandeld, terwijl hoofdstuk 4 en 5 de resultaten van zulke evaluaties behandelen. Hiermee draagt dit onderzoek zowel bij aan de ontwikkeling van deze evaluaties zelf, als aan het beter begrijpen van de effectiviteit van de interventies waar deze evaluaties op gericht zijn.

Contents

Acronyms & Abbreviations

3E+	Effectiveness, efficiency, equity and co-benefits
AFOLU	Agriculture, forestry and other land uses
AOI	Area of interest
BA	Before-After
BACI	Before-After-Control-Intervention
BAU	Business-as-usual
BFAST	Breaks For Additive Season and Trend
CI	Confidence interval
CIFOR	Center for International Forestry Research
DD	Deforestation and forest degradation
DID	Differences-in-differences
ERC	Ecosystem Restoration Concession
ESA	European Space Agency
FAO	Food and Agriculture Organization of the United Nations
FFI	Fauna and Flora International
FPIC	Free, prior, informed consent
FREL	Forest reference emission level
FRL	Forest reference level
GCS-REDD+	Global Comparative Study on REDD+
GFC	Global Forest Change
GHG	Greenhouse gas
GPG	Good Practice Guidance
HD	Hutan Desa, or Village Forest
ICDP	Integrated Conservation and Development Programme
INPE	Brazilian Institute for Space Research

IPCC	Intergovernmental Panel on Climate Change
IT	Intervention type
KCCP	Ketapang Community Carbon Pools
LiDAR	Light Detection And Ranging of Laser Imaging Detection And Ranging
LULCC	Land use and land cover change
MDA	Mean decrease in accuracy
MDD	Madre de Dios
MMU	Minimum mapping unit
MRV	Measurement, Reporting and Verification
NDCs	Nationally Determined Contributions
NDMI	Normalized Difference Moisture Index
NDVI	Normalized Difference Vegetation Index
NGO	Non-governmental organization
NIR	Near-infrared
OA	Overall accuracy
OECD	Organisation for Economic Co-operation and Development
PA	Producer's accuracy
PES	Payment for environmental services
PT.RMU	PT Rimba Makmur Utama
RED	Reducing emissions from deforestation
REDD+	Reducing emissions from deforestation and forest degradation and enhancing forest carbon stocks
RF	Random Forest model
RO	Research objective
RQ	Research question
SEPAL	System for Earth Observation Data Access, Processing and Analysis for Land Monitoring
SNV	Netherlands Development Organisation
SWIR	Short-wave infrared
TC	Tree cover
TI	Treatment intensity
UA	User's accuracy
UN	United Nations
UNFCCC	United Nations Framework Convention on Climate Change
USD	United States Dollar
VCS	Verified Carbon Standard

1 Introduction

1.1 Forests and climate change

Forest, and tropical forests in particular, provide a range of public and private goods and services. Products include timber and other building material, food, fuel and medicine. In terms of services, forests are an important element in the water cycle, prevent soil erosion and provide nutrients to the soil, regulate the microclimate, serve as carbon sinks, have existence values such as cultural and aesthetic values, and host 80% of the earth's biodiversity.

These goods and services can be provided concomitantly by forests. However, overexploitation can lead to forest degradation. This implies a lesser quality or amount of forest goods and services provision, or complete depletion when forests are deforested and converted to other land uses such as agriculture. Increased world population and corresponding increased demand for food supply has put a pressure on forest land and is expected to continue to do so in the coming decades (Blanco et al., 2014).

Forests can be considered a carbon sink when biomass accumulates in growing trees, while forests can also be a carbon source when forests are degraded or converted into other land uses. Since the 1970s, emissions from agriculture, forestry and other land uses (AFOLU) increased by 20% (Blanco et al., 2014), leading to a share of 23% of net anthropogenic global greenhouse gas (GHG) emissions in the period 2007-2016 (IPCC, 2019).

The climate change mitigation potential of halting or reversing deforestation and forest degradation is estimated to be 24-30% of the total mitigation potential (IPCC, 2014; Goodman and Herold, 2014), which covers not only the avoided emissions but also the carbon sequestration function of standing forests. This mitigation potential could be even higher when avoided emissions from forest change to agriculture are not displaced to other lands and ecosystems (i.e. leakage). In their most recent report, the Intergovernmental Panel on Climate Change (IPCC), stressed that in order to stay below the 2°C increase in global temperature, land-based mitigation is required, with most of the modelled pathways including different combinations of reforestation, afforestation, reduced deforestation and bioenergy (IPCC, 2019).

1.2 REDD+

1.2.1 Global agreements and local action

To counter forest change, a framework has been developed by the United Nations Framework Convention on Climate Change (UNFCCC) to reduce emissions from deforestation and forest degradation, and enhance forest carbon stocks in developing countries (REDD+). The role of forests in climate change mitigation was reiterated with the incorporation of REDD+ in the Paris Agreement (UNFCCC, 2015). In their Nationally Determined Contributions (NDCs), many tropical countries refer to REDD+ as mitigation strategy, while many others (including most non-tropical countries) refer to forest and land sector related mitigation strategies without explicit reference to REDD+ (figure 1.1).

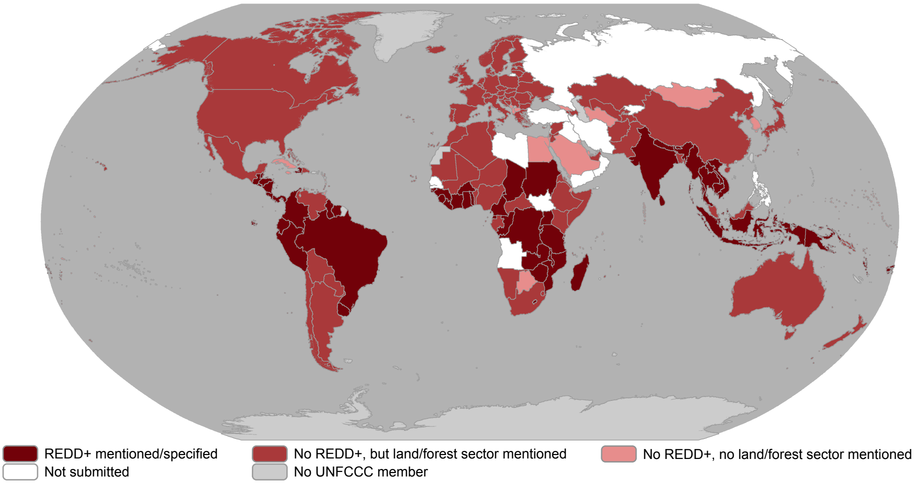


Figure 1.1: Coverage of REDD+ and the land/forest sector in countries' mitigation plans as reported in their NDCs (adapted from Pauw et al., 2016)

REDD+ as a framework is characterized by a phased evolution with a broadening scope. Reducing emissions from deforestation (RED) was launched at COP11 in Montreal in 2005. In 2007, the second "D" was added, representing forest degradation. Later, the extra "+" represented reducing emissions from conservation, sustainable development and enhancement of forest carbon stocks (Wertz-Kanounnikoff and Angelsen, 2009). In the 'Cancun Safeguards', agreements were made on 'free, prior, and informed consent' to protect the rights of indigenous people living in intervention areas, as well as agreements on regular reporting of social and environmental safeguards (UNFCCC, 2011).

REDD+ implementation at (sub)national level has three phases (readiness, policy reforms, and result-based action). Besides a broadening of scope, and thus objectives, the initial idea of sole result-based payments has also changed, since payments during the first and second phase are not linked to the direct measurable outcomes of the policy in terms of reduced emissions (Angelsen and McNeill, 2012). In the readiness phase, many existing Integrated Conservation and Development Programmes (ICDPs) were reframed under the umbrella of REDD+ (Sunderlin and Sills, 2012; Sunderlin et al., 2015; Simonet et al., 2015). These initiatives' objectives were built around a 'triple-win' scenario, in which objectives of conservation, development and climate change mitigation were combined (Milbank et al., 2018).

Since its inception, hundreds of subnational REDD+ initiatives have been implemented across the tropics (Simonet et al., 2015), which enabled both early action and evaluation on lessons learned by implementation on the ground (Sills et al., 2014; Duchelle et al., 2017). These initiatives were implemented by non-governmental organizations, private sector corporations and subnational governments. Originally, the theory of change behind REDD+ was mainly centred around results-based payments using the concept of payment for environmental services (PES). Yet in practice, these initiatives turned out to comprise of a basket of interventions, policies and programmes that combine incentives, disincentives and enabling measures (Angelsen et al., 2018). Since forces driving deforestation are complex, and space and time dependent, the interventions needed to address them effectively need to be tailored accordingly (Seymour and Harris, 2019).

1.2.2 Measurement, reporting and verification

Effective REDD+ requires accurate estimates of emissions from deforestation and forest degradation and changes therein. The international Good Practice Guidance (GPG) of the IPCC guide countries to set up national Measurement, Reporting and Verification (MRV) systems (IPCC, 2003; GOFC-GOLD, 2016). One approach to calculate carbon emissions from forest loss is by multiplying the activity data in a given area by an emission factor (Verchot et al., 2012; IPCC, 2006b). Activity data is the area of land changed from forest into another type of land use. Emission factors are proxies for terrestrial carbon stock densities per unit area. Both activity data and emission factors can be estimated in different ways.

Taking into account countries' differences in capacities (Romijn et al., 2015), and the availability and development of data sources and techniques needed for these estimates, the IPCC identifies different levels of detail and accuracy, so-called Tiers 1 to 3 in terms of emission factors. Here, Tier 1 entails default values per ecological zone, whereas Tier 3 would include spatially explicit wall-to-wall maps with regularly updated carbon stock values. Moving up in Tiers results in increased levels of detail, reduced uncertainty, and increased transparency of carbon stocks reported (Asner et al., 2009), but also entails increased costs and analytical complexity (UNFCCC, 2009). However, costs of activity data estimates have been reduced and quality of these estimates has improved, due to recent innovations in forest monitoring and remote sensing (section 1.3).

In this thesis, most focus is on the activity data side of the equation rather than emission factors, since at subnational level, it is expected that changes in emissions are mostly driven by activity data, rather than local variety in emission factors. This will be discussed in more detail in chapter 6.

1.2.3 Performance assessment

Performance assessment for REDD+ encompasses both carbon emission MRV and MRV on social and environmental safeguards and co-benefits. Performance assessment is important for several reasons (Wertz-Kanounnikoff and McNeill, 2012). First, it is necessary to monitor and measure the effects of projects and policies, in order to see what is (not) working. Consequently, one can improve the design of projects and policies. Second, once results are evaluated, one can use this as a basis for financial rewards. A third, more indirect value of performance measurement regards accountability and promoting effective REDD+ implementation.

While monitoring involves reporting on certain indicators, performance assessment further requires counterfactual thinking. Indicators allow for measurement of progress towards policy goal, but lack the ability to attribute progress to specific interventions (Ferraro, 2009). This attribution is closely linked to the REDD+ principle of *additionality*, that is, in case of reduced emissions to what extent this reduction would not have occurred in absence of the intervention. Countries measure the additionality by comparing actual emissions with a predefined forest reference emission level (FREL), which usually comprises of a business-as-usual (BAU) scenario (Herold et al., 2012). FRELs are therefore an indispensable element in performance assessments. Yet, in the absence of a politically agreed, uniform way of defining FRELs, it is of critical importance that countries, or any agent reporting REDD+ progress at national or subnational levels, are transparent about their FREL definition.

1.3 Remote sensing for REDD+

MRV systems and REDD+ performance assessments require spatially explicit data on the state of forests and forest change, which for an important part build upon satellite-based remote sensing data. Since the launch of the first Landsat satellite in the 1970s, earth observation missions have evolved rapidly (Belward and Skøien, 2014; De Sy et al., 2012). The Landsat archive opened in 2008 which led to the provision of decades of earth observation data for free. This allowed scientists to assess both anthropogenic and natural changes of the earth's surface at longer time series (Woodcock et al., 2008). In addition, the European Space Agency (ESA) provides access to data acquired by the more recently launched Sentinels. In general, more earth observation data are becoming available at lower costs and with higher spatial and temporal resolution. At the same time, free or relatively cheap, cloud computing platforms emerge, including FAO SEPAL, Google Earth Engine and Amazon Web Services, to deal with these increasing flows of data at larger scales (Petersen et al., 2018). This has led to a plethora of tools and datasets focussing on different aspects relevant for REDD+ monitoring, including global forest change (e.g. Hansen et al., 2013), pan-tropical biomass (Baccini et al., 2012; Saatchi et al., 2011; Avitabile et al., 2016), pan-tropical carbon emissions (Harris et al., 2012) and national-scale carbon using LiDAR (Asner et al., 2013b). REDD+ monitoring users may opt for readily available global datasets or open algorithms which may require more technical knowledge but are usually locally adjustable. To aid the user's decision, a better understanding of accuracies of different datasets and tools and their consistency is required.

1.4 Problem statement and research objectives

While especially in recent years monitoring of forests, carbon emissions, and changes therein has progressed at the national level (Romijn et al., 2015), it remains unclear how subnational efforts and outcomes would feed into these national monitoring systems. Further, these monitoring systems require data for which an increased number of datasets are available. Yet, dataset (dis)agreements are understudied. Also more transparency and better cooperation between the science and policy domain is required to measure –and realize– the mitigation potential of REDD+ activities (Grassi et al., 2017). Understanding the implications of different methodologies or choices in data sources for REDD+ outcomes is therefore vital if results-based payments are intended to be based on those estimates.

In recent scientific debates on conservation and development initiatives, there have been increased calls for a so-called “Conservation Evaluation 2.0”, in which conservation and livelihood outcomes of different policy tools applied across sites are assessed and compared jointly through *ex ante* matched treatment and control units (Miteva et al., 2012; Sims and Alix-Garcia, 2017). For comparison across initiatives and to achieve a higher level of learning, performance assessments should incorporate 3E+ outcomes, that is, effectiveness, efficiency, equity and co-benefits including poverty alleviation and biodiversity (Jagger et al., 2009). Despite calls for evidence based REDD+ and theoretically well-developed methodologies, these quasi-experimental differences-in-differences (DID) or BACI assessments are scarcely applied (Sills et al., 2017). Among other complicating factors, the timing of assessment and necessity of research design before interventions are applied, attribution of results to specific interventions, high multicollinearity, reliability of information, spatial misfit between scale of intervention effects and measured environmental outcomes, temporal lags, difficulties to define the appropriate indicators, and potential political sensitivity or privacy issues are making performance assessment difficult (Wertz-Kanounnikoff and McNeill, 2012; Rissman and Smail, 2015; Sills et al., 2017).

Through empirical studies on REDD+ performance assessment, learning-by-doing can be achieved to both operationalize and improve existing (theoretical) BACI methodologies and to inform policy-makers and other intervention implementers on first REDD+ outcomes. The Global Comparative Study on REDD+ (GCS-REDD+) of the Center for International Forestry Research (CIFOR) provides the stage for such an empirical study, in which 23 early-REDD+ initiatives across the tropics are being studied using a BACI study design. The chapters in this PhD thesis build upon the GCS-REDD+ data gathered through village and household level socio-economic surveys and focus group discussions at pre- and post-intervention times.

Although performance measurement is at the heart of the GCS-REDD+ study, developing a method for consistent performance measurement while comprehending the implications of the chosen methodology remains a considerable challenge. Stepwise approaches have been introduced in order to facilitate REDD+ implementation, reference level development and carbon monitoring (Herold et al., 2012; GOFC-GOLD, 2016). It remains unclear which trade-offs in terms of reasonable complexity and required accuracy are present when applying a certain monitoring method as opposed to alternative methods. Moreover, it is unclear how to promote interdisciplinary integration in order to better understand deforestation driver processes, their alignment with REDD+ interventions, and synergies and trade-offs between carbon and non-carbon outcomes. Temporal lags and spatial-diffusion processes complicate monitoring and measurement at the landscape level (Geoghegan, 1998) and thus complicate these alignment, synergy and trade-offs assessments. Jointly examining changes in forest area, carbon emissions, environmental and socioeconomic variables and outcomes is essential for deforestation driver and REDD+ performance assessment, but not straightforward, and experiments using empirical data are needed to unravel those complexities.

This study aims to fill several knowledge gaps related to both methodological issues and understudied REDD+ outcomes. The objective of this thesis is to explore and empirically test the use of environmental and socio-economic data sources to support subnational REDD+ performance assessment. To this end, the following research questions are defined:

1. What are the characteristics and consequences of different forest change assessment approaches?
2. How do the availability and choices of forest change datasets influence REDD+ measurements and corresponding uncertainties?
3. How can an integrated deforestation drivers assessment help understand driver-intervention alignment?
4. What are the relationships between changes in forests and non-carbon outcomes in the context of different REDD+ interventions?

1.5 Thesis outline

Chapters 2 to 5 constitute the core chapters of this thesis. These four chapters are inherently linked to the research questions as presented in section 1.4.

Chapter 2 presents a methodological assessment of different ways to assess forest cover change at the subnational level. It develops (1) a BACI method to assess the effectiveness of these REDD+ initiatives; compares (2) the results at the meso (initiative) and micro (village) scales; and compares (3) BACI with the simpler BA results. Hereby, it provides insights as to how the choice of assessment method influence REDD+ measurements.

Chapter 3 delves deeper into the forest cover change data required for performance assessment of REDD+. This chapter studies (1) the differences in accuracy between datasets of forest cover change; examines (2) if and how combinations of datasets can increase accuracy; and demonstrates (3) the effect of (not) doing accuracy assessments for REDD+ performance measurements. Like chapter 2, it discusses the implications of these results for REDD+ measurements.

Chapter 4 examines deforestation and forest degradation drivers and REDD+ interventions at five sites in Latin America and Southeast Asia. An interdisciplinary analysis of remotely sensed, spatially modelled, and locally reported drivers examines the complementarity of different driver assessment methods. It shows how a better understanding of the forest change drivers at the local scale in the intervention design phase can contribute to a better alignment of REDD+ interventions with forest change drivers.

Chapter 5 consists of an exploratory analysis of changes in forests and well-being in the context of forest conservation interventions in 17 subnational REDD+ initiatives across the tropics. A combination of socioeconomic surveys and remote sensing analyses was used to assess how interventions with different levels of treatment intensity associated with changes in deforestation trends, income, and perceived well-being at the household and village level.

The research conducted in this PhD contributes to the understanding of the complexities involved in REDD+ performance assessments at the subnational level. In chapter 2, 3 and 4, methodological factors are being addressed that are key to performance assessment design and implementation, whilst chapter 4 and 5 address the performance assessment outcomes (figure 1.2). Hereby, this study contributes to the development of performance assessment itself as well as to the understanding of the interventions' effectiveness which these assessments evaluate.

Chapter 6 presents the main conclusions of this thesis and revisits the research questions. This synthesis further discusses the implications of these results for performance assessments of climate change mitigation efforts in the future.

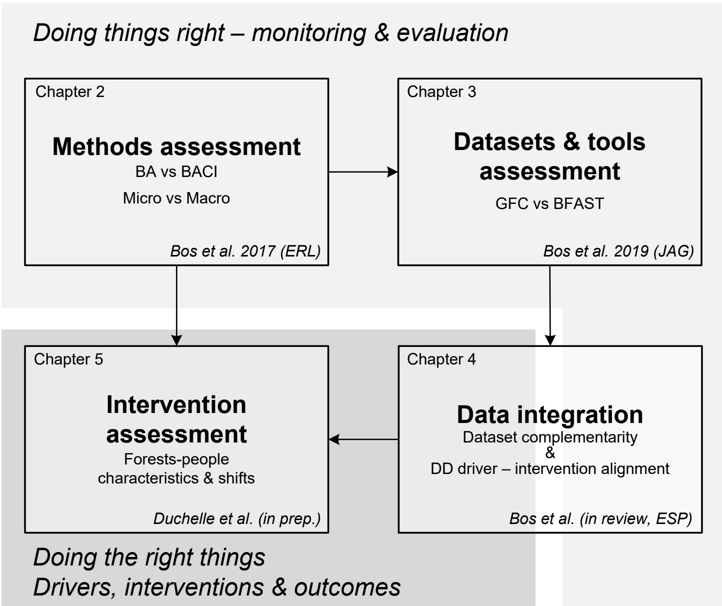


Figure 1.2: Linkages between the core chapters of the thesis

2 Comparing methods for assessing the effectiveness of subnational REDD+ initiatives

This chapter is based on:

Bos, A. B., Duchelle, A. E., Angelsen, A., Avitabile, V., De Sy, V., Herold, M., Joseph, S., de Sassi, C., Sills, E. O., Sunderlin, W. D., and Wunder, S. (2017). Comparing methods for assessing the effectiveness of subnational REDD+ initiatives. *Environmental Research Letters*, 12(7):074007

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 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 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.

2.1 Introduction

REDD+ has emerged as a key climate change mitigation strategy within the UNFCCC. Through the Paris agreement, the necessity for supporting and implementing REDD+ was reconfirmed and the role of forests as carbon sinks emphasized (UNFCCC, 2015). So far, approximately 40[1] countries mention either REDD+ or forests as part of the mitigation strategy in their NDCs. This importance makes it critical to monitor and evaluate the carbon effectiveness of REDD+.

The 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., 2013b).

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 BACI, or 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

[1]UNFCCC NDC registry http://www4.unfccc.int/ndcregistry/Pages/All.aspx, 5 December 2016

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 10 years). Hence, unlike BACI, it does not account for changes in drivers during the intervention period. This chapter 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.

2.2 Material and methods

2.2.1 Study area

Our study includes 23 subnational REDD+ initiatives in Brazil, Peru, Cameroon, Tanzania, Indonesia and Vietnam from CIFOR's GCS-REDD+ (figure 2.1). They differ greatly in terms of proponent type (government, NGO, private sector), size (ranging from 28 to approximately 160,000 km^2), 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).

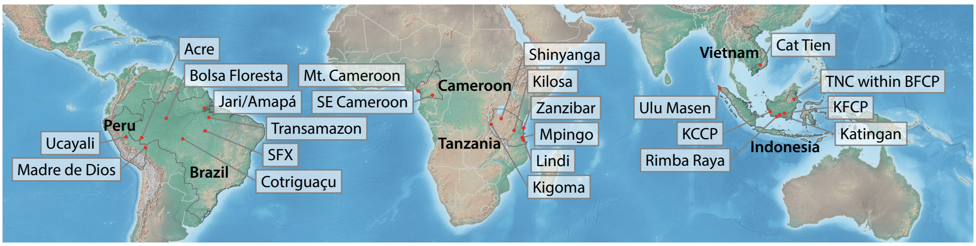


Figure 2.1: Initiatives included in the Global Comparative Study on REDD+

2.2.2 Tree cover data

We use the Global Forest Change (GFC) 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 30m). 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).

2.2.3 Performance assessment framework

For both approaches, we aggregate the time series data on annual tree cover loss into two periods (before and after) (figure 2.2. 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 (see appendix 6 of Sills et al., 2014)[2], thus the number of years in the after period ranges from two to nine (see table 2.1). The BA score α is calculated as follows:

$$\begin{gathered} \textit{BA score } \alpha = \bar{x}_{AI} - \bar{x}_{BI} \\ \textit{with } \bar{x}_{AI} = \frac{1}{n_a}\sum_{i=1}^{n_a} x_i \textit{ and } \bar{x}_{BI} = \frac{1}{n_b}\sum_{i=1}^{n_b} x_i \end{gathered} \tag{2.1}$$

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 (here: 2001) up until the intervention started, n_a and n_b the

[2] 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.

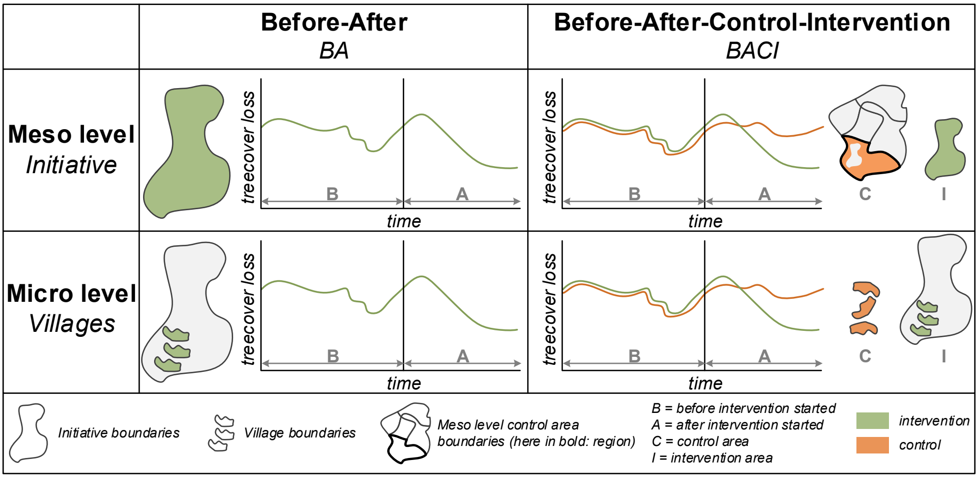


Figure 2.2: Theoretical framework for comparing performance assessment methods (BA and BACI) at the *meso* and *micro* level. Homogeneous trends in the *before* period like those presented here show the ideal situation.

number of years in respectively the after and before period. A BA score α 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:

$$\begin{aligned} \textit{BACI score } \beta &= (\bar{x}_{AI} - \bar{x}_{BI}) - (\bar{x}_{AC} - \bar{x}_{BC}) \\ \textit{with } \bar{x}_{AI} &= \frac{1}{n_a} \sum_{i=1}^{n_a} x_i \text{ , ... } \textit{etc.} \end{aligned} \tag{2.2}$$

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 2.3).

2.2.4 Levels of analysis

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 2.3), 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).[3]

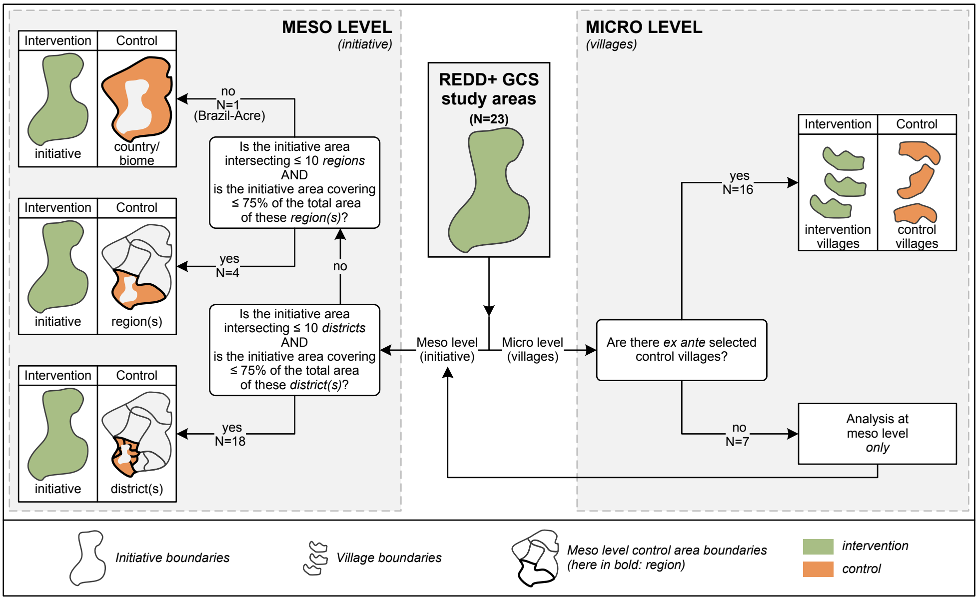


Figure 2.3: Decision tree for selecting control units at meso (left panel) and micro (right panel) levels.

For the micro level analysis (right panel, figure 2.3), we focused on 16 of the 23 REDD+ initiatives, known as "intensive sites" in the GCS-REDD+, 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

[3]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.

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 (see appendix A.1).

2

2.3 Results

2.3.1 General results

Table 2.1 [4] shows the summary statistics of the main variables introduced in section 2.2.3.

Table 2.1: 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 after period	23	2	9	6	6
both	n_b	years in before period	23	5	12	8	8
meso	α	BA score[a]	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[b]	16	-2.139	0.669	-0.271	0.048
micro	β	BACI score	16	-2.277	2.827	-0.449	-0.466

[a] In initiative area.
[b] In intervention villages.

The results of the BA α and BACI β performance scores were grouped into *good*, *neutral* and *poor*[5], 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 2.4).

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 2.1), 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)[6]. 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.

[4]See appendix A.2 for an extended version of the summary statistics.
[5]When grouping the scores, the following thresholds were used: good $\leq$ 0.1; 0.1>neutral<0.1; and poor $\geq$ 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.
[6]These results are not influenced by the difference in sample size between the meso and micro level (appendix A.3).

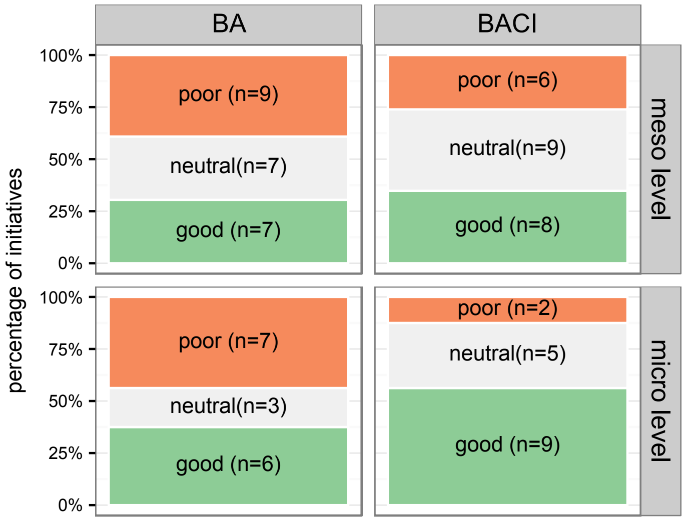


Figure 2.4: BA and BACI classified scores per analysis level, where n is the number of initiatives.

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.

2.3.2 Individual BA and BACI scores

To better understand the methodological differences, in this section we examine specific scenarios. Table 2.2 shows the occurrences of the prevailing factors that affect the BA and BACI scores, which we explain in more detail below.

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 2.2). 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 (appendix A.4). One such case is shown in figure 2.5 where meso-level *before* deforestation rates in the initiative area exceeded those in the corresponding control districts.

Table 2.2: Occurrences per analysis level of factors affecting the BA and BACI scores.

Level	n cases[a]	Bias in before 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.

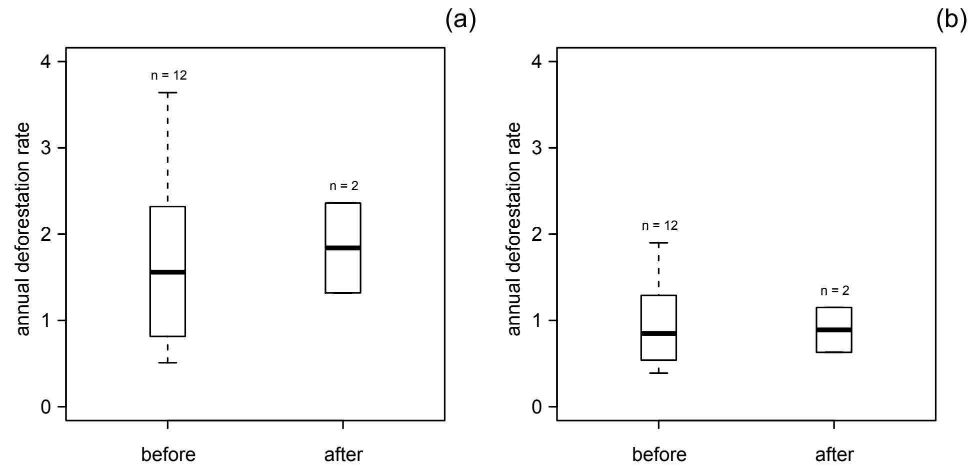


Figure 2.5: Annual deforestation rates (%) in the *before* and *after* period for the intervention (a) and control (b) areas for one initiative in Brazil, where n is the number of years per period. Upper and lower extremes of whiskers represent Q3+1.5*IQR and Q1−1.5*IQR respectively, where IQR=Q3-Q1.

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.

Peak years

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

Table 2.3: Evaluating BA and BACI score robustness to peak year influence

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

Bold 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.

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 2.3, 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.

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 2.6) 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).

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

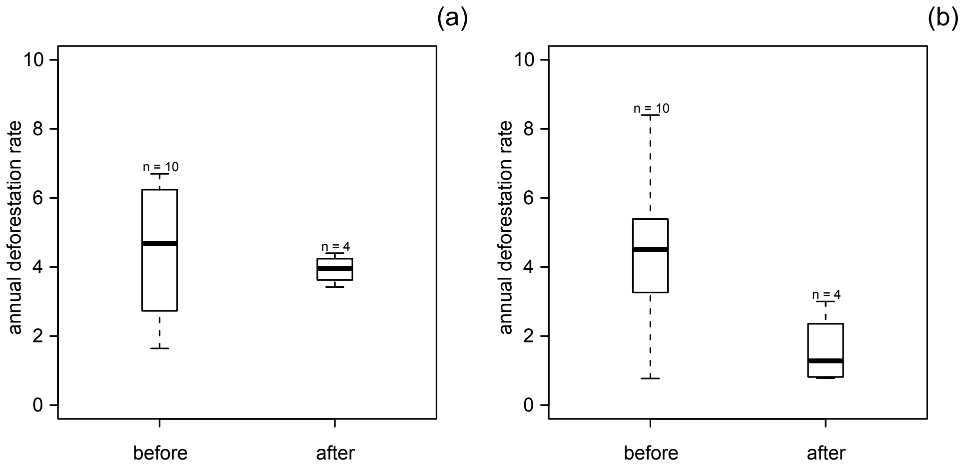


Figure 2.6: Annual deforestation rates (%) in the *before* and *after* period for the intervention (a) and control (b) areas for one initiative in Brazil, where n is the number of years per period. Upper and lower extremes of whiskers represent Q3+1.5*IQR and Q1−1.5*IQR respectively, where IQR=Q3-Q1.

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.

2.4 Discussion

We applied BA and BACI approaches at meso and micro levels to assess subnational-level REDD+ performance. Both approaches and levels of measurement have advantages and disadvantages for effectiveness assessment (table 2.4). 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.

Table 2.4: 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 - susceptible to external factors of influence, i.e. changes in deforestation could wrongfully be attributed to the intervention	+ able to discern additionality attributable to the intervention - requires careful *ex ante* control site selection and matching - high sensitivity of results to matching method
Aggregation level	
Meso level	**Micro level**
+ helps understanding trends within context + may indicate cases of leakage (but further analysis is then still required) - defining control areas may be more difficult	+ allows more precise comparison between intervention-targeted and non-targeted units - the notion of village is not universal, and delineating boundaries may be subjective - small changes may obscure “bigger picture” - sensitive to extreme events or single drivers

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-REDD+, 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 sites[7] 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 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 (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; GOFC-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

[7]Two sites at micro level, and two sites at both meso and micro level

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.

2.5 Conclusions

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.

2.6 Acknowledgements

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.

3 Global data and tools for local forest cover loss and REDD+ performance assessment: accuracy, uncertainty, complementarity and impact

This chapter is based on:

Bos, A. B., De Sy, V., Duchelle, A. E., Herold, M., Martius, C., and Tsendbazar, N.-E. (2019). Global data and tools for local forest cover loss and REDD+ performance assessment: Accuracy, uncertainty, complementarity and impact. *International Journal of Applied Earth Observation and Geoinformation*, 80:295–311.

Abstract

Assessing the performance of efforts to reduce emissions from deforestation and forest degradation (REDD+) requires data on forest cover change. Innovations in remote sensing and forest monitoring provide ever-increasing levels of coverage, spatial and temporal detail, and accuracy. More global products and advanced open-source algorithms are becoming available. Still, these datasets and tools are not always consistent or complementary, and their suitability for local REDD+ performance assessments remains unclear. These assessments should, ideally, be free of any confounding factors, but performance estimates are affected by data uncertainties in unknown ways. Here, we analyse (1) differences in accuracy between datasets of forest cover change; (2) if and how combinations of datasets can increase accuracy; and we demonstrate (3) the effect of (not) doing accuracy assessments for REDD+ performance measurements.

Our study covers five local REDD+ initiatives in four countries across the tropics. We compared accuracies of a readily available global forest cover change dataset and a locally modifiable open-source break detection algorithm. We applied human interpretation validation tools using Landsat Time Series data and high-resolution optical imagery. Next, we assessed whether and how combining different datasets can increase accuracies using several combination strategies. Finally, we demonstrated the consequences of using the input datasets for REDD+ performance assessments with and without considering their accuracies and uncertainties.

Estimating the amount of deforestation using validation samples could substantially reduce uncertainty in REDD+ performance assessments. We found that the accuracies of the various data sources differ at site level, although on average neither one of the input products consistently excelled in accuracy. Using a combination of both products as stratification for area estimation and validated with a sample of high-resolution data seems promising. In these combined products, the expected trade-offs in accuracies across change classes (*before*, *after*, *no change*) and across accuracy types (user's and producer's accuracy) were negligible, so their use is advantageous over single-source datasets. More locally calibrated wall-to-wall products should be developed to make them more useful and applicable for REDD+ purposes. The direction and degree of REDD+ performance remained statistically uncertain, as confidence intervals were overlapping in most cases for the deforestation estimates before and after the start of the REDD+ interventions. Given these uncertainties and inaccuracies and to increase the credibility of REDD+ it is advised to (1) be conservative in REDD+ accounting, and (2) not to rely on results from single currently available global data sources or tools without sample-based validation if results-based payments are intended to be made on this basis.

3

3.1 Introduction

Under the UNFCCC, REDD+ has been initiated as an important climate change mitigation strategy. Hundreds of government and non-government led REDD+ programs and projects have emerged at the subnational and local level over the past decade (Simonet et al., 2015). In order to track the performance of these initiatives, implementers must create or leverage MRV schemes for carbon stocks and carbon emissions. One approach to calculate carbon emissions is by multiplying the activity data in a given area by an emission factor (Verchot et al., 2012; IPCC, 2006a). Activity data is the area of land changed from forest into another type of land use.

The estimation of activity data evolved rapidly through innovations in remote sensing and forest monitoring, with algorithms and datasets with ever increasing levels of coverage, spatial and temporal detail, and accuracy. However, these datasets do not necessarily agree with each other, and more transparency and better cooperation between the science and policy domain is required to measure –and realize– the mitigation potential of REDD+ activities (Grassi et al., 2017). Estimates can differ due to many factors, including misalignment of reference levels and time periods, forest and deforestation definitions used, and (remote sensing) data sources used for a map product (e.g. different satellite data) (Melo et al., 2018). Although the resulting differences in estimates are expectable and understandable, the ambiguity leaves room for political manoeuvring around the data (Wong et al., 2016) which threatens accountability. On the positive side, it is becoming more common practice to systematically report map product's accuracies and uncertainties (e.g. Olofsson et al., 2013, 2014; Stehman, 2014), increasing both transparency and product comparability. To this end, a reference classification is needed. Accuracy is defined as the degree to which the produced map agrees with this reference classification (Olofsson et al., 2013), which generally requires a sample-based validation. The uncertainty of the corresponding area estimates of, in this case, deforestation, is then expressed by the variance, standard error, or confidence intervals (CIs) of these estimates. One could account for these uncertainties in the input data by being conservative about the subsequent REDD+ estimates, so as to prevent overestimation of the reduced emissions (Grassi et al., 2008).

Locally calibrated products are often favoured over global products, as this can considerably reduce the sample size for validation purposes (GFOI, 2016). Still, some widely used regional forest change datasets are found to be inaccurate by underestimating forest loss (Milodowski et al., 2017). Also, trade-offs exist between accuracy, local adjustability, and sample size needed on the one hand, and ease of use, processing time, knowledge and skills required on the other (Duchelle et al., 2015). While at the national level, in recent years the capacities of countries are increasing (Romijn et al., 2015), for local and subnational REDD+ initiatives it is often difficult and impractical to gain sufficient capacities and resources to perform proper area estimations. Here, the availability of open-source products provides an attractive opportunity. It remains understudied however, to what extent these readily available datasets and tools can contribute to challenges in the environmental domain and to REDD+ performance assessments in particular.

For local forest cover loss measurements, it is of vital importance to understand the differences in accuracies of forest cover loss maps derived from different products and tools. This supports the choice to use either more complex, time-consuming, but locally adaptable tools that provide the required high accuracies, or to opt for a readily available product with global coverage which might suffice in certain cases. In addition, accuracy assessment of combinations of products and tools can reveal their complementarities and show how uncertainties can be minimized while maximizing accuracies. In other words, in terms of increased accuracy and decreased uncertainty, a combined product may be better than the sum of its parts. An earlier study has focused on a comparison of available datasets in terms of in accuracy and uncertainty in one country (Melo et al., 2018), while others have studied the differences across several tropical countries (e.g. Turubanova et al., 2018). To the best of our knowledge, this is the first effort however, to compare different products at different (subnational) sites across the tropics, while exploring the potential and added value of combining those products.

Datasets used for REDD+ performance assessments should, ideally, be free of any confounding factors, but it is currently unclear how performance estimates are affected by data uncertainties. Hence, a systematic accuracy assessment is necessary to compare accuracies in various map products and to gain insight in the remaining uncertainty in deforestation area estimates. Furthermore, it remains understudied whether and how map products could complement each other and to what extent they are suitable for measuring the performance of REDD+. Therefore, the objectives of this study are to analyse if and how combinations of datasets can increase accuracy, and to understand how differences in accuracy between forest cover change datasets and its corresponding uncertainty influence REDD+ performance assessments. We defined the following research questions:

1. How do forest cover loss datasets differ in terms of accuracy?
2. What is the complementarity of these forest cover loss datasets in increasing accuracy?
3. How do map accuracy and area estimate uncertainty influence REDD+ performance assessment?

3.2 Material and methods

3.2.1 Study area

We use data from five local REDD+ initiatives located in four countries across the tropics (table 3.1). These initiatives are part of the Global Comparative Study on REDD+ (CIFOR, 2017) and were selected to represent a wide range of intervention types ((dis)incentives and enabling measures), implementer types (government, non-governmental organization, private sector), and geographies across the tropics. Furthermore, they vary in terms of size and environmental context, namely from dense primary rainforest to dry miombo woodlands (Sills et al., 2014). Data availability constraints affected the selection procedure, as the availability of both map products (section 3.2.3) was a prerequisite for this study.

Table 3.1: Site characteristics

Site	(Approx.) size (ha) of AOI[2]	Main ecozone(s)[3]	REDD+ start year	National forest definition[1] Tree cover (%)	National forest definition[1] MMU[4] (ha)
Peru	1,100,000	Tropical rainforest	2009	30	0.09
Tanzania	200,000	Tropical dry forest/ tropical shrubland	2010	10	0.50
Vietnam	800,000	Tropical rainforest/ tropical moist deciduous forest	2009	10	0.50
Indonesia-A	2,000,000	Tropical rainforest	2008	30	0.25
Indonesia-B	3,600,000	Tropical rainforest	2009	30	0.25

[1] Based on most recent submissions to UNFCCC (2019).
[2] Area of interest
[3] Source: FAO
[4] Minimum mapping unit.

3.2.2 Summary of workflow

The workflow and processing steps (figure 3.1) were repeated for each study site. We compared the accuracy of a tree cover change dataset, i.e. GFC, and a map developed using an open-source algorithm to detect forest cover change, i.e. Breaks For Additive Season and Trend (BFAST). For each study site and based on national forest definitions, we used the same forest mask using tree cover (TC) percentage and an area sieve using the minimum mapping unit (MMU) (table 3.1). We thus compared differences in change detection between the two input products, rather than differences in forest definitions applied. We considered three classes: *before*, *after* and *no change*. The transition between *before* and *after* is defined by the start year of each studied initiative. We combined the two products using different reclassification strategies, which led to a set of new combined change map products. We applied a stratified random sample on the change map and validated the original products and reclassified products using a set of visual tools. Accuracies were calculated using these validation samples, as well as the differences in accuracies relative to the two input map products. The uncertainty in the area estimates was expressed using the 95% and 50% CIs of those estimates. We compared the map estimates and reference-based area estimates. Finally, we assessed the influence of uncertainty in the area estimates and their trends on REDD+ performance measurements. All analytical steps are discussed in more detail below.

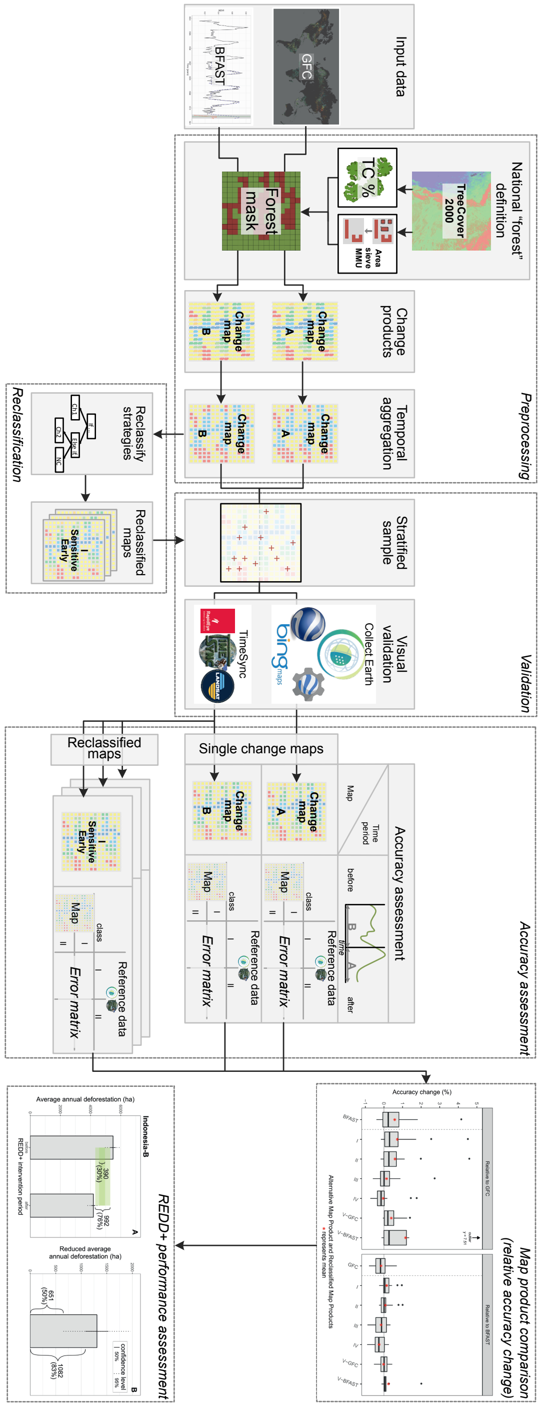


Figure 3.1: Workflow and processing steps

3.2.3 Input data

For the first map product, we used Global Forest Change (GFC) data (version 1.3), a Landsat-based time-series dataset of tree cover density in 2000 and annual tree cover loss for 2001-2015 (Hansen et al., 2013). The GFC product provides yearly forest cover loss data with global coverage. Together with baseline data on forest cover in 2000, users can relatively easily examine deforestation patterns using their own forest definitions. Data analysis using the Global Forest Watch tools does not require expert GIS knowledge. The other product is based on the BFAST algorithm (Verbesselt et al., 2010, 2012; DeVries et al., 2015), which requires a time series of local input data (here, Normalized Difference Vegetation Index (NDVI) and Normalized Difference Moisture Index (NDMI) based on Landsat satellite data). With this adaptable open-source deforestation detection algorithm, users can analyse deforestation patterns in their own time series data in, for example, a cloud processing environment. It is usually applied to smaller areas, as processing time increases with longer time series and larger area spans. Some degree of remote sensing knowledge and coding skills are necessary to apply the algorithm on the time series. The algorithm is highly flexible and can be adapted to the local (environmental) context and user needs. The user can calibrate the model by adjusting the parameters to the local context, resulting in change rasters with interannual precision. Both products allow the user to create forest cover change products with a temporal resolution of one year or shorter, and a spatial resolution of 30m. The main differences between the two products regard their flexibility, coverage, and ease-of-use (table 3.2).

Table 3.2: Comparison of GFC and BFAST products (with information from Hansen et al. (2013); Verbesselt et al. (2012); Gross et al. (2017)).

	GFC	**BFAST**
Type	2000 tree cover; loss; gain; and loss year raster products	Change detection algorithm
Sensor	Landsat ETM+	Depends on user input, here: Landsat ETM+
Spatial resolution	30m	Depends on user input, here: 30m
Temporal resolution	Year	Julian day, limited by user input and cloud coverage
Spatial coverage	Global	Site based; 'case studies'
Algorithm	Bagged decision tree model	Additive season and trend model
Advantages	Global coverage; easy to use; end product freely available	Locally modifiable; open source
Disadvantages	Algorithm not flexible; not near-real time	Requires user's input data; requires expert knowledge; computationally intensive
Source	http://earthenginepartners.appspot.com/ science-2013-global-forest	http://bfast.r-forge.r-project.org/
Reference	Hansen et al. (2013)	Verbesselt et al. (2012)

3.2.4 Pre-processing

We aligned our forest definitions with the corresponding countries' definitions[1]. These generally consist of a tree cover or crown percentage at the baseline year and a MMU (table 3.1). GFC's tree cover density layer for the year 2000 (TC2000) allowed us to create forest masks based on the nationally defined tree cover percentage thresholds. Next, we applied area sieves following the countries' defined MMU and applied these forest masks to both input products. We defined deforestation as a change from forested land (using the forest mask) to land that has been clear cut (i.e. bare soil)[2]. In addition to aligning forest and deforestation definitions, we needed to temporally align the data for the products to represent the same time periods (table 3.3). We then aggregated the change products into three classes, representing (1) the period *before* the REDD+ interventions started, (2) the period *after* the interventions started, and (3) *no change* (i.e. stable forest). All other pixels, (i.e. non-forest; forest cover change in other years etc.) were excluded from further analyses.

Table 3.3: Temporal alignment of change products per study site.

Site	Time frame GFC	Time frame BFAST	REDD+ start year	Aligned before period	Aligned after period
Peru	2001-2015	1999-2014	2009	2001-2008	2009-2014
Tanzania	2001-2015	2005-2015	2010	2005-2009	2010-2015
Vietnam	2001-2015	2005-2014	2009	2005-2008	2009-2014
Indonesia-A	2001-2015	2001-2014	2008	2001-2007	2008-2014
Indonesia-B	2001-2015	2001-2015	2009	2001-2008	2009-2015

3.2.5 Reclassification of change products

Since these datasets generally have their own strengths and weaknesses (table 3.2), we assessed whether joint products can lead to an accuracy increase. Therefore, we combined the two products at pixel level using five different reclassification strategies. The first four strategies are defined by differences in sensitivity to change and in timing of change detection (figure 3.2), based on the following decision rules:

I *Sensitive early* – Adopt value of change product that detects a disturbance the earliest, regardless of the other change product's detection;

II *Sensitive late* – Adopt value of change product that detects a disturbance the latest, regardless of the other change product's detection;

[1]Following the submissions to the UNFCCC's REDD+ platform (UNFCCC, 2019)

[2]Sometimes land use change from (natural) forest to forest plantation is considered degradation or even enhancement of carbon stock (e.g. in Vietnam's REDD+ Forest reference level (FRL) submission to the UNFCCC, 2016), but here it is considered deforestation, since at –at least- one point in time the forest was cleared which leads to a reflectance of bare soil.

III *Conservative early* – If any of the change products classifies the pixel as no change, then the decision for the reclassified product is no change. If both products detect change, trust the earliest detection;

IV *Conservative late* – If any of the change products classifies the pixel as no change, then the decision for the reclassified product is no change. If both products detect change, trust the latest detection.

A fifth strategy was added to represent a case in which the timing of change detection is irrelevant. Here, the two individual products were aggregated into two binary change-no change rasters, disregarding the year or corresponding period of change detection. Details are visualized in appendix B.1.

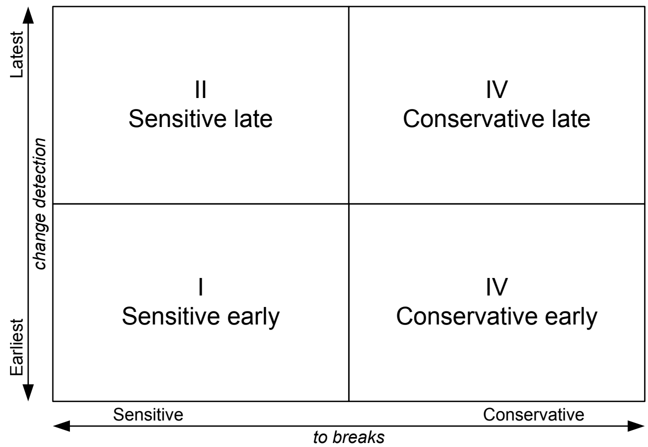


Figure 3.2: Rationale behind reclassification strategies

Table 3.4 shows the reclassification strata for each strategy, which formed the input for the stratified sampling (see next section). For each site, the five reclassification strategies resulted in six extra change maps, that is, four combined and two 'timeless' raster datasets, which were added to the accuracy assessment for comparison with the original GFC and BFAST products.

3.2.6 Validation

Sample size is important when designing validation schemes for comparative purposes (Foody, 2009). Although our individual aggregated change raster datasets consisted of three classes (*change before*, *change after* and *no change*), for simplification in the sampling design we considered them as having a binomial distribution (either change or no change) and used an alpha of 0.10, planned proportion estimate of 0.5 (i.e. conservative), and 0.05 margin of error leading to a sample size of 270 pixels per site (Foody, 2009; Cochran, 1977). We overlaid the

Table 3.4: Strata and classification values of different reclassification strategies

			Combination strategies				Timeless strategies	
			I	II	III	IV	V	
GFC	BFAST	Validation stratum	sensitive - early	sensitive - late	conservative - early	conservative - late	timeless-GFC	timeless-BFAST
before	before	1	before	before	before	before	change	change
after	before	3	before	after	before	after	change	change
no change	before	4	before	before	no change	no change	no change	change
before	after	3	before	after	before	after	change	change
after	after	2	after	after	after	after	change	change
no change	after	5	after	after	no change	no change	no change	change
before	no change	4	before	before	no change	no change	change	no change
after	no change	5	after	after	no change	no change	change	no change
no change	no change	6	no change	no change	no change	no change	no change	no change

two input change products with each three classes, resulting in nine possible combination values. These nine classes were aggregated into six strata (table 3.4). At each site, the 270 pixels were randomly selected across the strata, which led to 45 sample pixels per stratum (figure 3.3).

A validation survey was developed using Open Foris Collect (Open Foris, 2019). The survey and samples were loaded into Google Earth via CollectEarth and simultaneously visualised in R using the TimeSync package (Cohen et al., 2010). Each sample was visually checked through multiple available historical images within Google Earth (if any), the most recent Bing Maps image, the most recent image via Google Earth Engine, and false colour yearly composites of Landsat data within Google Earth Engine. Within R, a time series of RGB and false colour (Near-infrared (NIR), short-wave infrared (SWIR)1, red) snapshots were created with TimeSync. Together this allowed us to determine (1) whether there was any disturbance and, (2) if so, to find the timeliest disturbance date. In case of multiple disturbances within the time series, the first disturbance was recorded.

3.2.7 Accuracy assessment

After completing the validation survey, the visual judgements from the validation survey were compared with the findings from the GFC, BFAST and reclassified products. A map pixel was considered correct if both the status (*change* or *no change*) and time period (*before* or *after*) matched the visual judgement. Accuracies of the map products and the class area proportions were estimated while taking into account the inclusion probability of the samples per site. Since the sampling stratification was a combination of GFC and BFAST results, we followed the approaches detailed in Stehman (2014) which addresses estimating map accuracies and class areas when the sampling strata are different from the map classes. CIs of the estimation also followed the same method (Stehman, 2014; Cochran, 1977). For the remainder of this article, with 'map-based area estimates' we refer to area estimations directly calculated from the maps, whereas 'reference-based area estimates' refers to the areas as derived from the class

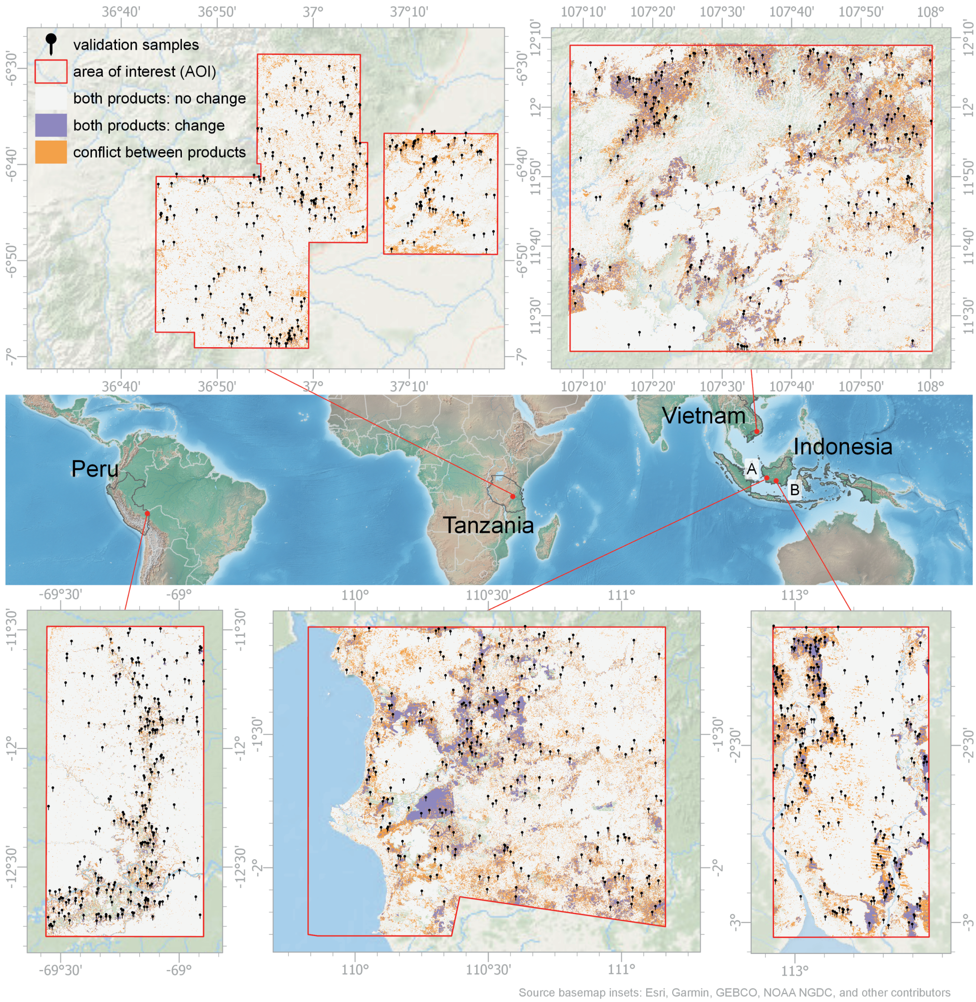


Figure 3.3: Study sites with validation samples and areas of agreement and conflict between the two input map products.

3

area proportions coming from the sample-based validation using reference data. Next, the differences in overall, producer's (inversely linked to errors of omission) and user's (inversely linked to errors of commission) accuracies were assessed by calculating the relative accuracy changes, which give insight in which reclassification strategy provides the largest increase in accuracy compared to the original input products. Relative accuracy change was calculated as follows:

$$RA(x) = \frac{(A_x - A_y)}{A_y} \tag{3.1}$$

Where x is the alternative map product, y is the original map product (either GFC or BFAST), and A is the corresponding accuracy (overall, producer's or user's accuracy).

3.2.8 Performance assessment

In this study, we simplify REDD+ performance by referring to the direction in deforestation trend over time, hence good REDD+ performance corresponds to reduced average annual deforestation. We compared the trends in average annual deforestation from before and after the start of the REDD+ intervention (Bos et al., 2017). The impact of ignoring data accuracy in REDD+ performance assessments was assessed by comparing the average annual deforestation per period for the map estimates and reference-based area estimates.

Trends and uncertainties were assessed in two ways. First, they were visually assessed by focusing on the overlap of the CIs of the deforestation estimates in the *before* and *after* period. Presence of such overlap would mean that direction and magnitude of REDD+ performance remains uncertain. Absence of such overlap would reveal the direction of deforestation trend and its magnitude with more certainty. In addition to the commonly used 95% CI, we applied a 50% CI. This means one accepts a 25% probability of overestimating the 'true' REDD+ value in the monitoring period, which is similar to the adjustment procedure under Article 5.2 of the Kyoto Protocol (UNFCCC, 2006; as cited in Grassi et al., 2008). Second, the trend uncertainty was calculated using the (joint) variances and CI of the trend itself (GOFC-GOLD, 2016).

The conservativeness principle (Grassi et al., 2008) was applied to a case with a decreasing trend, to examine the influence of different conservativeness standards on the final REDD+ estimate.

3.3 Results

3.3.1 Annual deforestation rates

Figure 3.4 shows an overview of the annual deforestation rates for both GFC and BFAST input products at each site before the accuracy was assessed and thus before the area estimates of deforestation using the reference data were calculated. Both products show overall higher annual deforestation rates in the Southeast Asian sites compared to the sites in Peru and Tanzania. The deforestation trends appear similar when comparing the two products at all sites. However, deforestation estimates in individual years differed considerably, especially so in Vietnam (2005) and Indonesia (2006 and 2007), which might indicate differences in timeliness of deforestation detection. In terms of REDD+ performance, these results reveal some ambiguity of the deforestation trends. In Peru, the GFC showed slightly increasing deforestation while according to BFAST deforestation was generally going down since the start of the REDD+ initiative. The site in Tanzania showed no clear performance while the steep drop in deforestation in site Indonesia-B after 2007 might indicate positive REDD+ performance.

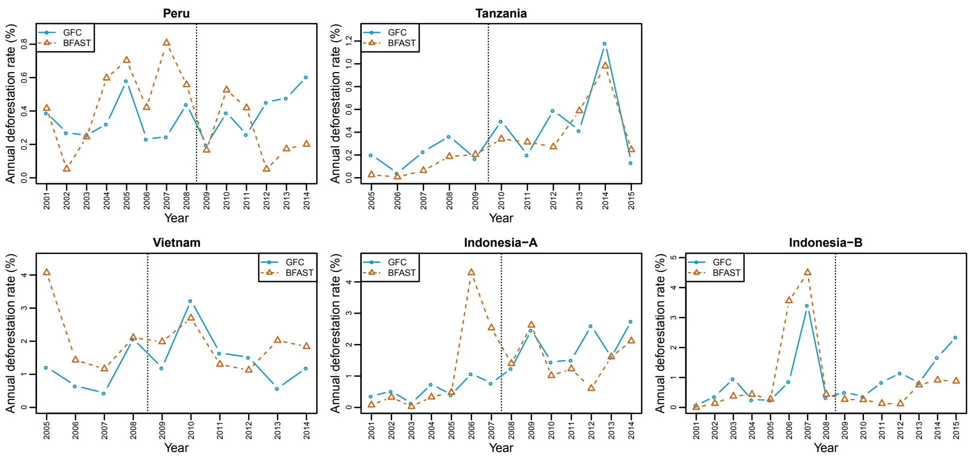


Figure 3.4: Site-based comparison of annual deforestation rates. Rates represent the deforestation detected by the input products as percentage of forest cover in 2000. Note that the x and y-scales differ per site. The vertical dotted line represents the start year of the REDD+ intervention(s) in the corresponding site and thus the transition from the *before* to *after* period.

3.3.2 Accuracy

Overall, user's and producer's accuracy

For all original and reclassified map products, the error matrices were calculated based on the comparison between the map class (*change before*, *change after*, *no change*) and the visually assigned class using the reference data. Figure 3.5 shows the overall accuracies (OAs), user's accuracies (UAs) and producer's accuracies (PAs) stemming from these error matrices.

While the OAs for all products were high, this result primarily stems from correctly detected no change areas (stable forest cover) which spans the majority of the areas studied. The UAs and PAs, and their corresponding errors of commission and omission respectively, were more informative for change related classes. In general, the lower PAs indicate that all products underestimate deforestation. When comparing the two input products, BFAST shows on average slightly higher accuracies (OA, UA and PA) compared to the GFC product. We found a general tendency of lower variation in accuracies of BFAST as compared to GFC across the sites.

However, whether GFC or BFAST performed better in terms of accuracy differs per study site (appendix B.2). In the Peruvian site, the GFC and BFAST accuracies were quite similar, although there were some notable differences in PA. The Tanzanian site was characterized by low accuracies in general, but BFAST seemed to perform better at distinguishing real deforestation impacts from seasonal effects, hence the difference in UA between the two products. In both Indonesian sites the PA of BFAST in the after class was lower compared to GFC, indicating that the most recent changes are not well detected by BFAST. In the Vietnamese site, this was the opposite, as the PA of BFAST outperformed GFC in the after class.

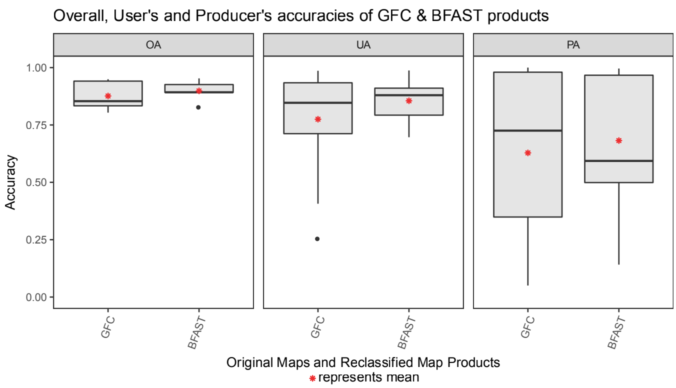


Figure 3.5: Overall (OA), user's (UA) and producer's (PA) accuracies of GFC and BFAST products. All classes (i.e. *change before*, *change after*, and *no change*) are included. Upper and lower extremes of whiskers represent $Q3 + 1.5*$interquartile range (IQR) and $Q1 - 1.5 * IQR$ respectively, where $IQR = Q3 - Q1$.

Relative accuracy change

Comparing accuracies of the original map products and the reclassified map products based on the five strategies (section 3.2.5), in four out of five sites[3] combining input maps following a *sensitive-early* strategy led to significantly higher accuracies compared to the original map products alone (figure 3.6, appendices B.3 and B.4). Still, in the Tanzanian site, none of the reclassification strategies led to higher accuracies[4] compared to (one of the) individual datasets, due to the poor performance of the GFC product in this study area.

[3]With the Tanzanian site being the exception.
[4]Increases in OA and PA in the change classes, with non to only slight (insignificant) decreases in UA, significance level 0.95 (appendix B.3).

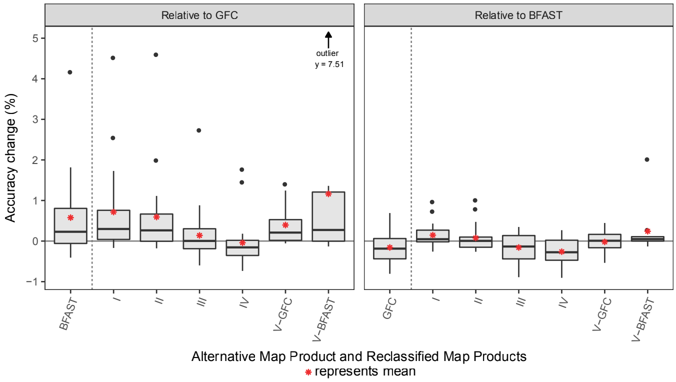


Figure 3.6: Relative change of the accuracies per alternative product. The figure includes the accuracies (only PA and UA) of the *before* and *after* change classes of all sites. Upper and lower extremes of whiskers represent $Q3 + 1.5*$interquartile range (IQR) and $Q1 - 1.5 * IQR$ respectively, where $IQR = Q3 - Q1$.

3

3.3.3 REDD+ performance assessment

Revealing the deforestation trend

To assess the influence of the map products' accuracies and area estimate uncertainties on REDD+ performance assessments, we visualize the average annual deforestation (in ha) in the *before* and *after* class (figure 3.7). The results show that both the magnitude and trend of deforestation delineated from the map products differed greatly from reference-based area estimates. In Peru, Tanzania and Vietnam, the map-based deforestation estimates of the reclassified map product are closer to the reference-based estimates than those of the two original products. In addition, in four out of five cases[5] the reclassified product reveals the same deforestation trend as the reference-based area estimates, although the magnitude of change differed (table 3.5). This reflects the added value of using a combined product over a single product, although accuracy assessment thus remains necessary. As table 3.5 shows, in three out of five sites the direction of the deforestation trend according to the best reclassified product was different from at least one of the individual products, which would have had major implications if results-based payments would be based on a single product alone and disregarding the product's map accuracies and estimate uncertainties.

The majority of the map-based estimates (both the two input products and reclassified product) fell outside the 95% CI of the reference-based area estimates of both change classes, which affirms the importance of doing a (sample-based) validation of the map products. At the Indonesia-B site, the accuracy assessment elucidated the direction of performance considerably, as the 95%CIs around the reference-based area estimates are relatively small. Here, the average

[5]Indonesia-A being the exception, here the area estimates showed a slight decrease, while the *I-sensitive early* product showed a slight increase.

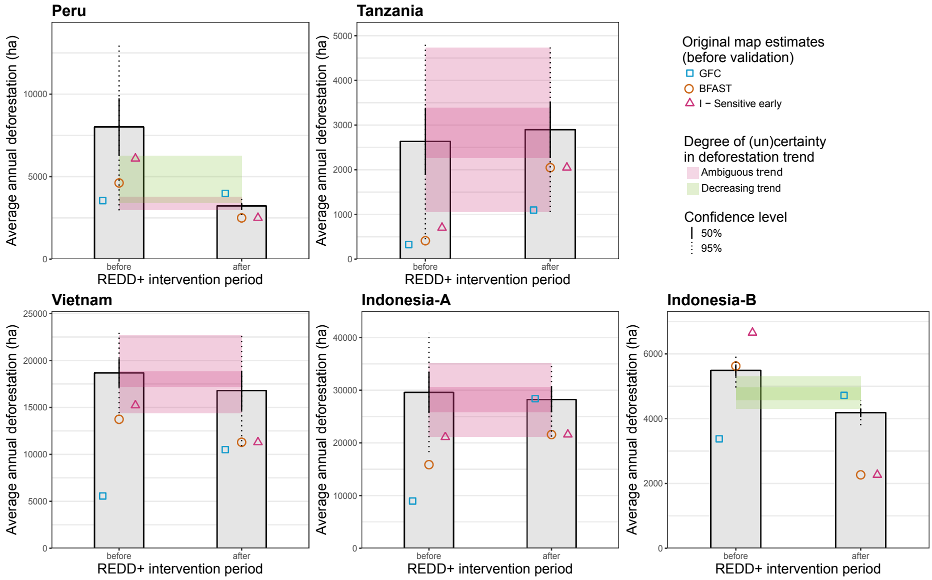


Figure 3.7: Influence of accuracy assessment and area estimates' uncertainty on REDD+ performance measurements. The grey bars represent the average annual deforested areas (reference-based area estimates), with 95% CIs. We corrected the CIs for differences in the number of years between the *before* and *after* period, assuming variances to be equally distributed in time. The selection of best performing reclassified product is based on the highest relative accuracy change, excluding the two V-timeless reclassified products, leading to I-sensitive early for all sites. The pink shaded areas represent the remaining degree of uncertainty, in which the direction of the deforestation trend remains ambiguous after considering the accuracy assessment. There is no overlap in the CIs of Peru (50%CI) and Indonesia-B (both 50%CI and 95%CI), hence the absence of a pink shaded area. The green shaded areas in those sites represent the downwards trend in deforestation, without overlap of CIs.

annual deforestation decreased from the *before* to the *after* class, while the corresponding CIs did not overlap, indicating a clear downwards trend in deforestation (green shaded area in figure 3.7, 3.5). At the site in Peru, both CIs in the *before* period are relatively large, but at a 50%CI a clear downwards trend in deforestation was found, as the CIs did not overlap. In all other sites, uncertainty in the direction of performance remained, since the CIs of the *before* and *after* period overlapped, as illustrated by the pink shaded area in figure 3.7.

Uncertainty of the deforestation trend

In addition to the visual assessment of uncertainty of the trend, we quantified the uncertainty of the trend's magnitude. Therefore, we estimated the trend uncertainty (two rightmost columns of table 3.5), which is based on the joint variance of the two monitoring periods, and is expressed in percent points (GOFC-GOLD, 2016). As an example, in Vietnam the reference-based area estimates revealed an average annual decrease in deforestation of 10% with a trend uncertainty

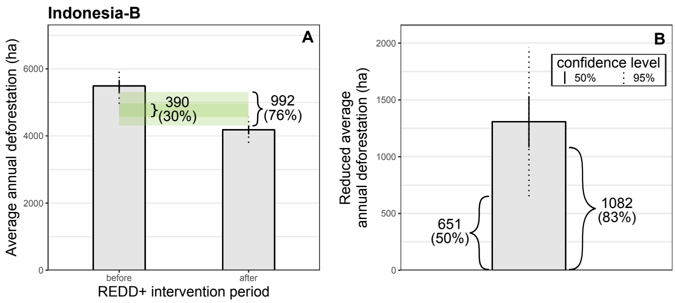


Figure 3.8: Conservativeness principle applied to calculate the REDD+ estimates for Indonesia-B. With approach A (left) one prevents overestimation of the reference estimates (*before* period) and underestimation of the assessment period (*after* period). Estimates in approach B (right) are derived from the uncertainty of the trend. Numbers next to curly brackets show the conservative REDD+ estimate (activity data only) in ha assessed at the 95%CI and 50%CI, and as percentages of the trend from the reference-based area estimates (grey bar in B).

of ±13% points at the 50%CI. Thus, at this confidence level an actual increase in deforestation of 3% is one of the possibilities. In Indonesia-B, the absence of overlapping CIs revealed a downwards trend. According to our area estimates this average annual decrease is 24% with a trend uncertainty of ±12 and ±4% points at a 95% and 50%CI respectively.

Applying the conservativeness principle to the REDD+ estimate – a case study

As illustrated above, deforestation estimates are subject to uncertainty, which is why one should be conservative when accounting for REDD+ in order to increase its credibility despite those uncertainties (Grassi et al., 2008). In other words, one should take into account the data uncertainties to prevent overestimation of the reduced emissions. Grassi et al. (2008) present four approaches to account for data uncertainty using the conservativeness principle, of which we apply two (A2 and B1 in Grassi et al. (2008), here referred to as A and B respectively) to our deforestation estimates of Indonesia-B. As figure 3.8 shows, both the approach and confidence level chosen have a great impact on the REDD+ estimate, with conservative estimates of reduced annual deforestation ranging from 390 to 1082ha, or 7 to 20% respectively.

3.4 Discussion

Most likely any deforestation map contains classification errors (Olofsson et al., 2013), and deforestation area estimates from these maps would thus differ from reality. We showed that a systematic accuracy assessment is critically important to value the usefulness of wall-to-wall forest cover change datasets for local REDD+ performance measurements. Distinguishing between overall, user's and producer's accuracy allowed comparison of different maps and helped to understand to what extent a map is likely to over- or underestimate real deforestation. The subsequent analysis of the variances and CIs showed to what extent the deforestation

	GFC[1]	BFAST[1]	I-Sensitive Early[1]	Area estimates[2]		
	change[3] (%)	change[3] (%)	change[3] (%)	change[3] (%)	Trend uncertainty with 95%CI (%.)[4]	Trend uncertainty with 50%CI (%.)[4]
Peru	12	-46	-59	-60[*]	63	22
Tanzania	238	397	192	10	109	37
Vietnam	89	-18	-26	-10	39	13
Indonesia-A	217	36	2	-5	45	16
Indonesia-B	40	-60	-66	-24[**]	12	4

[1] Original map estimates before accuracy assessment.
[2] Area estimates based on the reference data, see also figure 3.7.
[3] Degree of change (%) when comparing average deforestation in after period with average deforestation in *before* period. A negative number signifies a decrease in (average annual) deforestation over time.
[4] Uncertainty (U) of the trend in percent points is calculated as follows:
$U = \frac{CI(ha)}{def_{bef}(ha)}$. Where:
$CI(ha) = \sqrt{\frac{var_{bef}}{n_{bef}^2} + \frac{var_{aft}}{n_{aft}^2} * TotalArea^2} * z$;
var_{bef} and var_{aft} are the variance of the area proportion of the classes *before* and *after* respectively;
n_{bef} and n_{aft} are the number of years in the *before* and *after* period respectively;
z is 1.96 and 0.67 for the 95%CI and 50%CI respectively; and
$def_{bef}(ha)$ and $def_{aft}(ha)$ are the estimated deforestation in hectares in the *before* and *after* period respectively.
[*] no overlap between 50%CIs of the *before* and *after* estimates (here: decreasing deforestation trends).
[**] no overlap between 50%CIs and between 95%CIs of the *before* and *after* estimates (here: decreasing deforestation trends).

Table 3.5: Direction and degree of deforestation trend for the GFC, BFAST and reclassification strategy I map estimates, and for the area estimates using reference data. Trend uncertainty is indicated for the reference-based area estimates.

3

estimates remained uncertain. Furthermore, in this multi-site analysis, we assessed if, how and where a combination of forest cover change datasets can help to increase the accuracy and reduce the uncertainty of deforestation estimates for measuring the performance of local REDD+ initiatives.

We found high overall accuracies but striking differences in user's and producer's accuracies and area estimates, which is in line with findings from Melo et al. (2018) in Guinea-Bissau. In our multi-site study however, large regional differences appeared in the degree of discrepancy between the map products, with notable differences in the producer's accuracies particularly. Several recent changes were missed by BFAST leading to a lower PA in the after period, while BFAST's PA outperformed the GFC product in the first years of the monitoring period. Combining forest cover change datasets using a *sensitive-early* strategy generally improved accuracies and reduced uncertainties despite expected trade-offs between different types of accuracies. That is, although a *sensitive-early* strategy led, as expected, in three of the sites to slightly lower user's accuracies in the change classes because of a small increase in commission errors, the OAs and PAs increased more than the UAs deteriorated. Still, only in cases where the individual datasets showed reasonable to good accuracies, combining datasets led to a map product that was more accurate than the individual datasets, as low accuracies in one dataset could not be compensated by high accuracies in the other.

We found differences in timeliness of deforestation detection between GFC and BFAST, although these differences were not unidirectional across all sites. As stated in section 3.2.1, in the Indonesian sites, the lower PA of BFAST indicates omission errors in the after class, while in the Vietnamese site, BFAST appears to detect recent changes better than GFC does. Both GFC and BFAST appeared to have issues with a timely detection of deforestation due to mining, leading to errors of omission, while the visual validation with false-colour images showed easily detectable changes. More research is needed to verify if there is a correlation between the time series bands and corresponding vegetation and moisture indices, and their fitness to detect mining.

With our stratified sampling design (section 3.2.4) there was less risk of overlooking missed deforested pixels (i.e. missed omission errors), as conflicting pixels (in which change is detected in one, but not in the other product) were included in the sample as a separate stratum. On the downside, this might have led to an overestimation of omission errors due to the large area weight of stable forest classes in stratified sampling. Although our error matrices accounted for disproportional sampling of conflicting pixels, it is still likely that the producer's accuracies of both products were negatively influenced by this sampling design. At the same time, due to our sampling design we may have missed some omission errors in the non-forest class, i.e. pixels that were (erroneously) not included in the initial forest mask but in fact deforested. We focus on the (in)correct classification of change or no-change within the (initial) forest however, rather than the initial classification of forest or non-forest. We thus compared the change products in itself, and not differences in (or the accuracy of) forest masks. As the reference-based area estimates are only based on the reference samples due to the applied method (Stehman, 2014), the sampling design has a great influence on the results. Increasing the sample size further would reduce the uncertainty in the area estimates.

We focused on the uncertainty in performance assessments as caused by the underlying forest cover change dataset(s). Yet, uncertainty may come from more sources, including the precision and influence of the REDD+ initiative start year. We aggregated the (sub)annual deforestation detections into three classes, that is, *change before* REDD+, *change after* REDD+ started and *no change*. This rather sudden, and mainly theoretical, transition from the *before* to *after* class may have influenced our accuracies estimates. In practice, many local REDD+ initiatives are continuations of earlier integrated conservation and development projects, so interventions towards protecting local forests may predate the official start dates (Sunderlin et al., 2015). Since transitions in forests and forest use are often gradual processes too, this complicates performance assessments even further. Longer time series might be needed to clearly show the impact. Finally, all accuracies calculated are relative to the reference dataset, which is in this case created through the visually validated samples. Errors in the classification through visual validation were limited by using multiple time series data sources (e.g. RapidEye, Landsat TM) and multiple tools (i.e. TimeSync and CollectEarth).

It is important to note that for each site, a right-angled AOI was defined using the initiative's boundaries and a buffer (figure 3.3). Therefore, the AOIs included more than the 'pure' REDD+ intervention areas. Our objective was to explore the potential of combining activity datasets for accuracy improvement, and to demonstrate the implications of ignoring data uncertainties for performance measurements, rather than to calculate (change in) deforestation and corresponding carbon emissions for individual sites or to assess actual performance of specific initiatives. The results presented in section 3.3.3 should therefore not be used to assess the performance of these REDD+ initiatives as such.

3.5 Conclusions

We analysed the differences in accuracy and uncertainty between two forest cover change datasets for five sites and studied if and how combinations of datasets can increase accuracies and reduce uncertainties in the context of local REDD+ performance assessments. We demonstrated the use and usefulness of these global products to assess forest cover loss at local level.

How do forest cover loss datasets differ in terms of accuracy?
We found that accuracies differ at site level, although on average neither GFC nor BFAST excelled in accuracy. In the sites in Peru, Tanzania and Vietnam, BFAST performed better, while in the Indonesian sites, GFC achieved higher accuracies. Both GFC and BFAST underestimated deforestation, as reflected by the lower producer's accuracies and corresponding higher errors of omission.

What is the complementarity of these forest cover loss datasets in increasing accuracy?
Knowing the strengths and weaknesses of the individual products, we assessed their complementarity by overlaying the two products using different reclassification strategies. The strategy that led to the highest accuracy increases and uncertainty decreases differed per site. In four out of five cases, a *sensitive-early* strategy led to higher accuracies compared to the individual products. Only when both products' individual accuracies were already reasonable to good, a reclassification strategy resulted in higher accuracies. Products with low accuracies could not be ameliorated by any of our reclassification strategies.

How do map accuracy and uncertainty influence REDD+ performance assessment?
We show the influence of input data accuracies and remaining uncertainties in annual deforestation estimates on REDD+ performance assessment and demonstrate the importance of accuracy assessment. As the overlap in CIs indicated, in three out of five sites some degree of uncertainty in the deforestation trend remained, even after accuracy assessment. In one site, the accuracy assessment revealed a clear downwards trend in deforestation. In one other site, the (absence of a) clear downwards trend was dependent on the confidence level chosen. In three sites, the annual deforestation estimates of the reclassified product were closer to the reference-based estimates when compared to the estimates of GFC and BFAST. Still, these map-based estimates were mostly outside the 95%CI of the reference-based estimates, thus affirming the persisting need for validation. But even reference-based estimates are subject to uncertainty, thus leading to a need to be conservative in the accounting of corresponding REDD+ estimates.

3

The growing availability of global, readily available datasets and tools is of vital importance as local implementers' monitoring capacities are often limited. Our comparative study shows that consideration of and transparency about accuracies, (un)certainties and corresponding (dis)abilities of datasets and tools, is of key importance if results-based payments are to be based upon these performance measurements. Being conservative in REDD+ accounting could allow for these uncertainties and thus increase the credibility of the REDD+ estimates. To get insights in, and ultimately reduce, uncertainty, we showed that the value of sample-based accuracy assessments cannot be overstated.

3.6 Acknowledgements

The authors thank Tom Bewernick, Mathieu Decuyper, Ben DeVries and Erik van Schaik for their help with data processing; Mark Jorritsma for his assistance with the accuracy assessment; Sytze de Bruin and Valerio Avitabile for the helpful discussions; and we greatly appreciate the comments from two anonymous reviewers who helped improve the manuscript.

3

4 Integrated assessment of deforestation drivers and their alignment with subnational climate change mitigation efforts

This chapter is based on:

Bos, A. B., De Sy, V., Duchelle, A. E., Atmadja, S., de Bruin, S., Wunder, S., and Herold, M. Integrated assessment of deforestation drivers and their alignment with subnational climate change mitigation efforts *(in review, Environmental Science & Policy)*

Abstract

Efforts to reduce emissions from deforestation and forest degradation and enhancing forest carbon stocks (REDD+) have evolved over the past decade. Early REDD+ programs and local/subnational projects used various interventions (i.e. enabling measures, disincentives and incentives), implemented by government, the commercial and non-commercial private sector, but are currently understudied vis-à-vis their effectiveness to address site-specific drivers of deforestation and forest degradation (DD). We assess how well REDD+ interventions addressed DD at five project sites in Peru (1), Brazil (1), Vietnam (1) and Indonesia (2). Our study design includes an integrated assessment of remotely sensed, spatially modelled, and locally reported drivers. First, we observe follow-up land use from high resolution imagery as proxy for direct deforestation drivers. Second, spatial Random Forest modelling of DD drivers allows for influence quantification of topographic, climatic and proximity variables at each site. Third, we report direct and indirect DD drivers from pre-intervention surveys and semi-structured interviews with five REDD+ implementers, 40 villages and 1200 households. Data gathered included perceived changes in forest cover and quality, and their causes. We found general agreement between observed, modelled and reported local DD drivers, yet some were inadequately addressed by interventions. Intra-site differences in drivers underscores the importance of analysing micro-level DD drivers. Our interdisciplinary approach reveals the complexities of local direct and indirect DD drivers, and the complementarity of remotely sensed, spatially modelled and locally reported methods for driver identification. A better understanding of the alignment between DD drivers and REDD+ interventions is vital for practitioners and policy makers to enhance the effectiveness, efficiency, equity and co-benefits of REDD+ at the local level.

4.1 Introduction

Deforestation and other land use changes contribute significantly to carbon emissions (IPCC, 2006). Efforts to reduce emissions from deforestation and forest degradation and to enhance carbon stocks (REDD+) were embedded in the Paris Agreement (UNFCCC, 2015). To design effective policies, it is important to know: what land use change activities are happening; who are the agents linked to these changes; and what underlying forces are at play?

Numerous conceptual models can be used to understand the drivers of deforestation and forest degradation and their interactions. Geist and Lambin (2002) focus mainly on distinguishing proximate ("direct") and underlying ("indirect") causes, whereas Wood and Porro (2002) put more emphasis on the distinction between biophysical and socio-economic factors at different spatial scales. The approach of Kaimowitz and Angelsen (1998) is more similar to Geist and Lambin's, although the focus differs by concentrating on the economics behind the immediate and underlying factors. It is important to monitor drivers of deforestation and forest degradation at the local level because they differ across space and time (Rudel, 2007; Rudel et al., 2009; Defries et al., 2010; Hosonuma et al., 2012; De Sy et al., 2015; Curtis et al., 2018).

The methods to assess drivers are nested in different scientific disciplines. They range from visual assessment of land use, land cover, and changes therein (LULCC) (e.g. De Sy et al., 2015), socio-economic survey data collected in the field (e.g. Walker et al., 2002), to machine learning techniques assessing the relative importance of spatial factors explaining land cover change (e.g. Zanella et al., 2017). Each of these methods have their strengths and weaknesses in terms of the driver elements (e.g. agent, location, extent) that they can accurately assess. Remotely sensed imagery can help to identify the land cover following deforestation, which can then be used as proxy for the direct driver (De Sy et al., 2015). Recent technical innovations in remote sensing and forest-relevant monitoring techniques have resulted in national and global datasets with increasing levels of coverage, spatial and temporal detail and accuracy (Bos et al., 2017), which can capture changes in forest cover, including land uses following deforestation. Socio-economic data can complement these remote sensing techniques in helping to identify the agents or underlying factors at play. Spatial modelling with machine learning techniques, such as Random Forest modelling, provide powerful tools to reveal underlying spatial factors influencing DD. When used in isolation however, they lack the ability to provide a meaningful interpretation of these results. Rather than to compare the capabilities of each of the methods, we argue that an assessment of their complementarity is more valuable as combined, interdisciplinary approaches provide better understanding of the processes at stake than single-source approaches.

Information on drivers can help determine the appropriate policy interventions to address those change processes (Finer et al., 2018). As the activities leading to DD differ between continents and countries (Hosonuma et al., 2012; De Sy et al., 2015), there is no single intervention to address all drivers effectively (Seymour and Harris, 2019). Similarly, REDD+ interventions vary greatly in terms of type and implementer. Ideally, interventions are tailored to the local context (Godar et al., 2014; Austin et al., 2019), which requires an integrated assessment of relevant drivers. Information on drivers is therefore beneficial in all stages of the REDD+ design, implementation and evaluation De Sy et al. (2018). Incorporating this type of information is not straightforward, however, as recurrent monitoring is complex and costly.

The objectives of this study (represented by the linkages *between* the triangles in figure 4.1) are, (1) to assess the complementarity of different data sources in providing information on DD drivers; (2) to identify the most prevalent DD drivers in our study sites; and (3) to identify possible (mis)matches between the pre-identified DD drivers and REDD+ interventions. However, the aim of this study is not to assess how successful these interventions are in addressing these DD drivers, as this requires an impact assessment, which goes beyond the scope of this study. The topic of impact assessments is addressed in more detail in chapter 2 and 5. In order to reach our objectives, we will address the following research questions (represented by the elements *within* the triangles in figure 4.1):

1. To which land cover and land uses are forests converted, based on high resolution imagery?
2. What are the most important topographic, climatic and proximity variables explaining deforestation, based on a Random Forest Model?
3. What are the dominant locally reported direct and indirect DD drivers, based on household, village, and key informant interviews?
4. Which DD activities and agents are targeted by the REDD+ interventions?

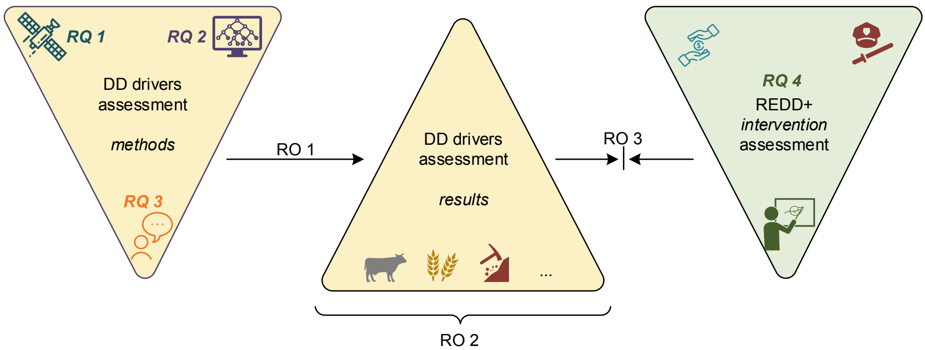


Figure 4.1: Relationship between research objectives (ROs), and research questions (RQs)

4.2 Material and methods

4.2.1 Conceptual framework

Figure 4.2 shows the conceptual framework of this study, which builds upon existing LULCC models as discussed in the previous section, but puts the activity inducing DD at the centre. In this way, we provide a holistic approach that can be used in both spatial and non-spatial assessments.

In our conceptual model, *drivers* are defined as the interplay of agents, land activities and underlying forces that lead to deforestation and forest degradation. *Agents* refer to entities performing land activities on the ground and include smallholders and communities, large agricultural land holders, large scale agribusinesses and logging or mining companies. *Activities* are human actions that lead to forest change (e.g. agricultural expansion, logging, infrastructure expansion) often referred to as *direct drivers* (see for example De Sy et al., 2018). *Environmental factors* consist of biophysical or topographic elements that allow or limit certain activities (e.g. slope, availability of soil minerals, etc.) but which in essence cannot be influenced by humans through policies or other interventions. *Underlying forces*, such as economic and political processes, are often complex and can interact with each other. They directly or indirectly influence the decision-making of the agents (e.g. farmers, government agencies, agricultural or mining companies etc.) who are performing the activities.

REDD+ interventions can be divided into three types (i.e. *enabling measures*, *disincentives* and *incentives*), which can be applied by different types of implementers (government, non-government, and private sector actors) at different levels (e.g. national and subnational programs vs. local level REDD+ projects). While incentives (e.g. payments for environmental services) and disincentives (e.g. command-and-control measures) aim to change agents' decision-making in terms of forest change activities, enabling measures (e.g. tenure clarification, environmental education) can influence agents, and thus indirectly their activities, or underlying forces.

4

4.2.2 Study areas

In our study we focus on five sites, located in four countries in Latin America and Asia (figure 4.3, table 4.1). These five sites are part of CIFOR's Global Comparative Study on REDD+ (GCS-REDD+), and were selected to represent a wide range of intervention types ((dis)incentives and enabling measures), implementer types (government, non-governmental organization (NGO), private sector), and geographies across the tropics (CIFOR, 2017). Further, data availability constraints concerning the availability of different forest change map products affected the final selection (Bos et al., 2019).

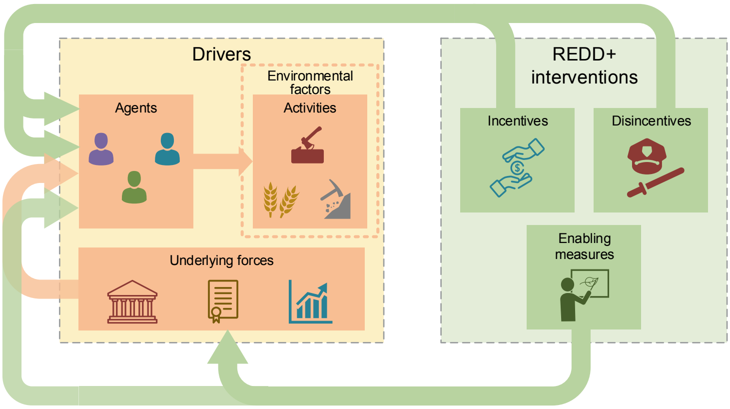


Figure 4.2: Conceptual framework. Orange arrows represent interactions between different driver elements. Green arrows represent how different REDD+ interventions envision to influence these elements or its interactions.

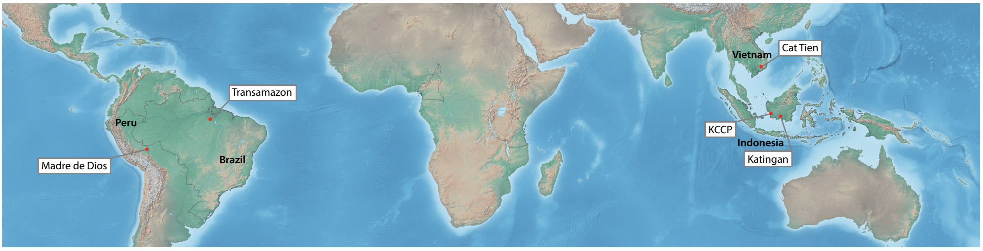


Figure 4.3: Study areas

Sustainable Settlements in the Amazon

From 2012 to 2017, this project targeted smallholders in the Transamazon Highway region (Eastern Brazilian Amazon) to promote sustainable agricultural practices and was implemented by the NGO Amazon Environmental Research Institute (IPAM). In 2000, forest cover was 95.4% but 19.2% points were lost during 2001-2012 (Sunderlin et al., 2014b). Smallholders sampled had 69% forest cover on their landholdings in 2010 (Duchelle et al., 2014), earning their income mostly from cropping and livestock, and clearing forest mostly for crops (Cromberg et al., 2014b). Interventions focused on more sustainable agriculture and economic compensation (Simonet et al., 2019).

REDD+ Project in Brazil Nut Concessions (BAM & FEPROCAMD)

The objective of the REDD+ project in Madre de Dios (MDD), Peru was to provide incentives for Brazil nut concessionaries to conserve the forests on which they depend. This area is heavily forested (99% in 2000) with very low deforestation, that is only 0.3% point loss from 2001 to 2012 (Sunderlin et al., 2014b). Brazil nut producers in the area glean most of their local income from forests, including Brazil nuts and timber (Garrish et al., 2014). The project began in 2009

Table 4.1: Site characteristics

Country	Site	Area of interest (in mln ha)	Ecozone (FAO)	REDD+ start year	REDD+ end year	Implementer type
Brazil	Transamazon[1]	4.8	TRF[6]	2013	2017	NGO
Peru	Madre de Dios (MDD)[2]	1.1	TRF[6]	2009	*Ongoing*	Private sector
Indonesia	KCCP[3]	2.0	TRF[6]	2008	*Ongoing*	NGO
Indonesia	Katingan[4]	3.6	TRF[6]	2009	*Ongoing*	Private sector
Vietnam	Cat Tien[5]	0.8	TRF[6]/ TMDF[7]	2009	2012	NGO

1 Sustainable Settlements in the Amazon (IPAM)

2 REDD+ Project in Brazil Nut Concessions (BAM & FEPROCAMD)

3 KetapangCommunity Carbon Pools (FFI)

4 Katingan Peatland Restoration & Conservation Project (PT.RMU)

5 Cat Tien National Park Pro-Poor REDD+ Project (SNV)

6 Tropical rainforest

7 Tropical moist deciduous forest

as a collaboration between the private company Bosques Amazonicos and the local Brazil nut producers' federation and targeted 405 concessionaries over 308,757 ha (BAM, 2012). It was validated by Verified Carbon Standard (VCS) in 2012 and sold 1.5 million verified carbon units through the voluntary market.

Ketapang Community Carbon Pools (FFI)

The Ketapang Community Carbon Pool (KCCP, referred to as Indonesia-A in chapter 3) is a forest carbon initiative of Fauna and Flora International (FFI) Indonesia Programme. The lowland and peat swamps in this area in West Kalimantan experienced 4.6% forest loss in the period 2001-2012, threatening biodiversity and carbon-rich tropical forests (Sunderlin et al., 2014b; Intarini et al., 2014). Started in 2008, the NGO focusses on arranging community forest rights for local villages, aiming to strengthen communities' tenure security and counter threats from large-scale external actors (Intarini et al., 2014).

Katingan Peatland Restoration & Conservation Project (PT.RMU)

The Katingan Peatland Restoration & Conservation Project, currently known as the Katingan Mentaya project and referred to as Indonesia-B in chapter 3, was founded in 2007, and is managed by PT Rimba Makmur Utama (PT.RMU), a private company based in Indonesia (Indriatmoko et al., 2014). The villages collaborating in the project are adjacent to Sebangau National Park. The REDD+ project site is largely forested and experienced 2.6% forest loss in the period 2001-2012 (Sunderlin et al., 2014b). The main project strategy is to protect an entire peat hydrological unit (i.e. 'peat dome') by converting the status of the land into a restoration concession and supporting communities with locally suitable and sustainable income-generating activities.

Vietnam - Cat Tien National Park Pro-Poor REDD+ Project (SNV)

This project (2009-2012) was initiated by SNV (the Netherlands Development Organisation) as a REDD+ readiness project to assess the opportunity for accessing the voluntary carbon market and to establish a forest carbon facility in participation with local villagers. In the project area, 58%–71% of villagers interviewed considered agriculture as their primary or secondary occupation (Huynh, 2014). Their largest proportion of land consist of secondary forest, followed by agriculture. Natural forests are owned by the government. The forest cover in this area is high (94.5% in 2000), with 5.3% forest loss from 2001-2012 (Sunderlin et al., 2014b). The REDD+ readiness interventions primarily focused on carbon monitoring and participatory forest monitoring trainings (Huynh, 2014).

4.2.3 Summary of workflow

The workflow of this study is visualized in figure 4.4. In this section, the elements are introduced briefly, and will be discussed in more detail in the following sections. This study consists of three parts, that is, (1) a DD drivers, (2) a REDD+ interventions, and (3) an alignment assessment. The DD drivers analysis uses three methods which build upon different data sources. Here, insights from high resolution imagery, spatial modelling and socio-economic surveys jointly provide insights in the DD drivers of the different study sites. The REDD+ intervention assessment builds upon village level survey data and a database containing information on REDD+ interventions in the different study sites. The DD drivers analysis formed the basis for the assessment of the complementarity of different (disciplinary) methods and datasets. Finally, we assessed the alignment of the DD drivers and REDD+ interventions.

4.2.4 Remotely observed land cover and land use patterns after DD using high-resolution imagery

4

For the first research question (figure 4.1, section 4.1), we used tree cover loss data based on a combination of the Global Forest Change (GFC) dataset (Hansen et al., 2013, version 1.3) and the Breaks For Additive Season and Trend (BFAST) algorithm (Verbesselt et al., 2010, 2012). For methods and sampling design regarding the forest loss detection, we refer to Bos et al. (2019)[1]. We define deforestation as a conversion from land above a certain tree cover percentage and covering more than a certain MMU[2] to land with very limited or no tree cover. Therefore, we follow the land cover definition of deforestation, which is more practical to assess, rather than a land use definition of deforestation (Seymour and Busch, 2016). Forest degradation refers to a decrease in quality of certain features of the forests while the predominant land cover and land use remains forest. In this study, degradation is exemplified by a reduction in tree cover, while still exceeding the threshold of the corresponding forest definition.

Follow-up land use or land cover after DD was used as proxy for the direct driver of

[1]In the original study, the sample size was 270 pixels for each of the sites, and included both forest loss and stable forest pixels. For this particular study, we only focussed on the forest loss pixels, which led to slightly different sample sizes for each of the sites, that is, n=197 for Brazil-Transamazon; n=203 for Peru-Madre de Dios; n=206 for Indonesia-Ketapang Community Carbon Pools (KCCP); n=203 for Indonesia-Katingan; and n=227 for Vietnam-Cat Tien.

[2]Following national forest definitions, source UNFCCC (2009). For specific thresholds used, see Bos et al. (2019). Forest definition used for Brazil-Transamazon is >=10% tree cover and an MMU of 0.5ha.

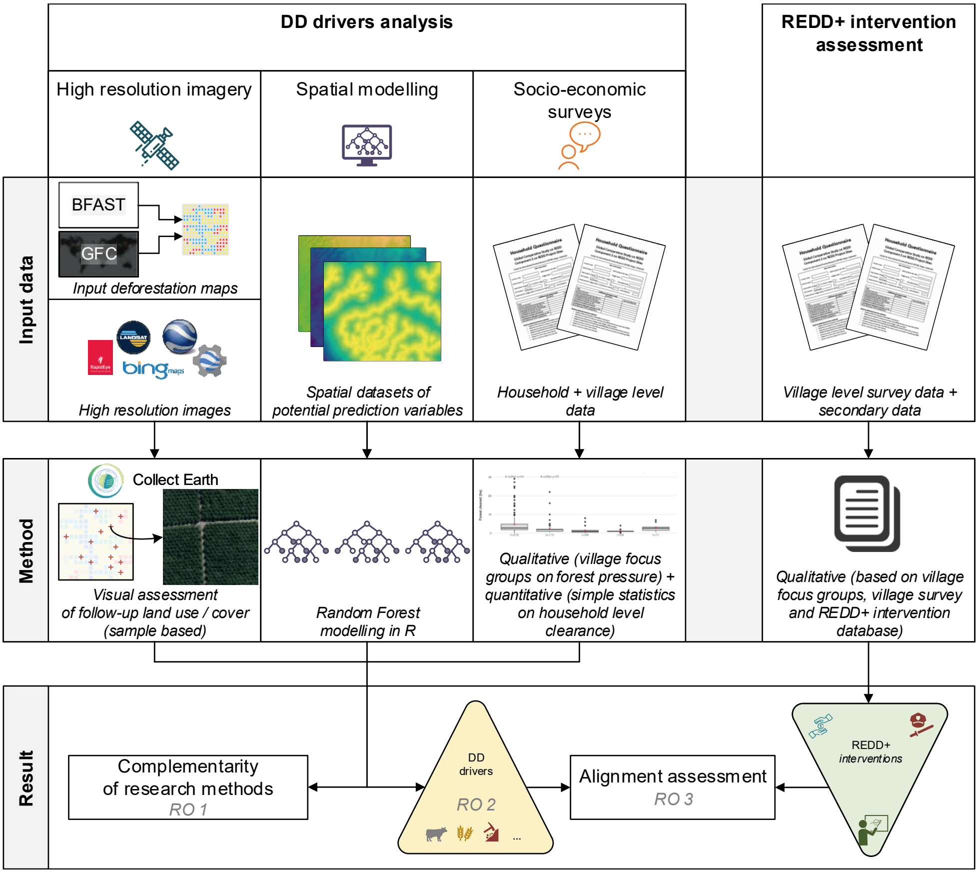


Figure 4.4: Workflow

deforestation[3]. To assess land use following deforestation, we assessed the forest loss samples from Bos et al. (2019), and determined follow-up land use using high-resolution imagery, consisting of Google Earth and RapidEye imagery. In addition, time series data of Landsat TM was assessed to clarify certain land use patterns. Although the spatial resolution of these data is limited (30m), in cases of large-scale land conversion (such as tree crop plantations) and limited availability of high-resolution imagery, Landsat TM was often sufficient to validate follow-up land use. For recording the follow-up land use, we developed a survey using Open Foris Collect (Open Foris, 2019). For each sample, the confidence level was recorded. When a sample's land

[3]We acknowledge that in certain areas, the first follow-up land use may not always reflect the main driver of forest clearance (e.g. in Amazonian areas where forest loss is often followed by cropping, but long term land use consists of pasture. Likewise, in Indonesia, deforestation can be followed by rice crops, while this is only temporary until their rubber trees mature), but emphasize that longitudinal high resolution imagery or other methods such as household level surveys may better reveal these type of processes.

use or land cover was confirmed with multiple imagery data sources, a high confidence level was given. Samples for which no decisive follow-up land use or land cover could be given due to data limitations or other reasons were assessed by an additional independent remote sensing expert or local expert. When uncertainty remained, samples were marked with a low confidence level. Follow-up land use classes were aggregated into four classes (table 4.2). The relative size of each class was calculated using Stehman's methods, while taking into account unequal sample class distributions (Stehman, 2014; Bos et al., 2019).

Table 4.2: Observations and corresponding aggregated classes of follow-up land use and land cover

	Aggregated class			
	Degradation	**Tree plantations**	**Agriculture**	**Other**
Observations	· Burned areas · Selectively logged areas (but remains predominantly forest) · Regrowth (partial)	· Rubber[1] · Palm oil[1] · Large scale tree plantations	· Crops · Cattle pastures · Shrub mix farm · Small scale agroforestry systems (incl. orchards, coffee etc.)	· Mining · Water ○ change in river flow ○ hydropower reservoir) · Road infrastructure · Buildings
Example				

1 To align with findings from the socioeconomic data we decided not to aggregate these under the "agriculture" class, as according to reported data, agents linked to these conversions often differ from agents for (subsistence or small scale cash crop) agriculture.

4.2.5 Spatial modelling of underlying factors associated with forest loss

For the second research question, we created a Random Forest model (RF) to assess the relative importance of predefined spatial variables to predict deforestation. A RF is a non-parametric method based on classification or regression tree learning. Unlike many other spatial models, RFs are known for their robustness, reduced risk of overfitting, capability to deal with non-linear relationships between prediction variables, and ability to address interactions without explicitly defining them in the model (Breiman, 2001). The forest loss data (response variable) used differed across the sites, and was based on the map product with the highest accuracy as found in Bos et al. (2019)[4]. The predictor variables used are described in table 4.3. These topographic, climatic and proximity variables are known to play a role in land use and land cover change processes (e.g. Kaimowitz and Angelsen, 1998; Wood and Porro, 2002; Geist and Lambin, 2002), but their relative importance may differ in different contexts. Variable importance of these predictor variables was used as proxy for underlying forces of deforestation. Classification trees were computed for a binary categorical response variable (forest loss and stable forest). For each

[4]The GFC dataset (Hansen et al., 2013) was used for Brazil-Transamazon, Peru-Madre de Dios, and Indonesia-KCCP. The combined dataset *sensitive-early* (Bos et al., 2019) was used for Indonesia-Katingan and Vietnam-Cat Tien.

Table 4.3: Prediction variables for Random Forest model

Variable	Type	Unit	Source
Elevation	Topographic	Meters	CIAT-CSI SRTM (Jarvis et al., 2008)
Slope	Topographic	Degrees	Derived from elevation, see above.
Annual precipitation	Climatic	Millimetres	WorldClim 2 (Fick and Hijmans, 2017)
Annual mean temperature	Climatic	Temperature Celsius	WorldClim 2 (Fick and Hijmans, 2017)
Distance to agriculture	Proximity	Meters	ESA Climate Change Initiative Land Cover Map (2015)
Distance to roads	Proximity	Meters	OpenStreetMap
Distance to waterways	Proximity	Meters	OpenStreetMap

of the sites, 5% of non-NA pixels were sampled for training data. To weigh all misclassifications equally in the trained RF, balanced training samples were generated so that 50% of the training samples consisted of forest loss, and 50% of stable forest. For each site, the random forest consisted of 500 classification trees. The spatial predictor variables selected for this study were elevation, slope, distance to roads, distance to waterways, distance to existing agriculture, average annual temperature and average annual precipitation. Following Breiman (2001) and using the randomForest package (Liaw and Wiener, 2018) in R, the relative variable importance using the mean decrease in accuracy (MDA) was calculated by (1) computing the out-of-bag statistic with the data for the i-th predictor variable intact, (2) permuting the data for the i-th predictor variable, (3) recalculating the out-of-bag statistic using the permuted data for the i-th predictor, (4) calculating the difference. This procedure was repeated for all seven prediction variables. Accuracies of the prediction maps were calculated following Olofsson et al. (2014).

4.2.6 Socio-economic survey data for perceived direct and indirect drivers of deforestation

For the third research question (figure 4.1, section 4.1) we used data on reported direct and indirect DD drivers. These data were gathered during semi-structured interviews with REDD+ implementers, village-level focus groups (mixed gender and women's only), and household surveys. The surveys were conducted in 2010-2011 and targeted approximately 1200 households in 40 villages. Data gathered included forest regulations; perceived causes of forest cover/quality change; and household level clearance of forests and its purpose (appendix C.1). A complete overview of the questions asked and methods applied can be found in the technical guidelines (Sunderlin et al., 2016, 2010).

Survey data from village focus groups and household interviews were cleaned, aggregated and visualized using R. Simple descriptive statistics were calculated for the main household variables, while a qualitative assessment was done for the data collected from the village surveys. The assessment focussed on the following themes and variables: area (size) per land use, purpose of clearing, principal crop and crop type after clearing, forest area and forest quality change and perceived (exogenous) causes of forest cover change.

4.2.7 Assessment of forest-based interventions and alignment with DD drivers

Data from a survey of village interventions were used to document the most relevant forest interventions at each site (Sunderlin et al., 2016). During the second phase of fieldwork (2013/2014), the research team first compiled a list of all interventions that aimed to conserve or restore forests that were documented in the study villages in earlier interviews with implementers and village focus group discussions. That list was refined with REDD+ implementers, and then with key informants in all study villages following the methods outlined in Sunderlin et al. (2016). Forest interventions included not only those implemented by the REDD+ proponent, but also national/subnational policies and programs that affected local forest use at the study sites. For each forest intervention, information on agent (target stakeholder (group)), sector (e.g. forest, agriculture), and level (national, subnational or local) were recorded to assess the degree of alignment with the DD drivers results as found in the earlier parts of the study.

4.3 Results

4.3.1 Forest change patterns observed with remote sensing

Table 4.4 gives an overview of the relative shares of forest area conversions. The aggregated classes in table 4.4 are broadly defined, but there are cross-site differences within those classes. That is, agriculture in Brazil is marked by pasture lands mainly, while in Indonesia-Katingan this is mainly cropland, including rice. In Indonesia-KCCP, tree plantations constitute of oil palm plantations, unlike the tree plantations in Vietnam Cat-Tien. Figure 4.5 shows some of these cross-site differences within the four classes. Site-specific findings are given below. appendix C.2 reports the confidence levels for the classes per site.

Table 4.4: Area proportion (%) of follow-up land use and land cover classes

	degradation	tree plantations	agriculture	other
Brazil-Transamazon	10%	0%	87%	3%
Peru-MDD	59%	0%	35%	6%
Indonesia-KCCP	32%	50%	15%	3%
Indonesia-Katingan	51%	24%	16%	9%
Vietnam-Cat Tien	14%	30%	42%	14%

Brazil – Transamazon

The class *degradation* (n=13) constitutes mostly (n=11) of samples that were characterised by regrowth after forest disturbance. All samples marked as *agriculture* (n=177), were pastural

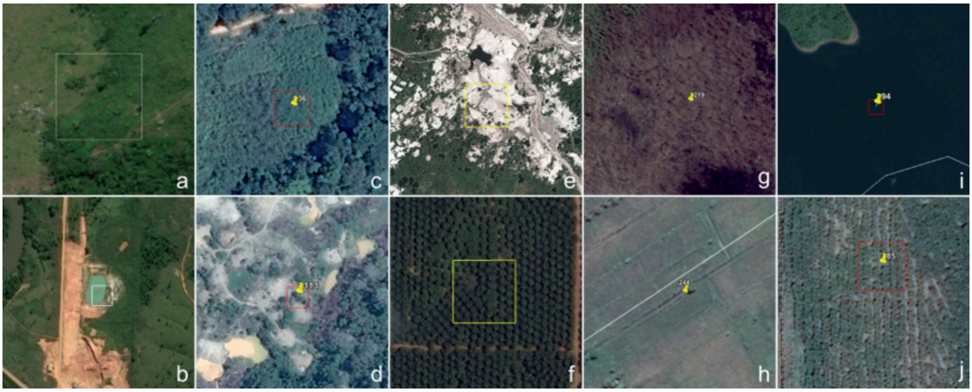


Figure 4.5: Examples of DD activities encountered. Brazil-Transamazon: (a) agriculture (pasture) and (b) other (infrastructure). Peru-MDD: (c) degradation and (d) other (mining). Indonesia-KCCP: (e) other (mining) and (f) tree plantation (oil palm). Indonesia-Katingan: (g) degradation (fire) and (h) agriculture (crops). Vietnam-Cat Tien: (i) other (hydropower reservoir) and (j) tree plantation.

lands, often marked with cattle and cattle tracks. The samples marked as *other* (n=7) were roads, buildings, or other infrastructures.

Peru – MDD

Samples classified as *degradation* were characterised by small scale disturbances after which some degree of regrowth was visible in the subsequent years. *Agriculture* consisted mainly of pastural lands (n=83) and to a lesser degree crops (n=16). Although the *other* class was relatively small (i.e. 6% of the total area of forest deforested, table 4.4), the spatial distribution of this class gave some clear insights (appendix C.3), with patches of mining, clearly distinguishable near the main river.

Indonesia-KCCP

Degradation in this site consisted of forest affected by fires, and logging after which regrowth occurred with a mixture of trees and small shrubs. *Tree plantations* consisted primarily of large-scale oil palm plantations, although often only marked several years after the deforestation disturbance was detected. *Agriculture* consisted of rice paddies and other crops. Conversions marked as *other* (n=11) were mostly cases of mining (n=7), and some conversions to infrastructure.

Indonesia-Katingan

Samples marked as *degradation* consisted mostly of partially logged plots and degraded forest at oil palm plantation edges. To a lesser degree, fires were noted, as well as some cases of partial regrowth after forest disturbance. *Tree plantations* consisted mostly (60 out of 62 cases) of oil palm plantations. Samples with *agriculture* were mostly small-scale croplands. The *other* class consisted of infrastructure (buildings) (n=3) and some cases of bare land (n=5) for which no other follow up land use was detected.

Vietnam – Cat Tien
Samples marked as *degradation* consisted of forests with clearly visible selective logging, and to a lesser degree recurrent disturbed forests with intermediate regrowth. A considerable amount of samples (n=44) were marked as large-scale monocultural *tree plantations*. *Agriculture* consisted mainly of cropland (n=110), including bushy crops, coffee and cashew trees. To a lesser degree, pastural lands were found (n=12) and some mixed areas with cropland and small-scale plantations (n=6). The *other* class consisted of infrastructure (buildings and roads, n=8), and flooded areas due to the building of a new hydropower dam (n=8).

4.3.2 Spatial modelling

Appendix C.4 shows the spatial distribution for each of the prediction variables, as well as comparisons between forest loss and stable forest pixels per site. Error matrices and corresponding error-adjusted areas were estimated and accuracies were calculated for all model predictions based on a comparison between the models' predictions and the input deforestation maps. The accuracies are listed in table 4.5, the error matrices can be found in appendix C.5. In general, the random forest models predicted deforestation well using the spatial layers as predictors, with overall accuracies exceeding 86%. The relative high overall accuracies build confidence in the RFs in general, as low accuracies in the models' predictions would also suggest that the variable importance findings would be less meaningful. The relatively low user's accuracies of forest loss class for Peru-MDD and Indonesia-KCCP indicate that the models overestimated forest loss at those sites. Here, the random forest models' predictions are thus well capable of modelling the spatial patterns of forest loss when using the available information from the prediction variables, but they are less capable of estimating the magnitude of forest loss.

4

Table 4.5: Accuracies of Random Forest Model predictions

	Brazil - Transamazon		Peru - MDD		Indonesia-KCCP		Indonesia-Katingan		Vietnam-Cat Tien	
	Loss	Stable	Loss	Stable	Loss	Stable	Loss	Stable	Loss	Stable
UA	93.5%	84.8%	34.1%	99.9%	57.3%	97.6%	74.7%	94.6%	66.3%	96.5%
PA	96.3%	75.5%	98.5%	89.8%	90.7%	85.0%	89.1%	86.4%	90.6%	85.1%
OA	91.8%		90.3%		86.0%		87.3%		86.4%	

UA = User's accuracy; PA = Producer's accuracy; OA = Overall accuracy

The MDA is an indicator of variable importance[5]. Figure 4.6 shows that in general, distance to existing agriculture, annual precipitation (i.e. micro-climate differences), and distance to

[5]MDAs refer to accuracies of single tree models, and should not be confused with the model accuracies of the map accuracies of the model's predictions as presented in table 4.5.

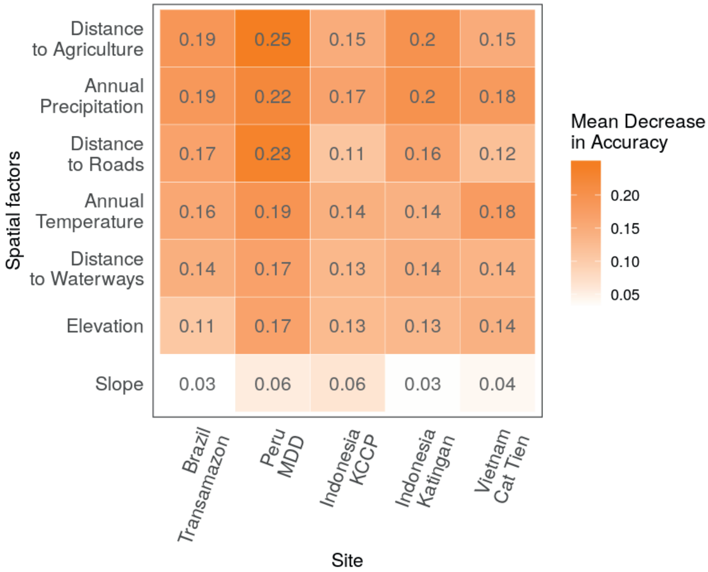


Figure 4.6: Random Forest models variable importance

roads are important spatial factors for explaining deforestation, although there are differences between sites. In Peru-Madre de Dios for example, the relative importance of distance to agriculture as deforestation predictor is higher than in Vietnam-Cat Tien, where it is ranked as the third-highest explanatory factor.

4

4.3.3 Socio-economic survey data for perceived direct and indirect drivers of deforestation

Perceived forest area and forest quality change at village level

During the mixed gender focus group discussions, the majority of villages reported a decrease in *forest area* in the past two years[6]. Forest quality was defined as the availability of goods and services of the forest related to density of woody material, forest health, and biological productivity and diversity, and is thus a proxy for forest degradation. The majority of the villages reported a decrease in *forest quality* in the past two years, with Vietnam-Cat Tien being the exception[7].

[6]There were eight village level focus groups at each site (total n=40). Decreased forest area was reported in six villages in Brazil-Transamazon and in Indonesia-Katingan, seven in Peru-MDD and in Indonesia-KCCP. Only in Vietnam-Cat Tien, a minority of the villages reported a decrease in forest area (n=2).

[7]For the same 40 village level focus groups, the majority of the villages in Brazil-Transamazon (n=7), Peru-MDD (n=8), Indonesia-KCCP (n=8) and Indonesia-Katingan (n=6) reported a decrease in forest quality. In Vietnam-Cat Tien, three villages reported no forest quality change, while the remaining five did not know or did not answer.

Perceived forest pressure sources at village level

During the mixed gender and women focus group discussions, participants were asked to report their perceived forest pressure sources in their village area and surroundings.

In Brazil-Transamazon pressures that were mentioned included incoming migrants who cleared for farmland. During the women's focus groups, the presence of logging companies (both small and large scale), seasonal migrants, people from neighbouring villages, and agro-industrial farms (cattle) were mentioned as forest pressure sources.

In Peru-MDD people mentioned unclear tenure rights in and outside Brail nut concession areas. In the women's focus group discussions, respondents mentioned the pressure from logging companies. One village noted that an agro-industrial firm (rubber) formed a pressure on the forest. We found differences in forest pressures between people who lived in areas close to the road, and people who live in more remote areas. In the former, pressures included incoming migrants for papaya plantations and timber (local logging companies). In the latter, people experience forest pressure due to gold mining.

In Indonesia-KCCP, respondents mentioned the presence of industrial companies for pulp and paper, soy and cattle. Also, some pressure due to swidden agriculture was mentioned. A large food estate project (rice, by the government), posed another pressure on the forested land. Some villages experience pressure due to the rise of oil palm plantations and illegal gold mining. Some villagers started rubber plantations as an effort to prevent conversion to oil palm by outsiders. Logging (legal) for housing infrastructure, logging by timber companies, and poaching were also mentioned, the latter in women's focus groups only.

In Indonesia-Katingan, pressures mentioned included palm oil companies, small scale (unregulated) mining and extension of agricultural lands and infrastructural developments due to stimulating government programs. In addition, women also mentioned poaching as putting a pressure on the forest resources, as well as people from both in- and outside the villages, small scale loggers, large scale cattle and rubber plantation firms, and timber plantation companies.

In Vietnam-Cat Tien, people mentioned that the government's forest allocation policies led to a conversion of degraded forest land to agroforestry systems by villagers.

Forest clearance and purpose by households

The previous section reported both exogenous and endogenous pressures on the villages' forests. Here, we focus on household level clearing of forests as reported in the household surveys, which represents endogenous pressures.

The results are visualised in figure 4.7. Whether or not households clear forests differs greatly between the sites. While in Brazil-Transamazon >75% of interviewed households report forest clearing in the past two years, in Vietnam-Cat Tien this is <5%. Also, the mean and median area of forest cleared differs widely, although there are large differences between the spreads within sites, as Brazil-Transamazon and Peru-MDD contain more outliers above the boxplot's maxima.

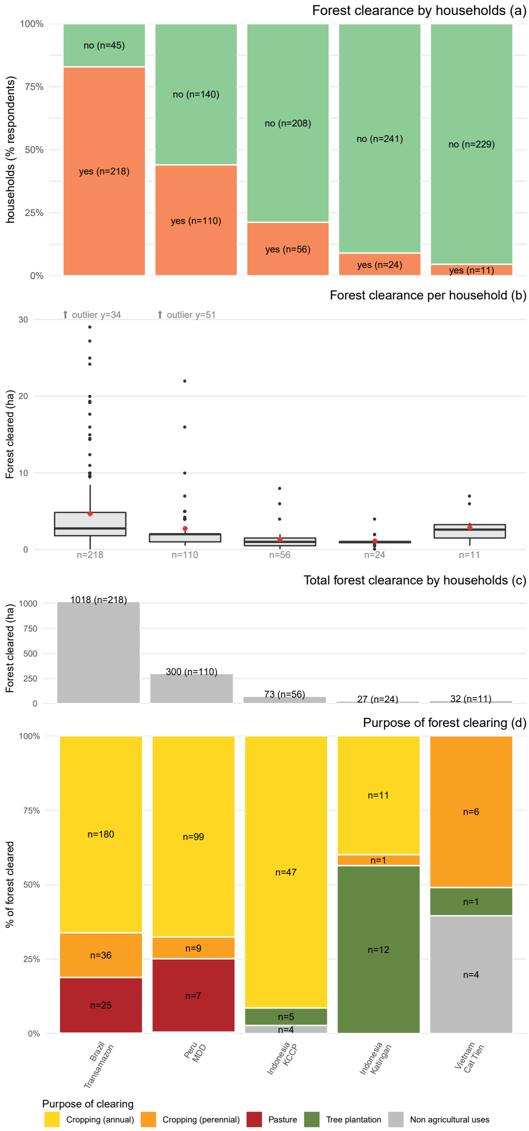


Figure 4.7: Reported forest clearance and purpose by households. (a) shows the response to the question "did your household clear any forest during the past 2 years?" (b) Forest clearance (area) by households that reported >0 ha clearance. Upper and lower extremes of whiskers represent Q3+1.5*IQR and Q1−1.5*IQR respectively, where IQR=Q3-Q1. (c) Total forest clearance by households (respondents only) per site (d) Follow up use of forest area cleared. n represents number of respondents per purpose category.

In all sites, household clearance is mostly for cropping. In the South American sites, relatively few households report a relatively large area cleared for pasture[8], while in the Southeast Asian sites this regards clearance for tree plantations such as oil palm and rubber plantations[9] (figure 4.7d). Still, it is worth noting that especially the reported clearance for tree plantation is sensitive to the moment of survey data collected, as households reported not to clear for tree plantations regularly. For example, in the survey round of a few years later (not reported here), the amount of forest cleared for tree plantations in Indonesia-KCCP was significantly larger compared to the results of the first survey round as presented in figure 4.7.

4.3.4 REDD+ Interventions

Site level descriptions of interventions

Brazil-Transamazon

Three main interventions were applied, which all focused on local small-to-medium sized farmers: 1) direct cash payments conditional on forest conservation and fire-free agricultural production; 2) investments in alternative production; and 3) support for farmers to comply with environmental regulations. Most interventions thus featured change in land use strategies (land-saving strategies) and compensated direct forest protection. At the same time, federal command-and-control policies had significantly curbed deforestation – from all sectors and actors alike (Börner et al., 2014). Yet, ultimately the Brazilian Forest Code was also reformed in ways that particularly pardoned smallholder deforestation, thus loosening somewhat command-and-control leverages on smallholders (Cromberg et al., 2014a; Simonet et al., 2019).

Peru-Madre de Dios

At the Madre de Dios site, Bosques Amazonicos, FEPROCAMD and a local Peruvian NGO provided extensive technical support to Brazil nut producers to help them comply with national forest management regulations, specifically related to the formulation of annual operational and 5-year management plans for their concessions. However, the main planned interventions of the REDD+ project – namely implementation of a forest monitoring and surveillance system, construction of a local nut processing plant to increase the market value of harvested nuts, and eventual payments from the sale of carbon credits (Garrish et al., 2014) – never came through due to expiration of operational funds for the project in 2014.

Indonesia-KCCP

[8] In Brazil-Transamazon, 25 households (10% of respondents) together reported approximately 190 ha of clearance for pasture, which equals 19% of reported forest area cleared. In Peru-MDD, 7 households (3% of respondents) together reported approximately 74 ha of clearance for pasture, which equals 25% of total reported forest area cleared.

[9] In Indonesia-KCCP, 5 households (2% of respondents) together reported approximately 4 ha of clearance for tree plantations, which equals 5% of total reported forest area cleared. In Indonsia-Katingan, 12 households (5%) together reported approximately 15 ha of clearance for tree plantations, which equals 56% of total reported forest area cleared. In Vietnam-Cat Tien, 1 household (<0.5%) reported 3 ha of clearance for tree plantations, which equals 9% of total reported forest area cleared.

The objective of KCCP is to protect biodiversity and reduce greenhouse gasemissions from deforestation and forest degradation (Intarini et al., 2014). The project's main intervention is attaining a designation for specific forest areas in groups of villages as a Hutan Desa, or Village Forest (HD), forming a forest carbon pool. This overcomes the issue of economies of scale, related to monitoring and establishing community-based REDD+ projects (Intarini et al., 2014). The tenure-based intervention is done in combination with support for village boundary mapping, land use planning, and reforestation. At the same time, there were existing government reforestation program and a forest monitoring activity by a separate NGO. By attaining the HD status, the tenure of specific villages over communally-managed forest areas are clarified. This paves the way for getting management rights of the forest. By 2011, six villages[10] in Ketapang district, West Kalimantan, have proposed and attained HD status from the central government. During the same year, the Indonesian Ministry of Forestry initiated a national moratorium on the issuance of new permits for forest utilization and conversion on peatlands and primary forests, partially overlapping our study area in KCCP (Indonesian Ministry of Forestry, 2011). This moratorium became permanent in August 2019, covering 66 million hectares of rainforest (Diela, 2019). By 2014, none of the studied villages had attained the next necessary step of the HD status, which is to secure management rights from the provincial governor. An important element of this initiative is FFI's role as facilitator, which was crucial in bringing diverse communities together Intarini et al. (2014).

Indonesia-Katingan

The main interventions of this initiative are: (i) prevent large-scale deforestation by attaining an Ecosystem Restoration Concession (ERC) over a carbon-dense peat dome between the Katingan and Mentaya rivers; (ii) provide incentives for communities living in areas surrounding PT.RMU's ERC to support the prevention of DD through various alternative livelihood interventions agreed upon with communities; (iii) restore degraded peat forests through forest restoration activities; and (iv) establish fire-fighting teams in communities. The Indonesian government granted PT RMU their first ERC covering an area of 108,225 ha in 2013, and granted an additional, 49,497 ha in 2016 (Indriatmoko et al., 2014). Between 2010 and 2018, the project generated 23.3 million Verified Carbon Units (VCUs) equivalent to 23.3 million tons of greenhouse gas emissions removed (VCS, 2015). The project is active as of the time of writing (September 2019). These ERC areas overlap with 34 village territories and are located inland and therefore do not significantly overlap with areas actively managed by communities nearer to the rivers. Nevertheless, communities (especially village leaders) are important in deciding on whether outside players have access to the ERC area. During our 2018 study, all the study villages had areas within their village territory that are managed by private land users, such as palm oil and timber production companies. Also, all villages received significant levels of Village Funds (Dana Desa), totalling IDR 1 billion (USD 70,240) or more per village, used mainly for village infrastructure and public facilities and not for forestry-related purposes[11].

[10]Including the four GCS-REDD+ intervention villages

[11]Based on 2018 rates: USD 1= IDR 14,237 from https://data.worldbank.org/indicator/PA.NUS.FCRF?locations=ID.

4

Vietnam-Cat Tien

Interventions in Vietnam-Cat Tien were implemented by government agencies and NGOs, and mainly focussed on forest protection through trainings (on forest protection, REDD+ carbon credits, agroforestry), alternative livelihoods provisions (focussing on cacao and cashew) and participatory monitoring (participatory forest management) activities. Government agencies such as the National Park management and NGOs targeted their activities to communities living in the buffer zone of a national park. In one intervention, an NGO assisted district government agencies, focussing on REDD+ policy making.

Overview of interventions across all sites

Table 4.6 shows an overview of the forest interventions as discussion in the previous section. Together with the information on deforestation drivers (section 4.3.1- 4.3.3), table 4.6 was used for the identification of (mis)matches between interventions and drivers (section 4.4.3).

4.4 Discussion and conclusion

In this final section, we summarize our findings on DD drivers, evaluate the complementarity of data sources in DD driver identification, recognize potential (mis)matches between DD drivers and REDD+ interventions, reflect upon our study design and results and conclude with some final remarks.

4.4.1 Drivers

Land patterns following DD observed by high resolution imagery

We detected both across- and within-site variability of land patterns following DD. While agriculture is the dominant DD in the sites of Brazil (mainly pasture) and Vietnam (mainly crops), in Indonesia-KCCP tree plantations (oil palm) are most prevalent. Degradation is the main forest change in Peru-MDD (selective logging) and Indonesia-Katingan (near oil palm). This is in line with findings from earlier studies in those four countries (e.g. Soares-Filho et al., 2006; Asner et al., 2013a; Gaveau et al., 2018; Khuc et al., 2018). Most of our sites showed within-site spatial variability in land patterns (appendix C.3). In Peru-MDD mining was found only close to the main river in the south, pastures mainly close to roads, while selectively logged areas were also observed further away from roads and rivers. In Indonesia-KCCP mining was found in the south west, while other conversion types were found across the site. In Indonesia-Katingan large areas of oil palm plantations were found in the north east, while degraded forest due to fires were mainly found along the two main rivers. In Vietnam-Cat Tien, crops were found in the north, tree plantations in the west, and deforestation in the east was associated with the establishment of a large hydropower dam.

The minimum wage for Central Kalimantan in 2018 was USD 170/month (https://www.karirpad.com/blog/daftar-umr-umk-ump-2018-di-seluruh-indonesia/). Public facilities include indoor toilets, health centres/posts, village hall, and water ambulance.

4

	REDD+interventionsaggregated by category																	
	Disincentives			**Incentives**						**Enabling measures**						**Other**		
	Restrictions on forest access/conversion			Forest enhancement			Livelihood enhancement			Environmental education			Tenure clarification					
	Agent	Sector	Level	Agent	Sector	Level	Agent	Sector	Level	Agent	Sector	Level	Agent	Sector	Level	Agent	Sector	Level
Brazil-Transamazon	LH/SH	*	N/S/L				SH	AC/AP/F	L	SH	AC/AP/F	L	SH	AC/AP/F	N/S/L			
Peru-Madre de Dios	SH	F	N/L				SH	F	L	SH			SH	F	N/L			
Indonesia-KCCP	SH	F	L	SH	F	N	SH	F/AC/TP	L	SH	F	L	SH	F	L			
Indonesia-Katingan	SH	F	L/S	SH	TP	L	SH	TP/Fi/F	L/N	SH	F	L	LH	F	L			
Vietnam-CatTien							SH	AC	S/L	SH	F	L				G	O	S

Agent (targeted)
SH small/medium landholdersor communities
LH large land holders
G government

Sector (targeted)
AC agriculture -crops
AP agriculture –pasture
F forestconservation/NTFPs etc
Fi fishery
TP tree plantation
O Other
* all sectors

Implementation level
N National
S Subnational jurisdiction
L local

Table 4.6: Overview of intervention types per site

Spatially modelled variables explaining DD

Overall, spatial factors' relative importance did not differ considerably between the sites. Distance to existing agriculture and distance to roads were found to be in the top three important spatial factors for explaining and predicting DD, with Vietnam-Cat Tien being the exception where climatic variability was more important. Local variability in annual precipitation turned out to be important across all sites, which can also be derived from the density plots (appendix C.4), which show distinct differences in precipitation between the groups of stable forest and forest loss pixels. Yet, local variability in annual precipitation may be correlated with other variables both included in, and excluded from these models. In general, the topographical factors of slope and elevation were least important.

Locally reported DD drivers

In all sites both endogenous and exogenous causes of DD were reported, although their degrees differed greatly across the sites. In the Southeast Asian sites, commercial and governmental large-scale conversions were reported and mainly comprised of agro-industrial activities such as oil palm, pulp and food plantations. In the South American sites, a greater mixture of endogenous and exogenous causes prevailed, including agriculture by small holders, settlement of migrants, presence of logging companies, agro-industrial firms and mining. In absolute terms, more household level clearing was reported at the Brazilian site than other sites in the study, though large differences between households exist. In the Transamazon, in terms of both household count and forest area cleared, annual crops were cited mostly as clearance purpose.

4.4.2 Complementarity of different data sources in providing DD drivers information

Each data source and method used has its advantages and disadvantages. Their ability to assess certain driver elements is shown in table 4.7. Human interpretation of high resolution remotely sensed imagery provides insights into the activities associated with different conversion types. When a proper sampling design is applied, a sample-based approach like ours allows for estimating the relative share of different conversion types. Increasing the number of DD classes may lead to more informative results but requires increased numbers of samples (Foody, 2009). Further, although going beyond the scope of this research, temporal changes in DD processes can be revealed when the same samples are assessed repeatedly over time. Mapping the different conversion type findings of the samples reveals within-site spatial patterns (appendix C.3).

Spatial modelling and random forest models in particular can reveal the relative importance of preselected underlying factors, while they can deal with non-linear relationships between prediction variables. Yet, the relationships between the prediction variable and spatial factors that turn out to be important may not be easily interpretable. While not part of this particular study, the spatially explicit prediction maps allow for identification of areas at immediate deforestation risk.

Table 4.7: Complementarity of methods and datasets

	Agent *who*	Activity *what*	Location *where*	Time *when*	Underlying factors *why*	Size (of activity) *how much*
High resolution imagery	×	✓	✓	✓	×	✓
Random Forest model	×	×	~	×	✓	~
Socio-economic surveys	✓	✓	~	~	~	×

✓ yes
~ to some extent
× no

Village and household level surveys further complement the previous methods, as they can provide insights in to the agents of specific DD activities. Further, local stakeholders can often help to identify the underlying factors at play. The spatial and temporal information about DD activities are often limited compared to remotely observed methods, but participatory mapping and recurrent surveys can be of added and unique value when combined with the spatial information on DD.

4.4.3 Alignment of DD drivers and REDD+ interventions

Site specific findings

In Brazil–Transamazon, local interventions generally focus on local small-to-medium sized farmers by promoting sustainable farming practices (incentives), while federal forest restrictive policies (disincentives) do not distinguish between agents and sectors. Both local interventions and federal restrictive policies thus seem to be aligned with the agriculture related DD drivers. Yet, the national policy partially pardons small-scale deforestation, thus somewhat contrasting federal policies.

In Peru-MDD, we found clear within-site spatial differences in DD drivers, which calls for a locally tailored approach. The REDD+ initiative focusses on Brazil nut concession owners north of the river, thus not targeting the large-scale mining near the main river. Further research is needed to verify whether other interventions target mining agents specifically or if indeed this driver is currently not addressed sufficiently. The REDD+ initiative indirectly addressed small-scale logging by adding value to Brazil nut concessions via increased prices for producers. Yet, limited logging under forestry regulation in Peru is allowed (Garrish et al., 2014).

In Indonesia-KCCP, the initiative's focus on tenure clarification is aimed as an empowerment tool for local communities, in order to keep exogenous agents out. In that sense, these interventions are in line with the exogenous threats coming from large scale palm oil companies. Mining was found to be a considerable, but very localised driver present in the south west of the area. This again calls for a locally tailored approach of REDD+ interventions, as mining was not addressed specifically by any of the interventions in this study.

Part of the initiative's focus in Indonesia-Katingan is fire prevention, to correspondingly reduce the impact of fires and thus prevent forest degradation. This is in line with the major threat we found in the area. Exogenous agents such as palm oil companies play an increasing role in the area's forest change activities, and is correspondingly putting a pressure on local communities. These exogenous drivers are not targeted directly by the interventions.

In Vietnam-Cat Tien, mostly secondary forests are being converted to agriculture and plantations (mainly orchards and cashew plantations). Interventions focused primarily on environmental education and stimulating sustainable livelihood practices through the provision of livelihood enhancements. Yet, reported clearance at household level was minimal, so conversions by other actors may have been addressed insufficiently.

General remarks regarding driver-intervention alignment

While national or subnational policy interventions across the sites mostly comprise of regulations to restrict forest access, the local initiatives often comprise of a mixture of interventions. These 'baskets' are regularly targeted towards individual smallholders or communities, and are often continuations of existing integrated development and conservation projects and programmes, as described in earlier studies based on the GCS-REDD+ data (Sunderlin et al., 2014a; Duchelle et al., 2017).

The village and household level survey data showed that exogenous agents played an important role in DD and high-resolution imagery revealed most conversion activities to be large-scale, while most REDD+ interventions mainly targeted local communities and smallholders. Still, incentives to smallholders are of value as they can compensate for disincentives affecting smallholders and large landholders alike.

Müller et al. (2013) argue that proper driver-intervention alignment does not necessarily mean that REDD+ should prioritise its activities on the largest driver at play (mechanized agriculture in that case), but rather should take into account its opportunity costs. In addition, one might argue that for effective and efficient REDD+ alone, driver-intervention alignment is not essential as, at least in theory, DD can be addressed by forest use restrictions combined with effective law enforcement. However, this is most likely not equitable in the sense that people who depend on the forest the most for their livelihoods, will most likely be disproportionally affected by restrictive interventions. Reduced DD may then lead to trade-offs in well-being and forest-related income (chapter 5). As Godar et al. (2014, p. 15595) acknowledge: "Beyond the technical difficulties and increased costs, efforts to curb deforestation in areas dominated by smallholders are politically and socially problematic because many smallholders depend on clearing small areas of forest for their livelihoods and subsistence". Therefore, understanding driver-intervention alignment is at the least essential to understand who is most likely to lose out from curtailing deforestation, and where trade-offs between carbon and well-being outcomes can be expected.

4.4.4 Study reflections

Dynamics of drivers over time

In addition to spatial variability in drivers, drivers can change over time, as a result of interventions or due to other processes. As Godar et al. (2014) argue, the changing (relative) contributions of specific actors to deforestation and degradation need to be examined in order to achieve further reductions in DD. These dynamics should be studied in more detail, and be taken duly into account when designing, implementing or evaluating REDD+ interventions.

In our study, the timeframes from our remote sensing and spatial modelling assessment (2001-2014/2015) differ from the timeframe addressed in the socio-economic surveys (conducted in 2010-2011, with reported forest clearing regarding the two years prior to the surveys). Remote sensing requires longer timeframes to detect follow-up land use and other DD patterns, which would complicate year-to-year comparison between remotely sensed patterns and reported drivers. We argue that for the purpose of method complementarity assessment, however, year-to-year alignment is of lesser importance, as the different data sources and corresponding methods focus on different driver elements.

Discrepancies in deforestation magnitudes and deforestation drivers

Although assessing their complementarity was the main reason for using multiple data sources, the results contain some, at least seemingly, discrepant findings regarding deforestation estimates and direct drivers categories.

We only report relative shares (in percentages) of forest change patterns observed by remote sensing, as the area of interests of the remote sensing analysis are based on rectangular buffers around the REDD+ initiative areas and therefore comprise most likely of more than the study villages' area of influence. In the absence of spatially explicit household areas, direct comparison of deforestation numbers in absolute terms would therefore be impossible. It is possible that household level clearance was under-, or over-reported, although multiple verification questions in the household survey limited this chance considerably.

In section 4.2.4 we already acknowledged that the follow-up land use after deforestation is not always the main driver of deforestation. Findings on 'drivers' from high resolution imagery can therefore seemingly contradict the findings from village and household surveys. In addition to the reasons addressed in the previous section, this would call for a longitudinal study on local land use patterns, in which corresponding DD drivers and changes therein would be repeatedly assessed.

Study limitations and further research

We acknowledge that in this study, we have put limited focus on the underlying forces influencing agents' land use decisions. Here, we limited ourselves to aspects of land tenure, while other potential underlying forces including commodity prices were largely ignored. We do argue however that REDD+ interventions may have limited influence on these (global) market prices, whereas strengthening land rights is at the core of many interventions as shown in this study.

In the drivers assessment part of this study, our main focus was to examine the complementarity of different data sources in addressing different driver elements. We therefore simplified the study design for each of the three methods. This means that especially in the spatial modelling part further research is needed. Among other things, future studies could experiment with feeding the RF with more or other spatial factors that potentially explain or relate to DD, such as distance to cities and markets, distance to palm oil mills, and other microclimate factors. In that way, the RF could further enhance the understanding of the relative importance of different spatial factors determining DD, and to further increase the accuracies of the prediction models, so as to identify future deforestation risk areas.

4.4.5 Concluding remarks

DD activities are the result of a complex interplay of agents, underlying forces and the environment. Our study showed that DD patterns differ across and within sites. This calls for a locally tailored approach when designing and implementing REDD+ interventions. We show that no single dataset or method can reveal all facets (who, what, where, why, when and how much) of DD drivers, while a combined assessment leads to a better understanding.

Despite the differences between sites, some general lessons can be drawn. The remote sensing analysis on DD classes showed that in most sites the predominant activity was large-scale agriculture or large-scale tree plantations. Household survey results showed that household-level forest clearance was mainly for annual crops. A basket of REDD+ interventions were applied in the study areas aiming to prevent forest conversions. Our results show that the local interventions mainly targeted households and small-scale processes, in contrast with the remote sensing findings that drivers were mostly large-scale.

In this interdisciplinary study, we have provided insights into the complexity of DD driver identification and complementarity of different driver related data sources at the local scale. Further, we have assessed the alignment of these identified drivers and REDD+ interventions. A better understanding of the alignment between DD drivers and REDD+ interventions is vital for practitioners and policy makers to enhance the effectiveness, efficiency, equity and co-benefits of REDD+ at the local level.

Acknowledgements

All respondents in the study sites are gratefully acknowledged, as well as the CIFOR staff who did the data collection in the field. The authors thank Mella Komalasari for her guidance with the M2 datasets, and Juultje Blom for her work on the Random Forest models.

5 Exploring changes in forest cover and well-being in the context of forest interventions

This chapter is based on:

Duchelle, A.E., Bos, A. B., de Sassi, C., De Sy, V., and Herold, M. Exploring changes in forest cover and well-being in the context of forest interventions *(to be submitted)*

Abstract

Although the aim of forest-based climate change mitigation interventions, such as REDD+, is to protect and enhance forests, there are legal, moral and practical reasons for making sure that this objective is achieved while at a minimum not harming, and ideally ensuring benefits for, local people. Here, we performed an exploratory analysis to examine the relationships between the treatment intensity of different types of forest interventions, changes in forest cover (loss) and changes in select measures of well-being (income and perceived well-being) in seventeen subnational sites across the tropics. Information on interventions, household income and perceived well-being was gathered from village and household level interviews from nearly 130 villages and 4,000 households. Global Forest Change data (2000-2018) was used to derive information on forest cover and forest cover change at the village level. We defined clusters of villages based on similar levels of intervention treatment intensities and deforestation trends to compare pre- and post-intervention characteristics.

We found that villages in the cluster with high treatment intensities and reduced post-intervention deforestation rates consisted mostly of Brazilian villages. These villages had higher income levels and deforestation rates in the pre-intervention period. In the post-intervention period, these villages were generally associated with an increase in income and its households reported a slightly better level of perceived well-being. In this analysis, we did not find clear differences in outcomes among different intervention types although restrictions in forest access and conversions were considered the least positive interventions in terms of effect on perceived well-being. We did not find indications of pronounced trade-offs between forest conservation and well-being outcomes at all villages and households as a whole, although trade-offs at specific villages and households could not be ruled out. This analysis provides one way of looking at the forest change and select well-being outcomes of different forest interventions, and their possible trade-offs.

5.1 Introduction

5.1.1 Background

The world's tropical forests hold important potential for mitigating climate change and contributing to development objectives. Halting deforestation, along with other 'natural climate solutions' such as restoring degraded lands and sustainably managing forests, could provide at least 37% of the cost-effective emissions mitigation needed by 2030 to keep global warming below 2°C (Griscom et al., 2017). Additionally natural forests and wildlands provide approximately 28% of total household income for communities in and around tropical forests in terms of food, fuelwood and fibre for consumption and sale, which is almost as much as income derived from agricultural crops (Angelsen et al., 2014).

When the REDD+ concept emerged ten years ago, it was considered a way to not only mitigate climate change, but also conserve biologically diverse forests, improve forest governance, and enhance the livelihoods of forest-dependent people (Brown et al., 2008). The focus on these so-called "non-carbon benefits" was enshrined in the REDD+ social and environmental safeguards negotiated under the UNFCCC.

Since its inception, REDD+ initiatives have grown at national, subnational and local scales (Duchelle et al., 2018b). Although a key characteristic of the UNFCCC REDD+ framework is its focus on the national level (or subnational in the interim), hundreds of local REDD+ projects were launched after the UN climate negotiations in 2007 called for "demonstration activities." These projects have been mostly implemented by NGOs or for-profit companies with an orientation toward voluntary carbon markets and a focus on smallholders (Simonet et al., 2015; Sills et al., 2014). Implementers of these initiatives apply intervention packages that in customized ways combine enabling measures (e.g. free, prior, informed consent (FPIC) and land tenure clarification), disincentives (e.g. restrictions on forest access or conversion), and incentives (conditional or non-conditional) with the aim to achieve better protection of forests (Sunderlin et al., 2015; Duchelle et al., 2017).

In REDD+ and other conservation interventions, there is a broad consensus that these conservation and climate change activities should be 'evidence-based'. Yet, approaches that enable impact evaluation through quasi-experimental approaches are hardly applied in conservation sciences (Sills et al., 2017). Although local REDD+ projects are fundamentally different than national- and subnational-level programs, they represent the natural laboratory for evaluation of interventions on the ground that are needed to inform the design and implementation of higher-level policies aimed at conserving forests while simultaneously promoting local well-being.

5.1.2 REDD+ project theory of change

Given the variety of possible impacts of REDD+ initiatives on local forests and people, it is important to understand what implementers set out to achieve, and through which types of interventions. Figure 5.1 illustrates the possible intended outputs, outcomes and impacts of the different types of REDD+ interventions. These elements are described in more detail below.

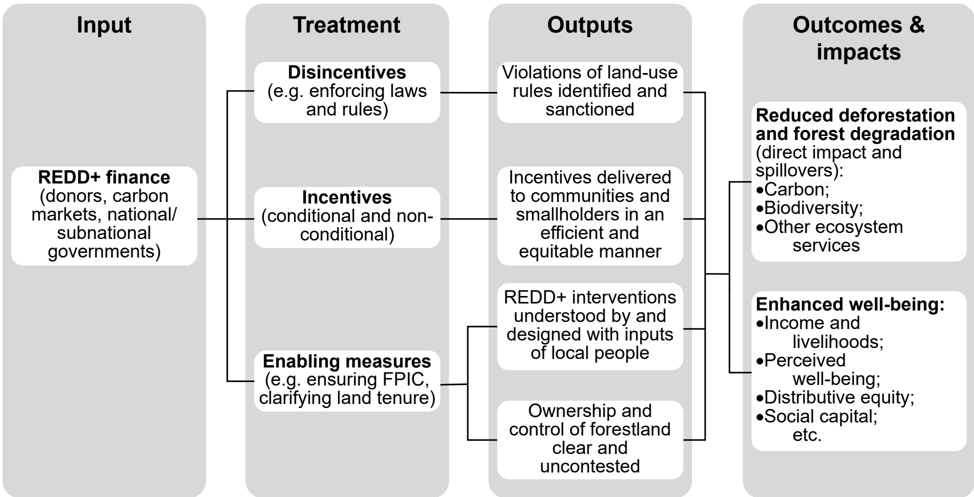


Figure 5.1: Theory of change for local REDD+ initiatives. Adapted from Duchelle et al. (2018a)

The primary input to local REDD+ initiatives is finance from donors, voluntary carbon markets, and governments. Donor-driven funding for REDD+ has been limited to a handful of donor countries and multilateral institutions and has thus far been insufficient to support the mitigation potential of tropical forests (Norman and Nakhooda, 2015; Atmadja et al., 2018). In terms of local REDD+ initiatives, fewer implementers have sold carbon credits than what was initially envisioned as the primary way in which REDD+ would be financed (Sunderlin et al., 2015). Implementing agencies have instead also relied on donor funding, which allows them to apply a package of land-use interventions to villages and individual households in the intervention areas. Subnational governments have also heavily subsidized local REDD+ initiatives (Luttrell et al., 2018).

The ‘treatment’ through REDD+ interventions is diverse. Enabling measures, such as FPIC, land tenure clarification and environmental education, help set the stage for responsible land stewardship, e.g. in terms of adequate information and assignments of rights. Disincentives include regulation and enforcement of restrictions in access to, and conversion of, forests. In theory, violations of forest and land-use rules should be identified and sanctioned through effective monitoring and enforcement by village associations and governmental agencies, and thus protect forests. Conditional incentives (e.g. PES) require participants to protect or improve local forests in exchange for benefits, unlike non-conditional livelihood support in which forest conservation and restoration activities are promoted by investing in productive alternatives (e.g. more sustainable agricultural practices) without consequences for non-adherence.

There are a variety of outputs associated with the different REDD+ interventions. For example, land tenure clarification activities should lead to clear and uncontested ownership of forestlands. Environmental education should lead to local awareness of REDD+ interventions, as well as to a better understanding of environmental legislation and the benefits of sustainable land use practices where relevant. Conditional and/or non-conditional livelihood enhancements should be delivered in an equitable manner (i.e. a substantial proportion of households should receive incentives and the distribution of incentives should reflect the foregone economic opportunities by all participants). Through effective monitoring and enforcement by village associations and governmental agencies, violations of forest and land-use rules should be identified and sanctioned.

Ultimately, the main desired outcomes of REDD+ interventions are: 1) reduced carbon emissions and biodiversity loss from deforestation and forest degradation in the intervention area; and 2) enhanced well-being of local land users in the intervention area. While there are many possible frameworks for conceptualizing and measuring well-being, the common impacts assessed in the recent REDD+ literature are income or livelihoods, project costs, perceived well-being, distributive equity, and social capital (Duchelle et al., 2018c). Beyond these, REDD+ interventions can also promote individual well-being through higher level impacts on increased land tenure security (Sunderlin et al., 2014c), local capacities, institutions and networks.

Since an important focus of REDD+ is to restrict deforestation and forest degradation activities, likely trade-offs between conservation and livelihood benefits must be examined. Local people are likely to incur opportunity costs (Rakatama et al., 2017), but also may benefit from forest conservation activities, especially when the damage is caused by external actors (Clements et al., 2014). In addition, REDD+ implementers may intentionally overcompensate local opportunity costs, which are difficult to quantify, to ensure that participating communities experience some net welfare gains (Ickowitz et al., 2017). These benefits may take time to materialize, however, as new activities start to pay off. One challenge is that certain groups often bear the costs of forest conservation, i.e. the poorest may be the most dependent on clearing forest for subsistence uses and adversely affected by conservation restrictions (Poudyal et al., 2018). There is also increasing evidence of elite capture in REDD+ benefit sharing schemes (Poudyal et al., 2016). At the same time, wealthier households often glean more absolute benefits from forests, meaning they would need higher compensation for foregone forest uses than poorer households (Ickowitz et al., 2017). To deliver maximum well-being benefits, incentives should be to a substantial proportion of households – and not just the village elites –, and local perceptions of equity (i.e. perceived fairness of benefits) should be taken into account (Loft et al., 2017).

In this context, we performed an exploratory analysis to examine the relationships between the treatment intensity of different types of forest interventions, changes in forest cover (loss) and changes in income and well-being at seventeen local REDD+ initiative sites in Brazil, Peru, Cameroon, Tanzania, Indonesia and Vietnam (Sills et al., 2017). Note that for this analysis we are looking at a broader set of forest interventions then those that are specifically labelled as a REDD+ interventions. Although the datasets used (described in the next sections) include a series of outcome variables broadly related to well-being, such as income (including cash and subsistence components), asset indices and key selected assets, perceived well-being, and perceived tenure security, among others, we chose to focus one objective measure of well-being (i.e. income) and one subjective measure (i.e. perceived well-being) for this exploratory analysis. We see this exploration as an important step towards the fourth objective of this dissertation, namely to begin to understand the relationships between forest change outcomes and select well-being outcomes of different REDD+ interventions. Based on lessons learned from this exploratory analysis, we provide recommendations for further steps toward a more rigorous impact and trade-off assessment of forest and well-being outcomes of REDD+ interventions. In summary, the objectives of this analysis are to (1) explore the relationship between the treatment intensity of different types of forest interventions and changes in forest cover; and (2) explore the relationship between the treatment intensity of different types of forest interventions and changes in household income and perceived well-being.

5.2 Material and methods

5.2.1 Study sites

For this analysis, we focused on seventeen local REDD+ sites in Brazil, Peru, Cameroon, Tanzania, Indonesia and Vietnam as part of CIFOR's GCS-REDD+ (figure 5.2). GCS-REDD+ is characterized by a quasi-experimental research approach through which socio-economic surveys were applied in nearly 130 villages and 4,000 households in 2010-2011 (pre-intervention) and 2013-2014 (post-intervention). See Sunderlin et al. (2016) and Sills et al. (2017) for a detailed description of the methods used for site selection, matching of intervention and control villages, and random sampling of households to be interviewed. For this particular analysis, we combined the results of these surveys with an analysis of Global Forest Change (GFC) data (2000-2018) to assess how different interventions are related to changes in forest cover, income, and perceived well-being at the village level. We describe the operationalization of and data collection associated with each of the three outcome variables in sections 5.2.4 (forest cover loss), 5.2.6 (changes in household income) and 5.2.7 (perceived changes in household well-being) below.

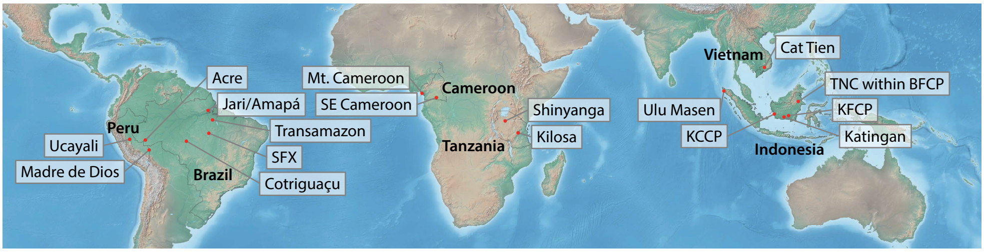


Figure 5.2: REDD+ initiatives analyzed through CIFOR's Global Comparative Study on REDD+

5.2.2 Categorization of village and household interventions

In the post-intervention period (2013-2014), we applied a survey to document all interventions in the study villages that intentionally aimed to protect or restore forests. For this survey, we first compiled a list of the main forest interventions applied at each site from project documents and the first research phase. Interventions included those that were applied by the REDD+ implementers themselves, as well as higher-level policy interventions that directly affected local forests and landholders (e.g. Brazilian command-and-control policy actions). A list of the top ten interventions per village was verified by the REDD+ implementer and village-level key informants before fieldwork began. In the household interviews during the post-intervention phase, we asked if the household had been involved in each intervention, and if so, how it affected their land use and perceived well-being.

For disincentives, we focused on regulation, monitoring or enforcement activities that restricted forest access or conversion (RFAC). At our study sites, these interventions included monitoring and enforcement of forest protection laws and regulations; imposition of fines; restrictions on local forest clearing and burning; and community monitoring to clarify boundaries and establish land use plans (Duchelle et al., 2017).

Incentives applied included conditional and non-conditional livelihood enhancements, as well as forest enhancements. Conditional transfers (e.g. PES) were comprised of direct cash payments and funds for sustainable production and infrastructure; they required participants to protect or improve local forests in exchange for getting this support. Non-conditional livelihood enhancements included support for a diversity of sustainable forestry, agriculture, and energy activities. Forest enhancement activities were primarily support for restoration of degraded lands with agroforestry systems, and enrichment planting with valuable timber and non-timber forest products (Duchelle et al., 2017).

For enabling measures, we focused on tenure clarification and environmental education activities. Tenure clarification (TC) activities aimed at resolving unclear or contested ownership and access rights over local forestlands, trees and carbon. At our study sites, TC activities included participatory forest mapping, land and resource conflict resolution, and regularization and change of tenure classification (Sunderlin et al., 2018). Environmental education (EE) activities were information sharing, including via FPIC processes, along with outreach and extension aimed at clarifying relevant environmental legislation and demonstrating the tangible benefits associated with protecting and/or enhancing local forests.

5.2.3 Treatment intensity of interventions

For this analysis, our measure of treatment intensity (TI) was based on a combination of households reached and households impacted by the different intervention types (ITs) described above. It bases on the rationale that the influence of an intervention does not solely depend on the amount of people targeted by an intervention alone (i.e. the *reach*), but also depends on the degree to which this intervention actually changed the behavior of the people targeted (i.e. the *impact*). The treatment intensity score per intervention type and per village was calculated as follows:

$$\begin{aligned} TreatmentIntensity_{IT,v} &= Reach_{IT,v} \times Impact_{IT,v} \\ &= \frac{\sum_{i=1}^{n_v} IT_i}{n_v} \times \frac{\sum_{j=1}^{m} LU_j}{\sum_{i=1}^{n_v} IT_i} \\ &= \frac{\sum_{j=1}^{m} LU_j}{n_v} \end{aligned} \tag{5.1}$$

Correspondingly, the total treatment intensity score per village was calculated as follows:

$$TotalTreatmentIntensity_v = \sum_{k=1}^{t} \left(\frac{\sum_{j=1}^{m} LU_j}{n_v} \right)_k \tag{5.2}$$

Where $IT_i \in [0, ..., \infty]$ refers to the exposure of a household to each of the interventions per intervention type. This is generally 1|0, but can be >1 in cases where a household is involved in multiple interventions of the same type. n_v is the number of household respondents per village. $LU_j \in [0, 1]$ refers to either a reported land use change behaviour (1) or not (0). $\sum_{i=1}^{n_v} IT_i$ is the sum of instances where households are exposed to one or more intervention of the same type. m is the number of instances per village where households reported to be involved in an intervention type. $\sum_{j=1}^{m} LU_j$ is the sum of instances where households reported a change in land use behaviour. t is the number of different intervention types, which is six in our study.

Figure 5.3 shows the average treatment intensity per country as derived from $TotalTreatmentIntensity_v$. Treatment intensity in Brazil and Peru is mainly defined by restrictions on forest access and conversion, whereas Indonesia is characterized mostly by non-conditional livelihood enhancements and forest enhancements.

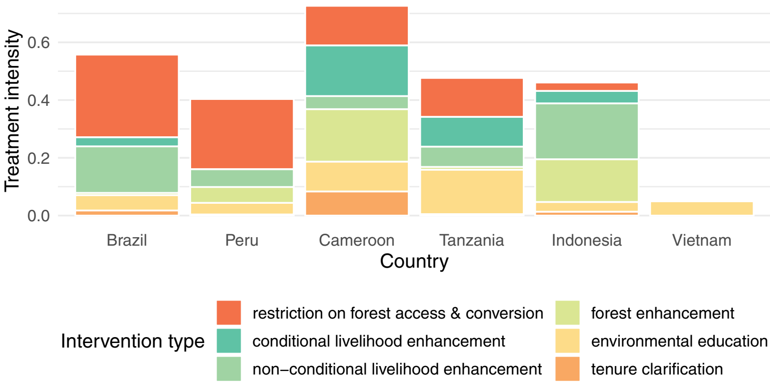


Figure 5.3: Average treatment intensity in all study villages per country

5.2.4 Change in forest cover loss as proxy for REDD+ effectiveness

We assessed annual forest cover loss in both the pre- and post-intervention period as well as the difference in average annual deforestation rates between those periods as proxy for REDD+ effectiveness.

We used Global Forest Change (GFC) data (version 1.6), which includes both a tree cover density in 2000 layer (*treecover2000*) and information on tree cover loss for 2001-2018 (*lossyear*) (Hansen et al., 2013). GFC provides data on global tree cover loss annually at a spatial resolution of 30m. Following FAO (2000), we defined forests as areas with >10% tree cover. The *treecover2000* layer was used to create forest masks using the 10% threshold. Accordingly, forest loss[1] was then defined as a conversion from >10% tree cover in 2000 to 0% in any of the years 2001-2018. We calculated the areas of forest loss (pixels) and average annual forest cover loss (%) relative to initial forest cover in 2000. See Bos et al. (2017) for an elaboration on the processing steps using the GFC data. Spatially explicit boundaries of the villages were based on government data; REDD+ implementer data; grid corner points as collected by field researchers using GPS devices; or a buffering of household GPS points.

Next, as proxy for the effect of the REDD+ interventions on the deforestation trends, we compared the average annual deforestation rates before and after the start year of the REDD+ initiative. To this end, we calculated the Before-After (BA) score for each of the villages, following Bos et al. (2017):

[1]We refer to *tree* cover loss data when discussing the raw input data (Hansen et al., 2013). Tree cover refers to the within-pixel coverage of trees. Once a forest mask is applied, based the FAO's definition of forests, and we discuss the tree cover loss in an aggregated area (e.g. a village area), we refer to *forest* cover loss.

$$\begin{gathered} BA\ score\ \alpha = \bar{x}_{after} - \bar{x}_{before} \\ with\ \bar{x}_{after} = \frac{1}{n_{after}} \sum_{i=1}^{n_{after}} x_i\ and\ \bar{x}_{before} = \frac{1}{n_{before}} \sum_{i=1}^{n_{before}} x_i \end{gathered} \tag{5.3}$$

Where $\bar{x}_{after}$ represents the average annual deforestation rate in the village area in the period since the intervention started (i.e. post-intervention), as a percentage of the total forest area in 2000; $\bar{x}_{before}$ represents the average annual deforestation rate in the intervention area in the period from the start year of measurement (here: 2001) up until the intervention started, n_{after} and n_{before} the number of years in respectively the after and before period. A negative α would thus signify a reduction in average annual deforestation, and thus potentially signify a positive REDD+ effect in terms of avoided deforestation.

5.2.5 Cluster building

As can be derived from figure 5.4, exploratory analysis showed no clear relationship between treatment intensity and REDD+ effectiveness, meaning that the expected negative relationship between BA score (i.e. reduced deforestation rates) and TI (i.e. high degree of interventions that changed behaviour) was not consistent across all study villages. To understand why some villages with high treatment intensities experienced reduced deforestation rates, while others were characterized by increased rates of deforestation, we used the TI and BA scores to define five clusters of villages with relatively homogeneous TI and BA scores. These homogeneously distributed clusters allow us to examine the contextual conditions and intervention types that might explain differences in (non)performance under similar levels of intervention application. In addition, it provides a framework to explore post-intervention changes in income and perceived well-being.

For the BA score classification, we used the thresholds from Bos et al. (2017) to differentiate between *good, neutral* and *poor* BA scores, representing decreased, stable and increased deforestation rates respectively. The TI class cut-off points were based on the median of all village-level TI scores, meaning that class *none* represents TI scores of 0, class *low* contains 0<TI<median(TI), and class high contains villages with TI>= median(TI). All possible combinations of the BA and TI classes were then aggregated into five clusters (table 5.1, figure 5.4.

Next, we used these clusters as the basis for analysis on: (1) pre-intervention characteristics comparisons (results in section 5.3.1), (2) intervention type composition comparisons (results in section 5.3.2), and (3) the effects of REDD+ interventions on household income and perceived well-being (results in section 5.3.3).

Table 5.1: Cluster definitions

BA class	TI class	Cluster	Description *(and underlying question)*
Good	High	A	High reduction in deforestation rates, with high treatment intensity *(as expected, but what made it work?)*
Good	Low/None	B	High reduction in deforestation rates, but with no or low treatment intensity *(why reduced forest loss, despite absence of interventions?)*
Neutral/Poor	High	C	Stable or increasing deforestation rates, but with high treatment intensity *(why increased forest loss, despite the interventions?)*
Neutral/Poor	Low	D	Stable or increasing deforestation rates, with no or low treatment intensity *(as expected, but what are the group characteristics?)*
Neutral/Poor	None	E	Increasing deforestation rates, without interventions *(as expected, but what are the group characteristics?)*

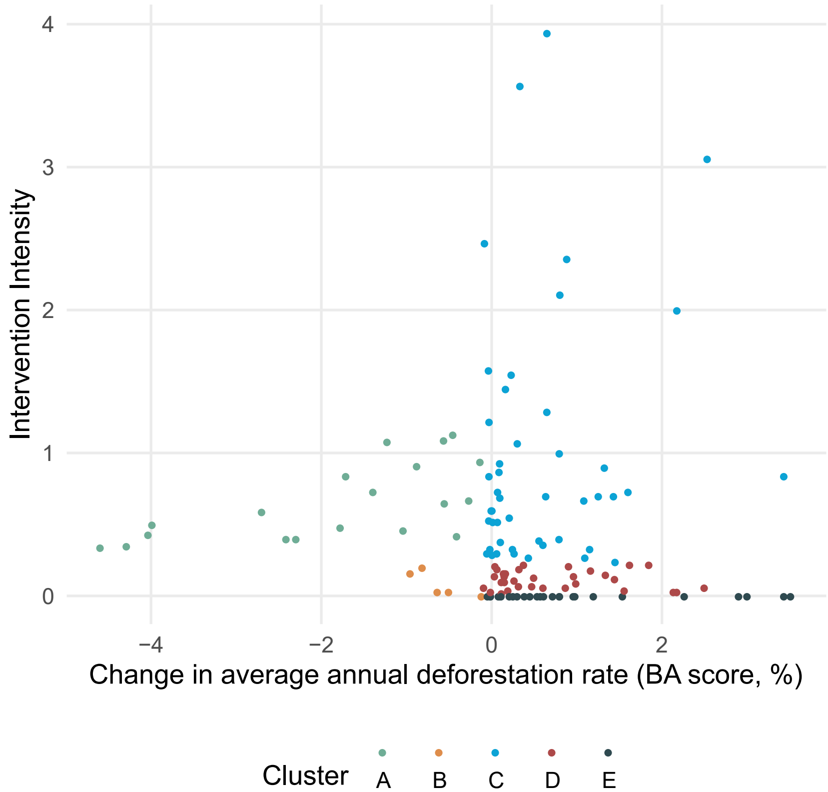


Figure 5.4: Clusters of villages based on change in average annual deforestation rate (BA score) and treatment intensity of interventions. A negative BA score represents a reduction in average annual deforestation rate.

5.2.6 Changes in household income

To understand the effects of REDD+ on household income, detailed income data (all cash and subsistence sources, following Angelsen et al. 2014) were collected for over 4,000 households in 130 villages in the 2010-2011 (pre-intervention) and 2013-2014 (post-intervention) periods, using a BACI study design and calculated as income per adult equivalent. Treated and control villages were reasonably well balanced at baseline (Sills et al., 2017), but we used matching on several household and village characteristics (de Sassi et al., in prep) to maximize accuracy in the comparison of intervention against control groups. For this particular analysis, we did not differentiate between the dichotomous classes of control and intervention, but rather focused on the degree of treatment intensity per village. Since forest interventions were applied by non-REDD+ implementers in control areas, villages assigned to the 'control' class in GCS-REDD+ can have a treatment intensity of >0. For our pre-intervention measure of household income, we use the data from 2010-2011. For post-intervention *change* in income, we calculated both absolute and relative difference between the 2010-2011 and 2013-2014 periods of data collection at the village level.

5.2.7 Perceived changes in household well-being

To complement our analysis of changes in household income, we also analyzed subjective measures of well-being. The Organisation for Economic Co-operation and Development (OECD) defines assessments of subjective well-being as "...measures of how people experience their life as a whole" (OECD, 2013, p. 10). Such measures of well-being have been increasingly adopted in national and international levels and gained credibility through the demonstrated correlation between these subjective measures and objective ones, such as income (Sunderlin et al., 2017; Krueger and Schkade, 2008). Following the methods outlined in Sunderlin et al. (2017), we used the data on households' perceptions of change in well-being (*better off now*, *about the same*, *worse off now*) in the two years prior to the post-intervention survey, and we assessed differences between clusters. In addition, and as described earlier, for all interventions in which households were involved, we asked them to evaluate the effect of each specific intervention on their household's well-being. Responses were categorized as *very negative*, *negative*, *mixed or no effects*, *positive*, or *very positive*. For both sets of questions on well-being, we did not impose a rigid definition of well-being to allow respondents to refer to their own conceptions of well-being (Sunderlin et al., 2016).

5

5.3 Results and discussion

In this section, we first show the differences and similarities in certain characteristics between the clusters in the pre-intervention period to assess whether theses clusters were already inherently different before interventions were applied (5.3.1). In this way, we aim to get a better understanding of the contextual conditions that might explain differences in (non)performance under similar levels of REDD+ intervention application. Next, for the two clusters with high TI (i.e. A and C), we show whether the composition of different intervention types differed between the two clusters (5.3.2). Then, we show the post-intervention changes in income and perceived well-being for the different clusters (5.3.3). We close with some study reflections (5.3.4).

5.3.1 Pre-intervention cluster characteristics

The main characteristics of the different clusters before interventions were applied are shown in figure 5.5.

Cluster A, which represents high TI and a decrease in average annual deforestation rates, is mainly observed in study villages in Brazil[2] (figure 5.5a). In fact, 46% of all Brazilian villages in our sample were part of cluster A[3]. Conversely, cluster E, which represents no TI and neutral or poor BA (E), is predominantly found in the Indonesian and Tanzanian study villages[4]. This cluster alone contained 39% of the Indonesian and 43% of the Tanzanian villages. The other clusters show a rather mixed country composition. Since some of the clusters appeared to be dominated by single countries, we tested whether differences between clusters were mainly country driven or not. The detailed results can be found in appendix D.1. Although we found some obvious differences between countries including, but not limited to, income (change) and initial forest cover, grouping villages by country did not result in distinct clusters of treatment intensities and change in average annual deforestation rates (figures D.1.1, D.1.2, D.1.3). Thus, country-origin alone is not the explanatory factor for reductions or increases in deforestation.

We applied Kruskal-Wallis Rank Sum tests (Hollander and Wolfe, 1973) and unpaired Wilcoxon Rank Sum tests using the *Stats* package in R to assess whether the clusters differed at baseline in terms of pre-intervention income, forest cover in 2000 and pre-intervention deforestation rates[5]. Detailed results can be found in appendix D.2.

For pre-intervention income (figure 5.5b), five cluster pairs turned out to be significantly different from each other, that is A-C, A-D, A-E, C-E and D-E. Overall, villages in cluster A showed higher initial incomes compared to most other clusters, and villages in cluster E showed a lower initial incomes compared to most other clusters. We presume that this was largely due to the differences in country composition between the clusters, rather than due to other cluster characteristics.

[2]Note that the number of initiatives and correspondingly the number of villages differs between countries, resulting in larger numbers of villages for Indonesia and Brazil, and fewer villages in Vietnam

[3]Nine out of these seventeen villages were classified as *intervention* in the original research design. The remaining six villages were *control* villages.

[4]The vast majority of the villages in cluster E (24 out of 26) were classified as *control* villages in the original research design. The remaining two villages were *intervention* villages.

[5]H_0 location parameters of the distribution of each variable are the same in each cluster. The alternative (H_A) is that they differ in at least one cluster.

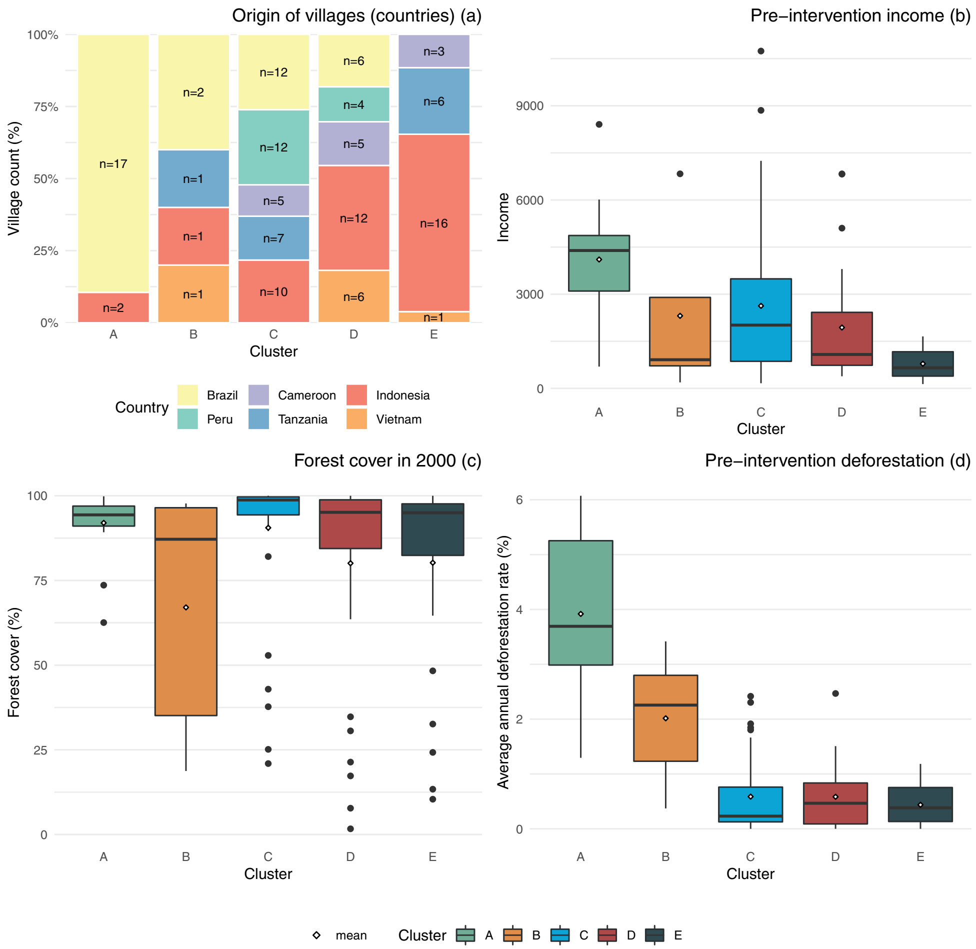


Figure 5.5: Pre-intervention cluster characteristics, showing (a) country composition, (b) pre-intervention income, (c) forest cover in 2000 and (d) pre-intervention deforestation rates. Forest cover in 2000 was defined as percentage of pixels with >10% tree cover in 2000 relative to the total village area. Upper and lower extremes of whiskers represent $Q3 + 1.5*$interquartile range (IQR) and $Q1 - 1.5 * IQR$ respectively, where $IQR = Q3 - Q1$.

For forest cover in 2000 (figure 5.5c), the p-value of the Kruskal-Wallis test was <0.05, while none of the one-to-one cluster pairs appeared to be significantly different. We thus conclude that a village's pre-intervention forest cover degree did not determine a good or poor BA score nor levels of REDD+ treatment intensity.

In contrast, for pre-intervention deforestation rates (figure 5.5d), we found that six cluster pairs differed significantly: A-C, A-D, A-E, B-C, B-D and B-E. Generally speaking, villages in cluster A showed higher deforestation rates than most other clusters, and cluster E villages showed lower deforestation rates than most others. This result might be driven by the unequal country distribution in the clusters (figure 5.5a) meaning that sites in Brazil are in general characterized by higher deforestation rates, but it may also indicate that areas with high deforestation pressures are more likely to reach 'good' reduced deforestation results from interventions than areas with lower deforestation pressures. While REDD+ interventions often target areas facing higher deforestation pressures (Sunderlin et al., 2018), initial deforestation rates in cluster C villages (similar treatment intensities as cluster A, but with neutral or poor BA scores) were significantly lower than in cluster A.

Thus, we conclude that apart from forest cover in 2000, the clusters differed already considerably before interventions were applied.

5.3.2 Intervention type differences between clusters

A detailed overview of the intervention type composition per cluster can be found in appendix D.3. To examine whether differences in BA scores were associated with differences in intervention type composition, we focused on a comparison of the clusters A and C, which have similar TI scores, but differ in BA. Using a Levene's test, we checked whether we should assume equal variances in intervention type specific TI scores between the clusters. The p-values of these Levene's tests were not significant (i.e. p-values > 0.05), thus we assumed equal variances in the corresponding two sample t-tests. Detailed results can be found in appendix D.4.

Comparing the treatment intensities for the separate intervention types of clusters A and C, none were found to be significantly different (figure 5.6). Thus, for our sample there is no single intervention type that explains the difference between good and poor performance in terms of reduced deforestation rates. Further research is needed to explore which combinations of intervention types might be most effective. In addition, other aggregation methods than our clusters may be tested as, for example, different intervention types might differ in effectiveness between different countries. Or, as discussed in chapter 4 of this thesis, these results might indicate that a reduction in deforestation rate is not dependent upon a specific intervention type, but this supposition evidently requires further research.

5.3.3 Post-intervention changes in income and well-being

Income change

We assessed the cluster differences in income change between the two data collection periods in both absolute and relative terms. The results can be found in figure 5.7. Again, we used Kruskal-Wallis Rank Sum tests and unpaired Wilcoxon Rank Sum tests to assess whether the clusters differed significantly. Detailed test results can be found in appendix D.5. One cluster pair appeared to differ significantly in absolute terms only (A-E) which is again a reflection of differences in country distribution across the clusters, whereas two cluster pairs differed

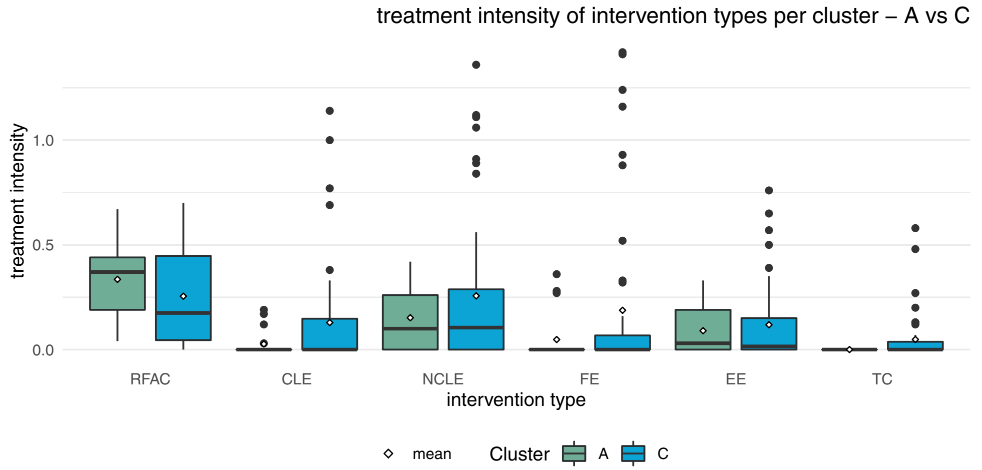


Figure 5.6: Boxplots of intervention type composition for clusters A and C. RFAC= Restrictions on forest access & conversion; CLE = conditional livelihood enhancements; NCLE = non-conditional livelihood enhancements; FE = forest enhancements; EE = environmental education; TC = tenure clarification. Upper and lower extremes of whiskers represent $Q3 + 1.5*$interquartile range (IQR) and $Q1 - 1.5 * IQR$ respectively, where $IQR = Q3 - Q1$.

significantly in both absolute and relative terms (A-C and A-D). Villages in cluster A experienced predominantly increases in income while in cluster C and D this was mixed (both decreases and increases) with medians and means close to 0. When assessing A and C as a single, merged cluster (both clusters with high TI), we find that a high TI is not necessarily associated with an increase (nor decrease) in income.

Perceived changes in household well-being

As mentioned earlier, in the post-intervention period (2013-2014), respondents were asked to evaluate both their household's general change in well-being as well as change in well-being associated with different interventions. Figure 5.8 shows the general perceived change in well-being, while figure 5.9 shows the reported effect of interventions on perceived well-being for clusters A and C, aggregated by intervention type. The results for the other clusters and for all households together can be found in appendix D.6. In addition, we clustered the effects on perceived well-being per country. These results can be found in appendix D.7.

Relatively speaking, slightly more respondents from households in cluster A reported an increase in perceived well-being than in other clusters. With regards to the effects of specific intervention types, we found that interventions were generally associated with a positive influence on well-being, particularly interventions of forest enhancement, conditional livelihood enhancements and environmental education. Of all intervention types, restrictions on forest access and conversion was reported to have the most negative influence on perceived well-being,

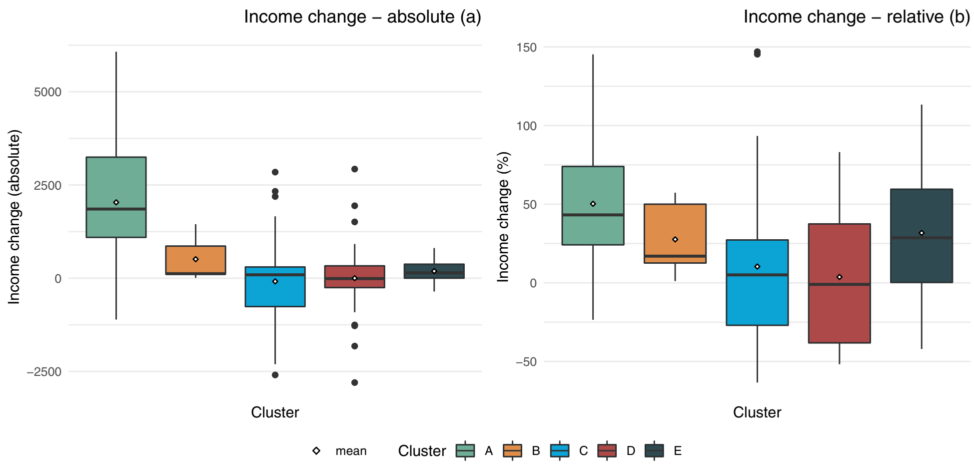


Figure 5.7: Income change per cluster from pre- to post-intervention phase (a) absolute and (b) relative terms. Upper and lower extremes of whiskers represent $Q3 + 1.5{*}$interquartile range (IQR) and $Q1 - 1.5 * IQR$ respectively, where $IQR = Q3 - Q1$.

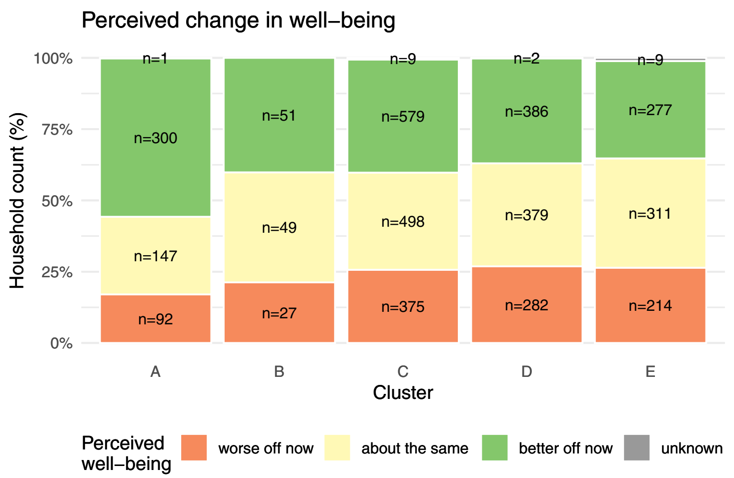


Figure 5.8: Respondents' general perceived change in well-being reported in the post-intervention phase. n represents number of respondents (households).

although the tendency was still towards the neutral or positive side. When comparing countries (appendix D.7), households sampled in sites in Brazil, Peru and Cameroon were more negatively affected by interventions, especially by RFAC, compared to households sampled in other countries.

5

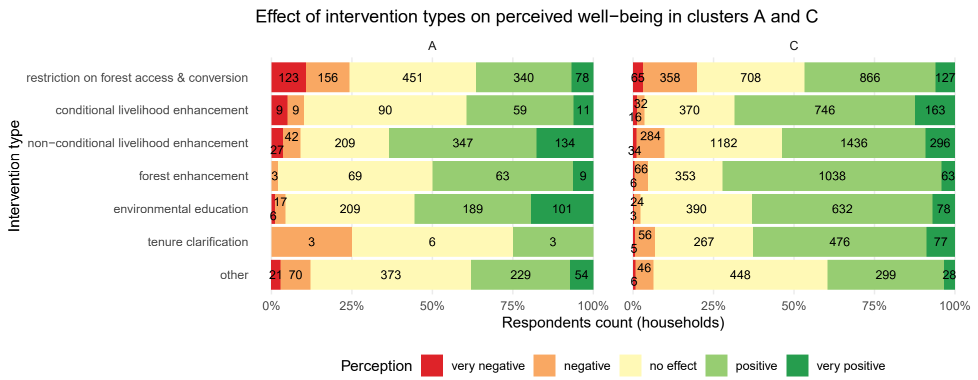


Figure 5.9: Effect of intervention types on perceived well-being in clusters A and C. Number inside the bars represent the number of respondents (households).

5.3.4 Reflections on exploratory analysis

We begin by reflecting on two methodological issues. First, although the thresholds for classifying the BA scores have been tested before (Bos et al., 2017), the treatment intensity thresholds that we used to define the clusters can be considered arbitrary. A sensitivity analysis would be needed to assess the effect the cut-off values used on the results. A second methodological caveat regards the mismatch in timeframes when assessing the change in socio-economic variables (income, perceived well-being) and the change in deforestation rates. The former is linked to the timing of the two surveys, that is, 2010-2011 for the pre-intervention period and 2013-2014 for the post-intervention period. The timeframe for the deforestation rates is 2001-2018, where the distinction between the *before* and *after* period is based on the start year of the associated REDD+ initiative (see section 5.2.4). Still, we argue that using these different timeframes in parallel is justified, as the effect on changes in land use behaviour inherently requires longer timeframes to detect.

Second, our analysis is not designed as a cause-effect examination of REDD+ interventions on forests and people, but is rather meant as an exploration of intervention treatment intensity, and forest and household/village characteristics that might explain differences in deforestation trends. Our focus on comparing clusters revealed some interesting findings as addressed above, but other levels of aggregation and further regression analyses are required to further enhance the understanding of the relations between REDD+ interventions and outcomes on forests and people, as described below. We have experienced the benefits of a carefully and thoroughly *ex ante* planned BACI research design, but also realize that in the absence of a controlled lab environment, one needs to be adaptive when it comes to the *ex post* evaluation. For this analysis, we diverted from the original BACI design of GCS REDD+ as in light of the particular objectives of this analysis, some of our assigned *control* villages could be considered *intervention* villages as they were not 'free' from forest interventions. Still, the results from our clustering method did show that the vast majority of villages in cluster E were both *ex ante* selected as control villages in GCS-REDD+, and in practice turned out to be evident controls. This result thus provides clear possibilities for further research to quantify the additionality of these interventions on reducing deforestation and to further enhance the understanding of trade-offs between carbon and well-being outcomes.

The findings of this exploratory analysis should be put into context of the GCS REDD+ impact evaluation results so far. First, we have observed no systematic negative impacts of REDD+ on local welfare across the study sites (Sunderlin et al., 2017; Duchelle et al., 2018a). Yet, it is clear that at sites where restrictions on forest access and conversion are applied more heavily (i.e. in Brazil), negative effects on subjective well-bring are more pronounced due to households being unable to clear forests to grow subsistence and cash crops (Sunderlin et al., 2017; Duchelle et al., 2017). At the same time, such restrictions were extremely effective in reducing forest clearing, especially among households at the Brazil sites that reported clearing larger areas of forest in the pre-intervention period (Duchelle et al., 2017). In this context, incentives for smallholders and communities (e.g. payments or infrastructure), significantly alleviated the burdens of land use restrictions (e.g. through law enforcement, protected areas) associated with some REDD+ initiatives (Duchelle et al., 2017).

5.3.5 Recommendations for further research

In this analysis, we did not find indications for general trade-offs between forest conservation and well-being outcomes. We did, however, find indications for potential trade-offs at the level of individual sites and villages that need to be studied further, possibly at the household level. For example, within the cluster where reduced deforestation rates were observed (A), some households encountered reductions in income as the lower whisker spread below 0 (figure 5.7). Likewise, some households in cluster A, but also in clusters with lower treatment intensities reported negative effects of certain interventions on their perceived well-being (figure 5.9). Conversely, we also observe villages with similar rates of forest loss but contrasting income well-being patterns, suggesting that these relationships may not always have similar behaviors.

Further research is needed to assess whether and under what conditions trade-offs (and possible synergies) between forest conservation and well-being outcomes exist at the village and household level. As next steps of the analysis, we will return to the BACI structure of the data to assess the impacts of treatment intensity of individual intervention types on deforestation, income and perceived changes in well-being at a higher level of aggregation (i.e. three categories instead of six: disincentives, incentives and enabling measures). Similarly, it may prove useful to better understand the relationship between our constructed measure of treatment intensity relative to the effect on perceived well-being of intervention types, as non-linear patterns in this relationship may help explain some of the divergent results. Results of the impact evaluation will also need to be better situated in context of the study countries and the specific theories of change of the different REDD+ initiatives.

5.4 Conclusions

This exploratory analysis provides insights into changes in forest cover and select measures of well-being in the context of different forest-based climate change mitigation interventions. We used data on forest cover change, treatment intensity and type of interventions, and two measures of well-being (income and perceived well-being) from seventeen sites across the tropics. Through assigning study villages into clusters with relatively homogenous deforestation and treatment intensity levels, we assessed select forest and well-being characteristics of these clusters in the pre-intervention period, and used the clusters to assess the effects of different treatment intensities on household incomes and perceived well-being.

Villages in the cluster with high treatment intensities and reduced deforestation rates consisted mostly of Brazilian villages. These villages had higher income levels and higher deforestation rates in the pre-intervention period compared to villages in other clusters. In the post-intervention period, they were generally associated with an increase in income with households reporting a slightly better level of perceived well-being compared to households in other clusters. In this particular analysis, we did not find indications of pronounced trade-offs between forest conservation and well-being outcomes, nor did we find clear differences in outcomes between different intervention types. At the individual village and household level however, there were indications for potential trade-offs as some of the villages and households in the cluster with high treatment intensities and reduced deforestation were associated with lower incomes and reported negative effects of interventions on perceived well-being.

In general, the lack of robust studies on forest and land use outcomes in the REDD+ literature makes it difficult to draw general conclusions about carbon versus well-being trade-offs (Duchelle et al., 2018c). While disincentives may be the most effective REDD+ instrument for conserving forests, other measures are needed to safeguard and enhance community well-being (Simonet et al., 2018; Duchelle et al., 2018a). This exploratory analysis is one way of looking at possible trade-offs based on treatment intensity of forest interventions. This type of information is particularly relevant in the context of REDD+ as it can provide valuable insights for policy makers and practitioners interested in developing REDD+ strategies that can provide both conservation and livelihood benefits.

Acknowledgements

All respondents in the study sites are gratefully acknowledged, as well as the CIFOR staff who did the data collection in the field.

6 Synthesis

6.1 Main findings

The main objective of this thesis was to explore and empirically test the use of environmental and socio-economic data sources to support subnational REDD+ performance assessment. As described in section 1.4, four research questions delineated this objective. Below, each of these questions is addressed, based on the findings from the corresponding chapters.

6.1.1 Forest change assessments – characteristics and consequences for REDD+

What are the characteristics and consequences of different forest change assessment approaches?

This question was addressed in chapter 2. To assess whether REDD+ interventions had an effect on their inherent goal -reducing emissions- forest changes over time need to be assessed, as well as changes in those trends compared to a certain counterfactual. In chapter 2, two forest change assessment methods were applied for comparison in data requirements and in measured REDD+ outcomes. The first, simpler Before-After (BA) method corresponds to the use of a conventional reference level, which is based on the assumption that expected future deforestation under a business-as-usual (BAU) scenario can be derived from historical deforestation (e.g. average of the past decade). The quasi-experimental Before-After-Control-Intervention (BACI) approach aims to control for confounding factors by assessing the unit of interest at two points in time (*before* and *after* the treatment) and in two different locations, that is, an area subjected to the 'treatment' (*intervention* area) and an area that is not (*control* area). In this way, when reductions in deforestation in the intervention area are observed, one can verify whether these changes are additional to what could have been expected without interventions.

A comparison of the BA and BACI methods shows that BA is simpler in terms of study design and data required. In principle, BA requires less decisions to be made and is thus less prone to subjective choices in the study design. Yet, (absence of) reductions in deforestation cannot be attributed to an intervention. BACI on the other hand is able to discern additionality attributable to interventions. A disadvantage of BACI is that it requires careful *ex ante* control site selection and matching, and that these results are sensitive to the decisions made in the matching process. These findings are not limited to the context of forest change assessments, however, but apply to other types of studies, such as socio-economic assessments on income and perceived well-being.

Our findings from chapter 2 showed that several factors affected both BA and BACI measurements in forest change assessments. For BACI to work, it is of vital importance that in the reference period, the control and intervention areas show similar rates of deforestation. Despite careful *ex ante* matching at village level, that was not the case in all of our study areas. Further, low absolute deforestation numbers and peak years influenced both our BA and BACI scores.

In terms of REDD+ performance outcomes, BA and BACI led to somewhat different conclusions. Overall, forest change assessment using BACI showed better REDD+ performance compared to BA, although the effect appeared to be more pronounced at the village (*micro*) level than at initiative level (*meso*). This thus reveals that especially at micro level, there were intervention villages with increasing deforestation rates during the intervention period, but these increases were less than in corresponding control villages. This raises the question whether increasing deforestation rates could still be considered 'good' REDD+ performance. On the other hand, when the magnitude of deforestation reduction in the intervention area is smaller than in the control area, a *good* BA score will correspond to a *poor* BACI score. The poor BACI score then indicates that the reduced deforestation in the intervention villages cannot be considered additional. While figure 2.4 on page 35 only indicates the 'net' differences between BA and BACI at the meso and micro level, figure 6.1 shows these internal 'shifts' in more detail. Figure 2.4 shows that at the meso level, BA and BACI do not appear to differ considerably. However, figure 6.1 indicates that these scores do differ greatly when looking at initiatives separately. At the micro level, the general tendency that BACI scores indicate better REDD+ performance is visible in figures 2.4 and 6.1 alike.

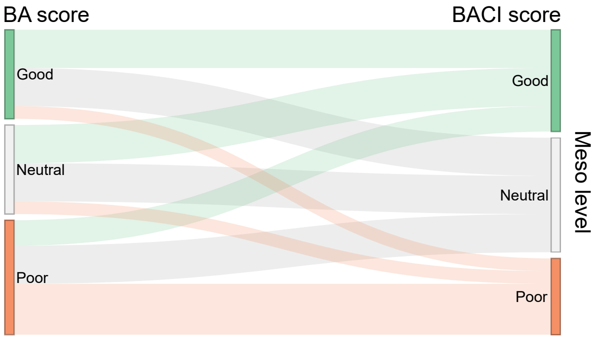
BACI was found to be more relevant at the local level, as constructing comparable control areas at scales higher than landscape level becomes problematic. Thus, national REDD+ monitoring would most likely not benefit from a BACI approach, as national level control areas would be difficult, if not impossible, to define.

6.1.2 Global data for local use – data uncertainty and consequences for REDD+

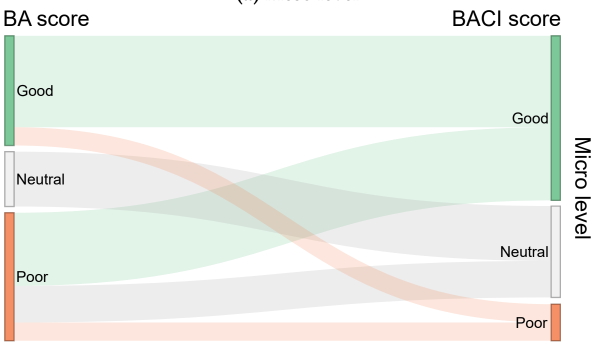
How do the availability and choices of forest change datasets influence REDD+ measurements and corresponding uncertainties?

This question was addressed in chapter 3. In recent decades, the remote sensing and forest monitoring arena is characterized by rapid developments in terms of ever-increasing levels of coverage, accuracy, and spatial, temporal and spectral detail. This resulted in a plethora of both global map products and advanced open-source algorithms alike. These products and tools are not always consistent or complementary, and data uncertainties affect REDD+ performance estimates in unknown ways. Therefore, in chapter 3, we compared and tested two of these products for the purpose of REDD+ performance assessment, of which one is a relatively easy to use global dataset and the other requires more technical know-how and computing power to implement, but is flexible and can be tailored to the local context.

Our results showed that the accuracies of the various data sources differ at site level, although on average neither one of the products excelled in accuracy consistently. Yet, both products underestimated deforestation, as reflected by the lower producer's accuracy and corresponding



(a) Meso level



(b) Micro level

Figure 6.1: Relations between individual BA and BACI scores at meso and micro level

higher omission errors. The use of a combination of both products as stratification for area estimation and validated with a sample of high-resolution data showed promising results. In these combined products, the expected trade-offs in accuracies across change classes (pre-intervention, post-intervention, no change) and across accuracy types (user's and producer's accuracy) were negligible, so their use is advantageous over single-source datasets. In four out of five sites, using a combination strategy based on the earliest detection of deforestation, led to higher accuracies than the individual products. Yet, the direction (i.e. reduced, stable or increased deforestation trends) and degree (i.e. magnitude of observed changes) of REDD+ performance remained statistically uncertain, as in three sites CIs of the deforestation estimates of pre- and post-intervention periods were overlapping. Although the developments of new, more detailed map products and tools are usually associated with lower uncertainties, a complete elimination of data uncertainty is unrealistic. Therefore, for REDD+ performance assessments, decisions need to be made concerning the degree of uncertainty that can be considered acceptable for accounting purposes, especially at the local level.

Estimating deforestation using a visual validation with a stratified sampling strategy could substantially reduce uncertainty in REDD+ performance assessments. The development of more locally calibrated wall-to-wall products could make them more useful and applicable for REDD+ purposes. Based on the results in chapter 3, it is advised to take note of and address these data uncertainties and inaccuracies through conservative and transparent REDD+ accounting, while not relying on results from single-source currently available global datasets alone.

6.1.3 Deforestation drivers and alignment with REDD+ interventions

How can an integrated deforestation drivers assessment help understand driver-intervention alignment?

This question was addressed in chapter 4. The study builds upon empirical research on early, local REDD+ projects in which different mixes of interventions (i.e. enabling measures, disincentives and incentives) were implemented by private sector and non-governmental organizations.

The study in chapter 4 contributes to reducing the knowledge gap on how to assess drivers at the local level in the context of REDD+. Our drivers assessment included a combination of optical remote sensing, spatial modelling and socio-economic survey methods to gain a better understanding of the complex interactions between driver elements (agents, activities, environmental factors and underlying forces) at the local level. We found considerable complementary power in combining different methods. The research in chapter 4 showed that no single dataset or method can reveal all facets (who, what, where, why, when and how much) of DD drivers, which stresses the added value of applying a combined assessment.

In terms of driver-intervention alignment, we found that higher level interventions focussing on restrictive use of forest resources generally do not target a particular DD driver, while in general local level interventions target primarily small holders and communities to either provide non-conditional livelihoods enhancements and forest enhancement practices to stimulate sustainable agricultural practices, or focus on tenure clarification to strengthen the position of local small holders and communities against large scale exogenous drivers. Yet, some of the drivers found in chapter 4 were inadequately addressed by the interventions included in the study, particularly those related to practices like illegal mining and illegal logging. The inter- and intra-site differences in drivers that were found underscore the importance of analysing micro-level DD drivers in a spatially explicit way, and of designing and implementing locally tailored interventions.

6.1.4 Changes in forest cover and well-being in the context of different REDD+ interventions

What are the relationships between changes in forests and non-carbon outcomes in the context of different REDD+ interventions?

This question is linked to the study in chapter 5. In that chapter, we applied a data-driven approach through which we tried to 'unpack' differences in forest change outcomes by evaluating both pre- and post-intervention characteristics in forest cover, forest cover change, and well-being. This exploratory analysis will feed into a more rigorous impact evaluation in the future, in order to understand what interventions worked in terms of reducing deforestation, why it worked, why there, and, in case trade-offs are observed, at what cost?

The study in chapter 5 comprised of an exploratory analysis to examine the relationships between the treatment intensity and type of forest conservation interventions, changes in forest cover (loss) and changes in income and well-being for different types of forest interventions (disincentives, incentives and enabling measures) in seventeen subnational sites across the tropics. Clusters of villages were defined based on similar levels of intervention treatment intensities and deforestation trends to compare pre- and post-intervention characteristics.

We found that villages in the cluster with high treatment intensities and reduced post-intervention deforestation rates primarily consisted of Brazilian villages, had higher income levels and deforestation rates in the pre-intervention period. In the post-intervention period, these villages were generally associated with an increase in income and its households reported a slight increase in perceived well-being. Differences in intervention types were not driving these outcomes. No clear trade-offs were found between forest conservation outcomes (i.e. a reduction in deforestation) and well-being outcomes across all study sites as a whole. Yet, trade-offs in specific villages and households could not be ruled out. Further research is needed at local scales and initiative level to apply a regression model and unpack impacts and trade-offs further. REDD+ strategies that can provide both conservation and livelihood benefits will ultimately be designed and improved based on empirical findings like these.

6.2 Reflection and outlook

6.2.1 Forest fires and tropical deforestation – a hot topic

At the time of writing (Third quarter of 2019), Brazil seems to head for a record-breaking year in terms of forest fires, with the highest number in fire counts since recordkeeping of fires started in 2013 (Gajanan, 2019). Meanwhile, in Indonesia, nearly a million people suffer from respiratory problems caused by thousands of wildfires on Sumatra and Kalimantan (Paddock et al., 2019). Fed by newspaper articles like these, international pressure is rising to counter these fires and tropical deforestation. Yet, scientists warn to not jump to conclusions, as the data on which these alarming newspaper headlines are based, are easily misinterpreted (Molijn, 2019).

To give an example, in the case of the Brazilian forest fires, the spatial resolution of imagery from the Brazilian INPE satellites is approximately 375m. Each time a satellite passes over a certain area and detects a fire, a fire count is recorded. One fire may encompass more than one pixel, one pixel can contain multiple fires, and a single prolonged fire can be counted multiple times when a satellite revisits the same area. To what extent these yearly fires are occurring in forests or rather on agricultural fields (i.e. for clearing biomass) is not easily detectable, and takes months to verify (Molijn, 2019).

Thus, estimating the amount and magnitude of forest fires -and deforestation in general- is complex, and the public debate would benefit from information which is presented with a transparent explanation vis-à-vis its abilities, limitations and accuracies.

6.2.2 Data-driven decisions in the policy cycle

The forest fire example in the previous section illustrates that (spatial) data on forests, forest change, and changes in forest trends play an important role in raising awareness and in defining a problem such as deforestation. In other stages of the policy cycle too, provision of transparent information is important for informed decision-making (De Sy et al., 2018).

In the policy option and selection stage (figure 6.2), based on spatially explicit information on deforestation hotspots and DD drivers, policy-makers can decide what areas and which agents to prioritise on, for example as part of national climate change mitigation plans written for the Paris Agreement (i.e. NDCs). In the implementation stage, repeatedly updated forest data allow for continuous progress tracking of interventions. Recent advances in remote sensing allow for near-real time alert systems targeting logging or fires, for example with the Brazilian DETER-B system (Diniz et al., 2015) and GLAD Forest Loss Alerts (Hansen et al., 2016). Finally, for evaluation and performance assessment purposes, forest observation data

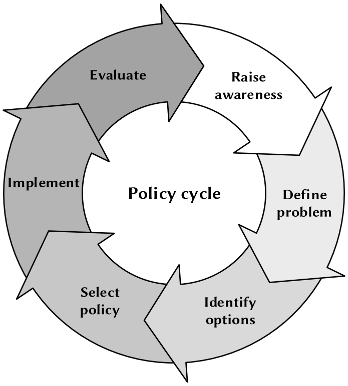


Figure 6.2: Stages in the policy cycle

plays a role at different scales, varying from estimating UNFCCC stock-takes at the global level, to national-level GHG and NDCs progress reporting, and locally, as discussed in this thesis, intervention effectiveness and impact evaluation. In order to proceed towards an optimal use of data and achieve effective policies, we have to consider whether we are doing things right in terms of data usage and assessment methodology (chapter 2,3,4), and whether we are focussing on the right things (drivers and interventions) in the first place (chapter 4,5, see also figure 1.2 on page 25). These two questions will be discussed in the following two sections respectively.

6.2.3 Are we doing things right? Monitoring & evaluation

The question *'are we doing things right'* refers to the methodological and data related factors to consider when monitoring forests, assessing drivers and evaluating REDD+ performance.

Chapters 2 and 3 have shown that both the assessment method -and corresponding reference level type- and datasets used may influence the direction and degree of the measured REDD+ effect considerably. This raises the issue that data could be framed in such a way, that it tells the story the reporting agency wants to present, thus undermining the values of integrity and trust. In light of the diversification of REDD+ implementers, transparent monitoring and evaluation should therefore be considered the cornerstones in building the credibility of REDD+. Additionally, independent (non-state) stakeholders including UNFCCC experts and local implementers require data for their own performance assessments, for verifying assessments of other agencies or governments, or for global stock-take purposes. Therefore, it is likely that scientists and other independent organizations will play an increasingly important role in the provision of data and tools.

Chapter 4 has shown that integrating different datasets or methods helps to better understand all facets of deforestation drivers (who, what, where, why, when and how much) compared to a single dataset alone. For evaluations on local deforestation drivers it is therefore advised not to rely on single-source datasets or methods. The value and use of drivers assessments will be discussed in more detail in the next section.

As discussed in chapters 2 and 5, in order to attribute observed changes to certain interventions applied, one requires a study design which includes control units as counterfactual, such as the BACI design. Inherent to this BACI design is the matching procedure, in which intervention and control sites are matched based on key factors that could have been potential confounders. Still, intervention and control villages differed in the sense that the majority of intervention villages, but a minority of control villages, had prior experience with conservation NGOs (Lin, 2012; Sills et al., 2017), which indicates that the location of local REDD+ initiatives is not random, but often a continuation of existing conservation efforts (Sills et al., 2014). This might be a confounding factor too, in the sense that people who have been taught sustainable agricultural or forest conservation practices prior to REDD+, might be more likely to show more sustainable land use behaviour regardless of the REDD+ intervention.

This phenomenon is not unique to conservation studies, as it has very similar characteristics to the so-called 'healthy donor effect' in medical studies (e.g. Ullum et al., 2015). In these studies, blood donors are often associated to have lower mortality rates and fewer health issues compared to non-donors. This cannot be attributed to the blood donation itself however, but is rather caused by the fact that blood donors are often inherently healthier than non-donors, as their health status is repeatedly being checked throughout their donation career leading to informed and health-conscious donors (Peffer, 2015). Thus, one could argue that intervention villages in conservation studies may be considered 'healthy donors' too, which needs to be accounted for in order to show unbiased REDD+ performance results.

6.2.4 Are we doing the right things? Interventions, drivers & outcomes

The question *'are we doing the rights things'* refers to the interventions, to assess whether these are targeting the drivers (chapter 4), and in terms of outcomes, to what extent these interventions affect forest change and well-being (chapter 5).

In the past decade in which REDD+ design and monitoring has developed, but also in this thesis, most emphasis has been on developing performance assessments methods for REDD+. Few studies have focussed on measuring actual REDD+ performance assessment on the ground (Duchelle et al., 2018c), perhaps because overall REDD+ progress has been much slower than anticipated (Angelsen et al., 2012). Scholars have attributed the overall underwhelming progress to a lack of (continued) finance (Norman and Nakhooda, 2015; Duchelle et al., 2018b) and the related low treatment intensity (Fischer et al., 2016) or the slow progress on land tenure and carbon rights reform (e.g. Dunlop and Corbera, 2016). Most local REDD+ initiatives have focused mainly on smallholders, even though the drivers of deforestation often operate at larger scales (Luttrell et al., 2018), so for enhancing the carbon performance it is argued to broaden the focus of REDD+ interventions to other agents of deforestation (Bos et al., 2017; Thaler and Anandi, 2017). One of the main rationales for setting up MRV systems and doing performance assessments has been to get results-based payments off the ground (Wong et al., 2016; Korhonen-Kurki et al., 2013; La Viña and de Leon, 2014), but perhaps obtaining an enhanced understanding of the complex and changing interactions of interventions and drivers is an additional value of doing performance assessments that is currently underestimated.

6.2.5 Towards robust performance assessments

The word 'towards' in the title of this thesis implies that the findings of this study contribute to the development of performance assessments, by providing insights in the implications of choosing certain datasets or methods, and by exploring ways to examine driver-intervention alignment and the intervention-outcome relations. But which promising developments in terms of data provision lie ahead and which features for setting up a robust REDD+ performance assessment system are still missing?

Developments in forest monitoring

As discussed earlier in this synthesis (section 6.1.2), the data and tools available for monitoring forest change are evolving rapidly. In spaceborne remote sensing, for example, advances in global radar data from Sentinel-1 allow near real-time forest monitoring (Reiche et al., 2016). Compared to optical remote sensing, radar has the ability to penetrate through clouds and haze, and to detect degradation (Olander et al., 2008). Increasingly available airborne and terrestrial LiDAR (Light Detection And Ranging) data help mapping the vertical forest structure, to estimate tree heights, tree volume and biomass (GOFC-GOLD, 2016; Hyyppä et al., 2012). Monitoring agencies including REDD+ countries need to have the technical capacity and funds to truly benefit from these technological advances (Petersen et al., 2018; Romijn et al., 2015). To this end, Norway recently pledged to spend 53 million United States Dollar (USD) on high resolution imagery (Solsvik, 2019) and granting collaborating REDD+ countries free access to bring transparent forest monitoring for conservation purposes to the next level.

This thesis has focussed mainly on the activity data (forest area change) component of carbon assessments (see also chapter 1), as it was assumed that when plot level data on carbon stocks and fluxes are not available, and when carbon data are thus limited to a single emission factor per site, the amount of forest loss (activity data) is driving the emissions from deforestation. Currently, there is not yet a single biomass dataset that serves all users' needs with regards to area coverage, timeliness and uncertainty requirements (Herold et al., 2019). Yet, developments in carbon data provision (see chapter 1), and developments in satellite missions (e.g. ESA's GlobBiomass project and the Sentinel satellites) allow to estimate emission factors with much greater detail and accuracies, so future studies will benefit from these developments to incorporate these in REDD+ performance assessments.

Scalar integration – a 'nested' approach

REDD+ is inherently a multi-level framework. As Korhonen-Kurki et al. (2012, p. 91) put it, "local people face global demands for climate change mitigation that must be effected through existing and emerging national and subnational institutions and structures". While most funds are organised at the national level through bilateral agreements, the fundamental idea of REDD+ was that carbon rights holders receive direct compensation for their activities and corresponding avoided emissions.

Just as the payments, also policy implementations have to cross different levels, as explained by Seymour and Angelsen (2009, p. 295): "National REDD+ institutions must make upward and downward linkages: transferring funds from the national to the local level, managing incentives (both policy measures and payments) and channelling information from the local to the national and international levels". Although national monitoring systems are being put in place with increased levels of detail and accuracy, in light of results-based payments there are calls to adjust the level of monitoring to the level in line to where the payments are made (Vatn and Angelsen, 2009). A 'nested' approach to REDD+ would entail that multiple levels of forest governance are linked, also through translocal and transnational networks (Visseren-Hamakers et al., 2012). In terms of carbon monitoring, this nesting would also prevent double counting and inefficient mitigation. This requires decisions on what to measure, at what scale and effort, and how reporting from local to international levels should have to take place (de Sassi et al., 2015). The broadening set of implementers makes this scalar integration difficult, yet it is crucial for attributing the performance to interventions at higher levels and to assess potential leakage.

Disciplinary integration

This thesis assessed outcomes in terms of potential social co-benefits or trade-offs by focusing on income and perceived well-being only. Certainly, this is a simplification, as other social elements have to be considered too, including, but not limited to, status and change in tenure security (Sunderlin et al., 2014c) and women's empowerment (Larson et al., 2018). Moreover, this thesis did not consider how to monitor the potential environmental co-benefits or trade-offs with respect to, for example, biodiversity (Larson et al., 2018).

In a recent call for an integrated monitoring approach, opportunities for scalar and disciplinary integration for REDD+ monitoring were suggested (de Sassi et al., 2015) (see figure 6.3). An example of a scalar integration is the combined use of both local level forest inventory plot data and higher level remotely sensed wall-to-wall maps in national forest monitoring systems. An example of disciplinary integration is the combined monitoring of carbon and socio-economic outcomes, to detect potential synergies or trade-offs (see also chapter 4). Since reductions in carbon emissions must be achieved while aiming for co-benefits and avoid doing harm to social and environmental values over time and across scales (Brown et al., 2008), integrated monitoring for REDD+ is essential.

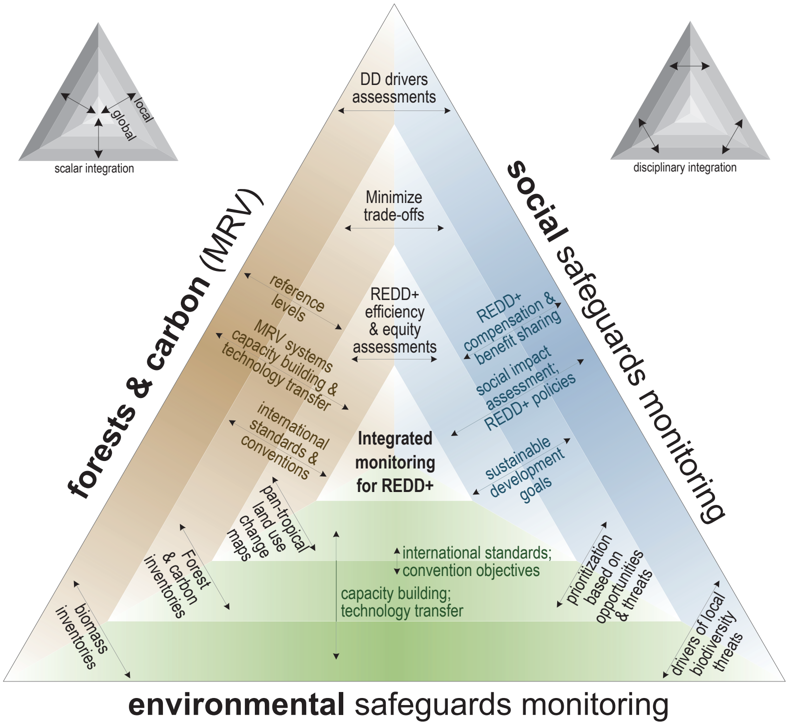


Figure 6.3: Opportunities for disciplinary and scalar integration for REDD+ monitoring. Concepts with coloured fonts represent cross-scalar elements. Concepts with black fonts represent cross-disciplinary elements. (Adapted from de Sassi et al., 2015)

This thesis has shown that REDD+ performance assessments at the subnational level are not set in stone yet, and that they are affected by choices in method, datasets and corresponding data uncertainties. Nevertheless, with recent and future technological developments and increasing experience through empirical studies on the ground, accurate, transparent and timely information about the state of the forests and the people who depend on them is within reach.

References

Angelsen, A., Brockhaus, M., Sunderlin, W., and Verchot, L. (2012). *Analysing REDD+: Challenges and choices.* Center for International Forestry Research (CIFOR), Bogor, Indonesia.

Angelsen, A., Jagger, P., Babigumira, R., Belcher, B., Hogarth, N. J., Bauch, S., Börner, J., Smith-Hall, C., and Wunder, S. (2014). Environmental Income and Rural Livelihoods: A Global-Comparative Analysis. *World Development*, 64(S1):S12–S28.

Angelsen, A., Martius, C., de Sy, V., Duchelle, A., Larson, A., Pham, T., and (eds.) (2018). *Transforming REDD+: Lessons and new directions.* Center for International Forestry Research (CIFOR).

Angelsen, A. and McNeill, D. (2012). The evolution of REDD+. In Angelsen, A., editor, *Analysing REDD+: Challenges and choices*, chapter 3, pages 31–48. Center for International Forestry Research (CIFOR), Bogor, Indonesia.

Asner, G. P., Knapp, D. E., Balaji, A., and Paez-Acosta, G. (2009). Automated mapping of tropical deforestation and forest degradation: CLASlite.

Asner, G. P., Llactayo, W., Tupayachi, R., and Luna, E. R. (2013a). Elevated rates of gold mining in the Amazon revealed through high-resolution monitoring. *Proceedings of the National Academy of Sciences*, 110(46):18454–18459.

Asner, G. P., Mascaro, J., Anderson, C., Knapp, D. E., Martin, R. E., Kennedy-Bowdoin, T., van Breugel, M., Davies, S., Hall, J. S., Muller-Landau, H. C., Potvin, C., Sousa, W., Wright, J., and Bermingham, E. (2013b). High-fidelity national carbon mapping for resource management and REDD+. *Carbon balance and management*, 8:7.

Atmadja, S., Arwida, S., Martius, C., and Pham, T. (2018). Financing REDD+: A transaction among equals, or an uneven playing field? In Angelsen, A., Martius, C., De Sy, V., Duchelle, A., Larson, A., and Pham, T., editors, *Transforming REDD+: Lessons and new directions*, chapter 3, pages 29–40. Center for International Forestry Research (CIFOR), Bogor, Indonesia.

Austin, K. G., Schwantes, A., Gu, Y., and Kasibhatla, P. S. (2019). What causes deforestation in

Indonesia? *Environmental Research Letters*, 14(2):024007.

Avitabile, V., Herold, M., Heuvelink, G. B., Lewis, S. L., Phillips, O. L., Asner, G. P., Armston, J., Ashton, P. S., Banin, L., Bayol, N., Berry, N. J., Boeckx, P., de Jong, B. H. J., DeVries, B., Girardin, C. A. J., Kearsley, E., Lindsell, J. A., Lopez-Gonzalez, G., Lucas, R., Malhi, Y., Morel, A., Mitchard, E. T. A., Nagy, L., Qie, L., Quinones, M. J., Ryan, C. M., Ferry, S. J. W., Sunderland, T., Laurin, G. V., Gatti, R. C., Valentini, R., Verbeeck, H., Wijaya, A., and Willcock, S. (2016). An integrated pan-tropical biomass map using multiple reference datasets. *Global Change Biology*, 22(4):1406–1420.

Baccini, A., Goetz, S. J., Walker, W. S., Laporte, N. T., Sun, M., Sulla-Menashe, D., Hackler, J., Beck, P. S. A., Dubayah, R., Friedl, M. A., Samanta, S., and Houghton, R. A. (2012). Estimated carbon dioxide emissions from tropical deforestation improved by carbon-density maps. *Nature Climate Change*, 2(3):182–185.

Bellot, F.-F., 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.* FORCLIME Forests and Climate Change Programme, Jakarta, Indonesia.

Belward, A. S. and Skøien, J. O. (2014). Who launched what, when and why; trends in global land-cover observation capacity from civilian earth observation satellites. *ISPRS Journal of Photogrammetry and Remote Sensing*.

Blanco, G., Gerlagh, R., Suh, S., Barrett, J., de Coninck, H., Morejon, C. D., Mathur, R., Nakicenovic, N., Ahenkora, A. O., Pan, J., Pathak, H., Rice, J., Richels, R., Smith, S., Stern, D., Toth, F., and Zhou, P. (2014). Drivers, Change and Mitigation. In Edenhofer, O., Pichs-Madruga, R., Sokona, Y., Farahani, E., Kadner, S., Seyboth, K., Adler, A., Baum, I., Brunner, S., Eickemeier, P., Kriemann, B., Savolainen, J., Schlömer, S., von Stechow, C., Zwickel, T., and Minx, J., editors, *Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, page 412. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Börner, J., Wunder, S., Wertz-Kanounnikoff, S., Hyman, G., and Nascimento, N. (2014). Forest law enforcement in the Brazilian Amazon: Costs and income effects. *Global Environmental Change*, 29:294–305.

Bos, A. B., De Sy, V., Duchelle, A. E., Herold, M., Martius, C., and Tsendbazar, N.-E. (2019). Global data and tools for local forest cover loss and REDD+ performance assessment: Accuracy, uncertainty, complementarity and impact. *International Journal of Applied Earth Observation and Geoinformation*, 80:295–311.

Bos, A. B., Duchelle, A. E., Angelsen, A., Avitabile, V., De Sy, V., Herold, M., Joseph, S., de Sassi, C., Sills, E. O., Sunderlin, W. D., and Wunder, S. (2017). Comparing methods for assessing the effectiveness of subnational REDD+ initiatives. *Environmental Research Letters*, 12(7):074007.

Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1):5–32.

Brown, D., Seymour, F., and Peskett, L. (2008). How do we achieve REDD co-benefits and avoid doing harm? *Moving Ahead with REDD: Issues, Options and Implications*, pages 107–118.

CIFOR (2017). Global Comparative Study on REDD+ Subnational REDD+ Initiatives.

Clements, T., Suon, S., Wilkie, D. S., and Milner-Gulland, E. (2014). Impacts of Protected Areas on Local Livelihoods in Cambodia. *World Development*, 64(S1):S125–S134.

Cochran, W. G. (1977). *Sampling techniques*. Wiley, New York.

Cohen, W. B., Yang, Z., and Kennedy, R. (2010). Detecting trends in forest disturbance and recovery using yearly Landsat time series: 2. TimeSync — Tools for calibration and validation. *Remote Sensing of Environment*, 114(12):2911–2924.

Cromberg, M., Duchelle, A. E., and Rocha, I. d. O. (2014a). Local Participation in REDD+: Lessons from the Eastern Brazilian Amazon. *Forests*, 5(4):579–598.

Cromberg, M., Duchelle, A. E., Simonet, G., and de Freitas, A. C. (2014b). Sustainable Settlements in the Amazon, Brazil. In Sills, E., Atmadja, S., de Sassi, C., Duchelle, A., Kweka, D., Resosudarmo, I., and Sunderlin, W., editors, *REDD+ on the ground: A case book of subnational initiatives across the globe*, chapter 7, pages 124–143. Center for International Forestry Research (CIFOR), Bogor, Indonesia.

Curtis, P. G., Slay, C. M., Harris, N. L., Tyukavina, A., and Hansen, M. C. (2018). Classifying drivers of global forest loss. *Science*, 361(6407):1108–1111.

de Sassi, C., Joseph, S., Bos, A. B., Duchelle, A. E., Ravikumar, A., and Herold, M. (2015). Towards integrated monitoring of REDD+. *Current Opinion in Environmental Sustainability*, 14:93–100.

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. *Current Opinion in Environmental Sustainability*, 4(6):696–706.

De Sy, V., Herold, M., Achard, F., Beuchle, R., Clevers, J. G. P. W., Lindquist, E., and Verchot, L. (2015). Land use patterns and related carbon losses following deforestation in South America. *Environmental Research Letters*, 10(12):124004.

De Sy, V., Herold, M., Brockhaus, M., Di Gregorio, M., and Ochieng, R. (2018). Information and policy change: Data on drivers can drive change - if used wisely. In Angelsem, A., Martius, C., Sy, V. D., Duchelle, A., Larson, A., and T.T., P., editors, *Transforming REDD+: Lessons and new directions*, chapter 5, pages 55–67. CIFOR.

Defries, R. S., Rudel, T., Uriarte, M., and Hansen, M. (2010). Deforestation driven by urban population growth and agricultural trade in the twenty-first century. *Nature Geoscience*.

DeVries, B., Verbesselt, J., Kooistra, L., and Herold, M. (2015). Robust monitoring of small-scale forest disturbances in a tropical montane forest using Landsat time series. *Remote Sensing of Environment*.

Diela, T. (2019). Indonesia president makes moratorium on forest clearance permanent - Reuters.

Diniz, C. G., Souza, A. A. d. A., Santos, D. C., Dias, M. C., da Luz, N. C., de Moraes, D. R. V., Maia, J. S. A., Gomes, A. R., Narvaes, I. d. S., Valeriano, D. M., Maurano, L. E. P., and Adami, M. (2015). DETER-B: The New Amazon Near Real-Time Deforestation Detection System. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(7):3619–3628.

Duchelle, A. E., Cromberg, M., Gebara, M. F., Guerra, R., Melo, T., Larson, A., Cronkleton, P., Börner, J., Sills, E., Wunder, S., Bauch, S., May, P., Selaya, G., and Sunderlin, W. D. (2014). Linking Forest Tenure Reform, Environmental Compliance, and Incentives: Lessons from REDD+ Initiatives in the Brazilian Amazon. *World Development*, 55:53–67.

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. *Ecology and Society*, 22(3).

Duchelle, A. E., de Sassi, C., Sills, E. O., and Wunder, S. (2018a). People and communities: Well-being impacts of REDD+ on the ground. In Angelsen, A., Martius, C., De Sy, V., Duchelle, A., Larson, A., and Pham, T., editors, *Transforming REDD+: Lessons and new directions*, chapter 11, pages 131–141. Center for International Forestry Research (CIFOR), Bogor, Indonesia.

Duchelle, A. E., Herold, M., and de Sassi, C. (2015). Monitoring REDD+ Impacts: Cross Scale Coordination And Interdisciplinary Integration. In *Sustainability Indicators in Practice*, chapter 4, pages 55–79. De Gruyter Open, Berlin, Germany.

Duchelle, A. E., Seymour, F., Brockhaus, M., Angelsen, A., Larson, A. M., Moeliono, M., Wong, G. Y., Pham, T. T., and Martius, C. (2018b). REDD+: Lessons From National and Subnational Implementation - Ending Tropical Deforestation: a Stock-Take of Progress and Challenges.

Duchelle, A. E., Simonet, G., Sunderlin, W. D., and Wunder, S. (2018c). What is REDD+ achieving on the ground? *Current Opinion in Environmental Sustainability*, 32:134–140.

Dunlop, T. and Corbera, E. (2016). Incentivizing REDD+: How developing countries are laying the groundwork for benefit-sharing. *Environmental Science & Policy*, 63:44–54.

FAO (2000). FRA 2000 on definitions of forest and forest change. Technical report, Food and Agriculture Organization of the United Nations, Rome.

Ferraro, P. J. (2009). Counterfactual thinking and impact evaluation in environmental policy. *New Directions for Evaluation*, 2009(122):75–84.

Fick, S. E. and Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37(12):4302–4315.

Finer, B. M., Novoa, S., Weisse, M. J., Petersen, R., Mascaro, J., Souto, T., Stearns, F., and Martinez, R. G. (2018). Combating deforestation: From satellite to intervention.

Fischer, R., Hargita, Y., and Günter, S. (2016). Insights from the ground level? A content analysis review of multi-national REDD+ studies since 2010. *Forest Policy and Economics*, 66:47–58.

Foody, G. M. (2009). Sample size determination for image classification accuracy assessment and comparison. *International Journal of Remote Sensing*, 30(20):5273–5291.

Gajanan, M. (2019). A Record Number of Fires Are Currently Burning Across the Amazon Rainforest.

Garrish, V., Perales, E., Duchelle, A. E., and Cronkleton, P. (2014). The REDD Project in Brazil Nut Concessions in Madre de Dios, Peru. In Sills, E., Atmadja, S., de Sassi, C., Duchelle, A., Kweka, D., Resosudarmo, I., and Sunderlin, W., editors, *REDD+ on the ground: A case book of subnational initiatives across the globe*, chapter 8, pages 147–165. Center for International Forestry Research (CIFOR), Bogor, Indonesia.

Gaveau, D. L., Locatelli, B., Salim, M. A., Yaen, H., Pacheco, P., and Sheil, D. (2018). Rise and fall of forest loss and industrial plantations in Borneo (2000-2017). *Conservation Letters*, page e12622.

Geist, H. J. and Lambin, E. F. (2002). Proximate causes and underlying driving forces of tropical deforestation. *BioScience*, 52(2):143–150.

Geoghegan, J. (1998). "Socializing the Pixel" and "Pixelizing the Social" in Land-Use and Land-Cover Change. In *People and Pixels: Linking Remote Sensing and Social Science*, pages 51–69.

GFOI (2016). Integration of remote-sensing and ground-based observations for estimation of emissions and removals of greenhouse gases in forests. Technical report, FAO, Rome.

Godar, J., Gardner, T. A., Tizado, E. J., and Pacheco, P. (2014). Actor-specific contributions to the deforestation slowdown in the Brazilian Amazon. *Proceedings of the National Academy of Sciences*, 111(43):15591–15596.

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. Technical report, GOFC-GOLD Land Cover Project Office, Wageningen University, Wageningen, the Netherlands.

Goodman, R. C. and Herold, M. (2014). Why Maintaining Tropical Forests Is Essential and Urgent for a Stable Climate. CGD Working Paper, page 51. Washington, DC.

Grassi, G., House, J., Dentener, F., Federici, S., den Elzen, M., and Penman, J. (2017). The key role of forests in meeting climate targets requires science for credible mitigation. *Nature Clim. Change*, 7.

Grassi, G., Monni, S., Federici, S., Achard, F., and Mollicone, D. (2008). Applying the conservativeness principle to REDD to deal with the uncertainties of the estimates. *Environmental Research Letters*, 3(3):035005.

Griscom, B. W., Adams, J., Ellis, P. W., Houghton, R. A., Lomax, G., Miteva, D. A., Schlesinger, W. H., Shoch, D., Siikamäki, J. V., Smith, P., Woodbury, P., Zganjar, C., Blackman, A., Campari, J., Conant, R. T., Delgado, C., Elias, P., Gopalakrishna, T., Hamsik, M. R., Herrero, M., Kiesecker, J., Landis, E., Laestadius, L., Leavitt, S. M., Minnemeyer, S., Polasky, S., Potapov, P., Putz, F. E., Sanderman, J., Silvius, M., Wollenberg, E., and Fargione, J. (2017). Natural climate solutions. *Proceedings of the National Academy of Sciences*, 114(44):11645–11650.

Gross, D., Achard, F., Dubois, G., Brink, A., and Prins, H. H. T. (2017). Uncertainties in tree cover maps of Sub-Saharan Africa and their implications for measuring progress towards CBD Aichi Targets. *Remote Sensing in Ecology and Conservation*, pages 1–19.

Hansen, M. C., Krylov, A., Tyukavina, A., Potapov, P. V., Turubanova, S., Zutta, B., Ifo, S., Margono, B., Stolle, F., and Moore, R. (2016). Humid tropical forest disturbance alerts using Landsat data. *Environmental Research Letters*, 11(3):034008.

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., Loveland, T. R., Kommareddy, A., Egorov, A., Chini, L., Justice, C. O., and Townshend, J. R. G. (2013). High-resolution global maps of 21st-century forest cover change. *Science*, 342(6160):850–3.

Harris, N. L., Brown, S., Hagen, S. C., Saatchi, S. S., Petrova, S., Salas, W., Hansen, M. C., Potapov, P. V., and Lotsch, A. (2012). Baseline map of carbon emissions from deforestation in tropical regions. *Science*, 336(6088):1573–6.

Herold, M., Angelsen, A., Verchot, L. V., Wijaya, A., and Ainembabazi, J. H. (2012). A stepwise framework for developing REDD+ reference levels. In Angelsen, A., editor, *Analysing REDD+: Challenges and choices*, chapter 16, pages 279–300. Center for International Forestry Research (CIFOR), Bogor, Indonesia.

Herold, M., Carter, S., Avitabile, V., Espejo, A. B., Jonckheere, I., Lucas, R., McRoberts, R. E., Næsset, E., Nightingale, J., Petersen, R., Reiche, J., Romijn, E., Rosenqvist, A., Rozendaal, D. M. A., Seifert, F. M., Sanz, M. J., and De Sy, V. (2019). The Role and Need for Space-Based Forest Biomass-Related Measurements in Environmental Management and Policy. *Surveys in Geophysics*, 40(4):757–778.

Herold, M. and Skutsch, M. M. (2009). Measurement, reporting and verification for REDD+ : Objectives, capacities and institutions. In Angelsen, A., editor, *Realising REDD+: National strategy and policy options*, chapter 7, pages 84–100. Center for International Forestry Research (CIFOR), Bogor, Indonesia.

Hijmans, R. J. (2016). raster: Geographic Data Analysis and Modeling.

Hollander, M. and Wolfe, D. A. (1973). *Nonparametric statistical methods.* John Wiley & Sons, New York SE.

Hosonuma, N., Herold, M., De Sy, V., De Fries, R. S., Brockhaus, M., Verchot, L., Angelsen, A., and Romijn, E. (2012). An assessment of deforestation and forest degradation drivers in developing countries.

Huynh, T. B. (2014). Cat Loc Landscape – Cat Tien National Park Pro-Poor REDD+ Project, Vietnam. In Sills, E., Atmadja, S., de Sassi, C., Duchelle, A., Kweka, D., Resosudarmo, I., and

Sunderlin, W., editors, *REDD+ on the ground: A case book of subnational initiatives across the globe*, pages 401–416. Center for International Forestry Research (CIFOR), Bogor, Indonesia.

Hyyppä, J., Yu, X., Hyyppä, H., Vastaranta, M., Holopainen, M., Kukko, A., Kaartinen, H., Jaakkola, A., Vaaja, M., Koskinen, J., and Alho, P. (2012). Advances in Forest Inventory Using Airborne Laser Scanning. *Remote Sensing*, 4(5):1190–1207.

Ickowitz, A., Sills, E., and de Sassi, C. (2017). Estimating Smallholder Opportunity Costs of REDD+: A Pantropical Analysis from Households to Carbon and Back. *World Development*, 95:15–26.

Imbens, G. W. and Wooldridge, J. M. (2009). Recent Developments in the Econometrics of Program Evaluation. *Journal of Economic Literature*, 47(1):5–86.

Indonesian Ministry of Forestry (2011). Decree of the Ministry of Forestry of the Republic of Indonesia no. SK.323/Mehut-II/2011 on the Establishment of an Indicative Map of the Moratorium of the Issuance of New Permits for Forest Utilization, Forest Area Utilization, and Changes in Land Use.

Indriatmoko, Y., Atmadja, S., Utomo, N. A., Ekaputri, A. D., and Komalasari, M. (2014). Katingan Peatland Restoration and Conservation Project, Central Kalimantan, Indonesia. In Sills, E. O., Atmadja, S. S., de Sassi, C., Duchelle, A. E., Kweka, D. L., Resosudarmo, I. A. P., and Sunderlin, W. D., editors, *REDD+ on the ground: A case book of subnational initiatives across the globe*, chapter 18, pages 309–328. Center for International Forestry Research (CIFOR), Bogor, Indonesia.

Intarini, D. Y., Resosudarmo, I., Komalasari, M., Ekaputri, A. D., and Agustavia, M. (2014). Ketapang Community Carbon Pools, West Kalimantan, Indonesia. In Sills, E. O., Atmadja, S. S., de Sassi, C., Duchelle, A. E., Kweka, D. L., Resosudarmo, I. A. P., and Sunderlin, W. D., editors, *REDD+ on the ground: A case book of subnational initiatives across the globe*, chapter 19, pages 329–347. Center for International Forestry Research (CIFOR), Bogor, Indonesia.

IPCC (2003). *Good Practice Guidance for Land Use, Land-Use Change and Forestry.* Institute for Global Environmental Strategies (IGES), Hayama, Kanagawa, Japan.

IPCC (2006a). Agriculture, Forestry and Other Land Use. In Eggleston, H. S., Buendia, L., Miwa, K., Ngara, T., and Tanabe, K., editors, *2006 IPCC Guidelines for National Greenhouse Gas Inventories*, volume 4, chapter Vol. 4. IGES, Japan.

IPCC (2006b). Generic methodologies applicable to multiple land-use categories. In Eggleston, S., Buendia, L., Miwa, K., Ngara, T., and Tanabe, K., editors, *IPCC Guidelines for National Greenhouse Gas Inventories - Vol 4 AFOLU*, chapter 2, pages 2.1–2.59. Institute for Global Environmental Strategies (IGES), Hayama, Japan.

IPCC (2014). *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.* IPCC, Geneva, Switzerland.

IPCC (2019). Special Report on Climate Change and Land. Technical report, IPCC.

Jagger, P., Atmadja, S., Pattanayak, S. K., Sills, E., and Sunderlin, W. D. (2009). Learning while doing: Evaluating impacts of REDD+ projects. In Angelsen, A., editor, *Realising REDD+: National strategy and policy options*, chapter 22, pages 281–292. Center for International Forestry Research (CIFOR), Bogor, Indonesia.

Jagger, P., Sills, E., Lawlor, K., and Sunderlin, W. D. (2010). A guide to learning about livelihood impacts of REDD+. Technical report, Center for International Forestry Research (CIFOR), Bogor, Indonesia.

Jarvis, A., Reuter, H. I., Nelson, A., and Guevara, E. (2008). Hole-filled SRTM for the globe Version 4. Technical report, CGIAR-CSI.

Kaimowitz, D. and Angelsen, A. (1998). *Economic models of tropical deforestation: a review*. Center for International Forestry Research (CIFOR).

Khuc, Q. V., Tran, B. Q., Meyfroidt, P., and Paschke, M. W. (2018). Drivers of deforestation and forest degradation in Vietnam: An exploratory analysis at the national level. *Forest Policy and Economics*, 90:128–141.

Korhonen-Kurki, K., Brockhaus, M., Duchelle, A. E., Atmadja, S., and Pham, T. T. (2012). Multiple levels and multiple challenges for REDD+: Lessons from the field. In Angelsen, A., editor, *Analysing REDD+: Challenges and choices*, chapter 6, pages 91–110. Center for International Forestry Research (CIFOR), Bogor, Indonesia.

Korhonen-Kurki, K., Brockhaus, M., Duchelle, A. E., Atmadja, S., Thu Thuy, P., and Schofield, L. (2013). Multiple levels and multiple challenges for measurement, reporting and verification of REDD+. *International Journal of the Commons*, 7(2):344–366.

Krueger, A. B. and Schkade, D. A. (2008). The reliability of subjective well-being measures. *Journal of Public Economics*, 92(8-9):1833–1845.

La Viña, A. G. and de Leon, A. (2014). Two Global Challenges, One Solution: International Cooperation to Combat Climate Change and Tropical Deforestation. Technical report, Center for Global Development, Washington, DC.

Lambin, E. F., Geist, H. J., and Lepers, E. (2003). Dynamics of Land-use and Land-cover Change in Tropical Regions. *Annual Review of Environment and Resources*, 28(1):205–241.

Larson, A. M., Solis, D., Duchelle, A. E., Atmadja, S., Resosudarmo, I. A. P., Dokken, T., and Komalasari, M. (2018). Gender lessons for climate initiatives: A comparative study of REDD+ impacts on subjective wellbeing. *World Development*, 108:86–102.

Liaw, A. and Wiener, M. (2018). Breiman and Cutler's Random Forests for Classification and Regression. Package 'randomForest'. Technical report.

Lin, L. (2012). *Geography of REDD+ at Multiple Scales: Country Participation and Project Location*. PhD thesis, North Carolina State University.

Loft, L., Le, D. N., Pham, T. T., Yang, A. L., Tjajadi, J. S., and Wong, G. Y. (2017). Whose Equity Matters? National to Local Equity Perceptions in Vietnam's Payments for Forest Ecosystem Services Scheme. *Ecological Economics*, 135:164–175.

Luttrell, C., Sills, E., Aryani, R., Ekaputri, A. D., and Evinke, M. F. (2018). Beyond opportunity costs: who bears the implementation costs of reducing emissions from deforestation and degradation? *Mitigation and Adaptation Strategies for Global Change*, 23(2):291–310.

Melo, J. B., Ziv, G., Baker, T. R., Carreiras, J. M. B., Pearson, T. R. H., and Vasconcelos, M. J. (2018). Striking divergences in Earth Observation products may limit their use for REDD+. *Environmental Research Letters*, 13(10):104020.

Milbank, C., Coomes, D., and Vira, B. (2018). Assessing the Progress of REDD+ Projects towards the Sustainable Development Goals. *Forests*, 9(10):589.

Milodowski, D. T., Mitchard, E. T. A., and Williams, M. (2017). Forest loss maps from regional satellite monitoring systematically underestimate deforestation in two rapidly changing parts of the Amazon. *Environmental Research Letters*, 12(9):094003.

Miteva, D. A., Pattanayak, S. K., and Ferraro, P. J. (2012). Evaluation of biodiversity policy instruments: What works and what doesn't? *Oxford Review of Economic Policy*, 28(1):69–92.

Molijn, C. (2019). De Amazone brandt - maar hoe fel? [The Amazon is burning - but how fierce?].

Müller, R., Pistorius, T., Rohde, S., Gerold, G., and Pacheco, P. (2013). Policy options to reduce deforestation based on a systematic analysis of drivers and agents in lowland Bolivia. *Land Use Policy*, 30(1):895–907.

Nepstad, D. C., Verssimo, A., Alencar, A., Nobre, C., Lima, E. S., Lefebvre, P., Schlesinger, P., Potter, C., Moutinho, P., Mendoza, E., Cochrane, M., and Brooks, V. (1999). Large-scale impoverishment of Amazonian forests by logging and fire. *Nature*, 398(6727):505–508.

Norman, M. and Nakhooda, S. (2015). The State of REDD+ Finance. *SSRN Electronic Journal*.

OECD (2013). *OECD Guidelines on Measuring Subjective Well-being*. OECD Publishing, Paris.

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. 3(2).

Olofsson, P., Foody, G. M., Herold, M., Stehman, S. V., Woodcock, C. E., and Wulder, M. A. (2014). Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment*, 148:42–57.

Olofsson, P., Foody, G. M., Stehman, S. V., and Woodcock, C. E. (2013). Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation. *Remote Sensing of Environment*, 129:122–131.

Open Foris (2019). Open Foris - Free open-source solutions for environmental monitoring.

Paddock, R. C., Suhartono, M., and Ifansasti, U. (2019). A Blood-Red Sky: Fires Leave a Million Indonesians Gasping.

Pauw, W., Cassanmagnano, D., Mbeva, K., Hein, J., Guarin, A., Brandi, C., Dzebo, A., Canales, N., Adams, K., Atteridge, A., Bock, T., Helms, J., Zalewski, A., E., F., Lindener, A., and Muhammad, D. (2016). NDC Explorer.

Peffer, K. (2015). *Blood donation and cardiovascular disease - Addressing the healthy donor effect*. PhD thesis, Radboud Universiteit Nijmegen.

Petersen, R., Davis, C., Herold, M., and De Sy, V. (2018). Ending Tropical Deforestation: Tropical Forest Monitoring: Exploring the Gaps Between What is Required and What is Possible for REDD+ and Other Initiatives. Technical report, World Resources Institute, Washington DC.

Poudyal, M., Jones, J. P., Rakotonarivo, O. S., Hockley, N., Gibbons, J. M., Mandimbiniaina, R., Rasoamanana, A., Andrianantenaina, N. S., and Ramamonjisoa, B. S. (2018). Who bears the cost of forest conservation? *PeerJ*, 6:e5106.

Poudyal, M., Ramamonjisoa, B. S., Hockley, N., Rakotonarivo, O. S., Gibbons, J. M., Mandimbiniaina, R., Rasoamanana, A., and Jones, J. P. G. (2016). Can REDD plus social safeguards reach the 'right' people? Lessons from Madagascar. *Global Environmental Change - Human Policy Dimensions*, 37:31–42.

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 Biology*, 6(7):1368–1369.

Rakatama, A., Pandit, R., Ma, C., and Iftekhar, S. (2017). The costs and benefits of REDD + : A review of the literature. *Forest Policy and Economics*, 75:103–111.

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. *International Journal of the Commons*, 9(2):909.

Reiche, J., Lucas, R., Mitchell, A. L., Verbesselt, J., Hoekman, D. H., Haarpaintner, J., Kellndorfer, J. M., Rosenqvist, A., Lehmann, E. A., Woodcock, C. E., Seifert, F. M., and Herold, M. (2016). Combining satellite data for better tropical forest monitoring.

Rissman, A. R. and Smail, R. (2015). Accounting for results: how conservation organizations report performance information. *Environmental management*, 55(4):916–29.

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 Ecology and Management*, 352:109–123.

Rudel, T. K. (2007). Changing agents of deforestation: From state-initiated to enterprise driven processes, 1970-2000. *Land Use Policy*.

Rudel, T. K., Defries, R., Asner, G. P., and Laurance, W. F. (2009). Changing drivers of deforestation and new opportunities for conservation. *Conservation Biology*.

Saatchi, S. S., Harris, N. L., Brown, S., Lefsky, M., Mitchard, E. T. A., Salas, W., Zutta, B. R., Buermann, W., Lewis, S. L., Hagen, S., Petrova, S., White, L., Silman, M., and Morel, A. (2011). Benchmark map of forest carbon stocks in tropical regions across three continents. *Proceedings of the National Academy of Sciences of the United States of America*, 108(24):9899–904.

Seymour, F. and Angelsen, A. (2009). Summary and conclusions: REDD wine in old wineskins? In Angelsen, A., editor, *Realising REDD+: National strategy and policy options*, chapter 23, pages 293–304. Center for International Forestry Research (CIFOR), Bogor, Indonesia.

Seymour, F. and Busch, J. (2016). *Why Forests? Why Now?: The Science, Economics, and Politics of Tropical Forests and Climate Change.* Center for Global Development (CGD), Washington, DC.

Seymour, F. and Harris, N. L. (2019). Reducing tropical deforestation. *Science*, 365(6455):756–757.

Sills, E., Atmadja, S., de Sassi, C., Duchelle, A., Kweka, D., Resosudarmo, I., and Sunderlin, W. (2014). *REDD+ on the ground: A case book of subnational initiatives across the globe.* Center for International Forestry Research (CIFOR), Bogor, Indonesia.

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. *Global Environmental Change*, 43:148–160.

Simonet, G., Bos, A. B., Duchelle, A. E., Pradnja Resosudarmo, I. A., Subervie, J., and Wunder, S. (2018). Forests and carbon : The impacts of local REDD+ initiatives. In Angelsen, A., Martius, C., De Sy, V., Duchelle, A. E., Larson, A. M., and Pham, T. T., editors, *Transforming REDD+*, chapter 10, pages 117–130. Center for International Forestry Research (CIFOR), Bogor, Indonesia.

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. Technical Report 32, Climate Economics Chair, Paris, France.

Simonet, G., Subervie, J., Ezzine-de Blas, D., Cromberg, M., and Duchelle, A. E. (2019). Effectiveness of a REDD+ Project in Reducing Deforestation in the Brazilian Amazon. *American Journal of Agricultural Economics*, 101(1):211–229.

Sims, K. R. and Alix-Garcia, J. M. (2017). Parks versus PES: Evaluating direct and incentive-based land conservation in Mexico. *Journal of Environmental Economics and Management*, 86:8–28.

Smith, E. P. (2002). BACI design. In El-Shaarawi, A. H. and Piegorsch, W. W., editors, *Encyclopedia of Environmetrics*, pages 141–148. John Wiley & Sons, Ltd., Chichester, vol. 1 edition.

Soares-Filho, B. S., Nepstad, D. C., Curran, L. M., Cerqueira, G. C., Garcia, R. A., Ramos, C. A., Voll, E., McDonald, A., Lefebvre, P., and Schlesinger, P. (2006). Modelling conservation in the

Amazon basin. *Nature*, 440(7083):520–523.

Solsvik, T. (2019). Better satellite images to help save rainforests, Norway says.

Stehman, S. V. (2014). Estimating area and map accuracy for stratified random sampling when the strata are different from the map classes. *International Journal of Remote Sensing*, 35(13):4923–4939.

Sunderlin, W., de Sassi, C., Ekaputri, A., Light, M., and Pratama, C. (2017). REDD+ Contribution to Well-Being and Income Is Marginal: The Perspective of Local Stakeholders. *Forests*, 8(4):125.

Sunderlin, W., Ekaputri, A. D., Sills, E., Duchelle, A. E., Kweka, D., Diprose, R., Doggart, N., Ball, S., Lima, R., Enright, A., Torres, J., Hartanto, H., and Toniolo, A. (2014a). *The challenge of establishing REDD+ on the ground: Insights from 23 subnational initiatives in six countries.* Center for International Forestry Research (CIFOR), Bogor, Indonesia.

Sunderlin, W., Larson, A., Duchelle, A., Sills, E., Luttrell, C., Jagger, P., Pattanayak, S., Cronkleton, P., A.D., E., de Sassi, C., Aryani, R., and G. Simonet (2016). *Technical guidelines for research on REDD+ subnational initiatives.* Center for International Forestry Research (CIFOR), Bogor, Indonesia, 2nd edition.

Sunderlin, W., Pratama, C., Bos, A., Avitabile, V., Sills, E., de Sassi, C., Joseph, S., Agustavia, M., Pribadi, U., and Anandadas, A. (2014b). REDD+ on the ground: The need for scientific evidence. In Sills, E., Atmadja, S., de Sassi, C., Duchelle, A., Kweka, D., Resosudarmo, I., and Sunderlin, W., editors, *REDD+ on the ground: A case book of subnational initiatives across the globe*, chapter 1, pages 2–21. Center for International Forestry Research (CIFOR), Bogor, Indonesia.

Sunderlin, W. D., de Sassi, C., Sills, E. O., Duchelle, A. E., Larson, A. M., Resosudarmo, I. A. P., Awono, A., Kweka, D. L., and Huynh, T. B. (2018). Creating an appropriate tenure foundation for REDD+: The record to date and prospects for the future. *World Development*, 106:376–392.

Sunderlin, W. D., Larson, A. M., Duchelle, A., Sills, E. O., Luttrell, C., Jagger, P., Pattanayak, S., Cronkleton, P., Ekaputri, A. D., de Sassi, C., and Aryani, R. (2010). Technical guidelines for research on REDD+ project sites. Technical report, Center for International Forestry Research, Bogor, Indonesia.

Sunderlin, W. D., Larson, A. M., Duchelle, A. E., Resosudarmo, I. A. P., Huynh, T. B., Awono, A., and Dokken, T. (2014c). How are REDD+ Proponents Addressing Tenure Problems? Evidence from Brazil, Cameroon, Tanzania, Indonesia, and Vietnam. *World Development*, 55:37–52.

Sunderlin, W. D., Sills, E., Duchelle, A., Ekaputri, A., Kweka, D., Toniolo, M., Ball, S., Doggart, N., Pratama, C., Padilla, J., Enright, A., and Otsyina, R. (2015). REDD+ at a critical juncture: assessing the limits of polycentric governance for achieving climate change mitigation. *International Forestry Review*, 17(4):400–413.

Sunderlin, W. D. and Sills, E. O. (2012). REDD+ projects as a hybrid of old and new forest conservation approaches. In *Analysing REDD+: Challenges and choices*.

Thaler, G. M. and Anandi, C. A. M. (2017). Shifting cultivation, contentious land change and forest governance: the politics of swidden in East Kalimantan. *The Journal of Peasant Studies*, 44(5):1066–1087.

Turubanova, S., Potapov, P. V., Tyukavina, A., and Hansen, M. C. (2018). Ongoing primary forest loss in Brazil, Democratic Republic of the Congo, and Indonesia. *Environmental Research Letters*, 13(7):074028.

Ullum, H., Rostgaard, K., Kamper-Jørgensen, M., Reilly, M., Melbye, M., Nyrén, O., Norda, R., Edgren, G., and Hjalgrim, H. (2015). Blood donation and blood donor mortality after adjustment for a healthy donor effect. *Transfusion*.

UNFCCC (2006). Good practice guidance and adjustments under Article 5, paragraph 2, of the Kyoto Protocol.

UNFCCC (2009). Cost of implementing methodologies and monitoring systems relating to estimates of emissions from deforestation and forest degradation, the assessment of carbon stocks and greenhouse gas emissions from changes in forest cover, and the enhancement of fores. Technical Report Technical Paper FCCC/TP/2009/1, Bonn, Germany.

UNFCCC (2011). The Cancun Agreements: Outcome of the work of the Ad Hoc Working Group on Long-term Cooperation Under the Convention. Decision 1/CP.16. Report of the Conference of the Parties on its Sixteenth Session. FCC/CP/2010/7 Add.1.

UNFCCC (2015). Adoption of the Paris Agreement. *Conference of the Parties on its twenty-first session*, 21932(December):32.

UNFCCC (2019). REDD+ Web platform - Submissions.

Vatn, A. and Angelsen, A. (2009). Options for a national REDD+ architecture. In Angelsen, A., editor, *Realising REDD+: National Strategy and Policy Options*, chapter 5, pages 57–74.

VCS (2015). Verra Project Database.

Verbesselt, J., Hyndman, R., Newnham, G., and Culvenor, D. (2010). Detecting trend and seasonal changes in satellite image time series. *Remote Sensing of Environment*, 114(1):106–115.

Verbesselt, J., Zeileis, A., and Herold, M. (2012). Near real-time disturbance detection using satellite image time series. *Remote Sensing of Environment*, 123:98–108.

Verchot, L. V., Anitha, K., Romijn, E., Herold, M., and Hergoualc'h, K. (2012). Emissions factors: Converting land use change to CO2 estimates. In Angelsen, A., editor, *Analysing REDD+: Challenges and choices*, chapter 15, pages 261–278. Center for International Forestry Research (CIFOR), Bogor, Indonesia.

Visseren-Hamakers, I. J., Gupta, A., Herold, M., Peña-Claros, M., and Vijge, M. J. (2012). Will REDD+ work? The need for interdisciplinary research to address key challenges. *Current Opinion in Environmental Sustainability*, 4(6):590–596.

Walker, R., Perz, S., Caldas, M., and Silva, L. G. T. (2002). Land Use and Land Cover Change in Forest Frontiers: The Role of Household Life Cycles. *International Regional Science Review*, 25(2):169–199.

Wertz-Kanounnikoff, S. and Angelsen, A. (2009). Global and national REDD+ architecture: linking institutions and actions. In Angelsen, A., editor, *Realising REDD+: National strategy and policy options*, chapter 2, pages 13–14. Center for International Forestry Research (CIFOR), Bogor, Indonesia.

Wertz-Kanounnikoff, S. and McNeill, D. (2012). Performance indicators and REDD+ implementation. In Angelsen, A., editor, *Analysing REDD+: Challenges and choices*, chapter 13, pages 233–246. Center for International Forestry Research (CIFOR), Bogor, Indonesia.

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? In Angelsen, A., editor, *Moving Ahead with REDD: Issues, Options and Implications*, chapter 9, pages 87–98. Center for International Forestry Research (CIFOR), Bogor, Indonesia.

Wong, G., Angelsen, A., Brockhaus, M., Carmenta, R., Duchelle, A., Leonard, S., Luttrell, C., Martius, C., and Wunder, S. (2016). Results-based payments for REDD+: Lessons on finance, performance, and non-carbon benefits. *CIFOR Infobrief*, 138.

Wood, C. H. and Porro, R. (2002). *Deforestation and land use in the Amazon*. University Press of Florida, Gainesville.

Woodcock, C. E., Allen, R., Anderson, M., Belward, A., Bindschadler, R., Cohen, W., Gao, F., Goward, S. N., Helder, D., Helmer, E., Nemani, R., Oreopoulos, L., Schott, J., Thenkabail, P. S., Vermote, E. F., Vogelmann, J., Wulder, M. A., and Wynne, R. (2008). Free access to landsat imagery.

Zanella, L., Folkard, A. M., Blackburn, G. A., and Carvalho, L. M. T. (2017). How well does random forest analysis model deforestation and forest fragmentation in the Brazilian Atlantic forest? *Environmental and Ecological Statistics*, 24(4):529–549.

Appendices

Appendix A Supplementary material for chapter 2

A.1 Village boundary delineation

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

A.2 General results extended

Table A.2 (page 139) shows an extended version of the summary statistics in section 2.3.1.

A.3 BA and BACI classified scores for intensive sites only

Figure A.3 (page 140) reports results at both the meso and micro level for the 16 'intensive' sites only, which as described in section 2.2.4 include both intervention and matched control villages. These results are mostly consistent with the results presented in figure 2.4, confirming our finding (presented in section 2.3.1) 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 A.3 confirms that this finding is not due to the difference in sample size for the meso and micro level analysis reported in figure 2.4.

A.4 Test results for bias detection

Table A.4 (page 141) shows the results of the Levene's and T-Tests to test for bias in the *before* period (see section 2.3.2).

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 after period	23	2	9	6	6
both	n_b	years in before period	23	5	12	8	8
meso	$\bar{x}_{AI}$	Average annual deforestation rate in intervention area during after period	23	0.037	1.840	0.522	0.430
meso	$\bar{x}_{BI}$	Average annual deforestation rate in intervention area during before period	23	0.021	1.620	0.479	0.370
meso	$\bar{x}_{AC}$	Average annual deforestation rate in control area during after period	23	0.065	1.930	0.664	0.605
meso	$\bar{x}_{BC}$	Average annual deforestation rate in control area during before period	23	0.048	1.620	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 after period	16	0.073	3.933	0.928	0.605
micro	$\bar{x}_{BI}$	Average annual deforestation rate in intervention area during before period	16	0.068	4.514	1.199	0.489
micro	$\bar{x}_{AC}$	Average annual deforestation rate in control area during after period	16	0.106	2.479	1.023	0.862
micro	$\bar{x}_{BC}$	Average annual deforestation rate in control area during before 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

Table A.2: General results table extended.

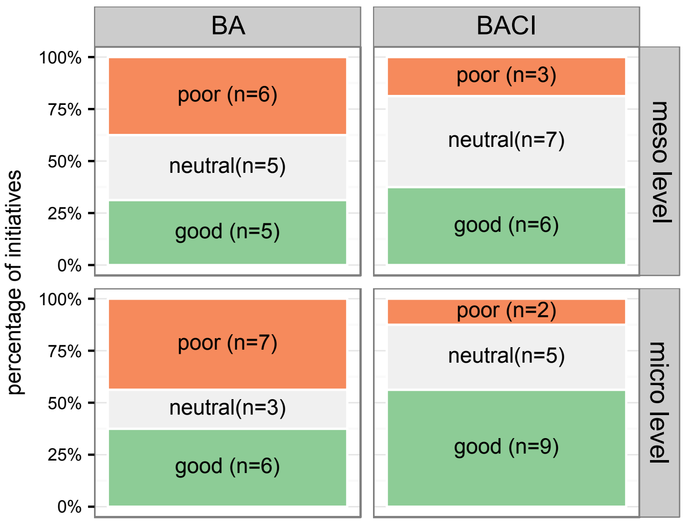


Figure A.3: BA and BACI classified scores with equal sample sizes for both levels.

Table A.4: 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			
	p-value Levene's test	*p*-value two sample *t*-test[a]	*p*-value Welch *t*-test[b]	Possible bias[c]	*p*-value Levene's test	*p*-value two sample *t*-test[a]	*p*-value Welch *t*-test[b]	Possible bias[c]
Brazil-Acre	0.4413	0.8487	N/A	FALSE	0.1062	0.1359	N/A	FALSE
Brazil-Cotriguacu	0.7500	0.4233	N/A	FALSE	0.5460	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	0.0100	N/A	0.0001	TRUE	0.2856	0.0267	N/A	TRUE
Peru-Ucayali	0.0001	N/A	0.0004	TRUE	0.4320	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	0.6630	N/A	FALSE	0.5390	0.5952	N/A	FALSE
Vietnam-Cat Tien	0.8567	0.8992	N/A	FALSE	0.0740	0.2737	N/A	FALSE

[a] Equal variances assumed. [b] Unequal variances assumed. [c] Using confidence level of 0.95.

Appendix B Supplementary material for chapter 3

B.1 Decision trees for reclassification strategies

Figure B.1 (page 143) shows the decision trees that were used to define the reclassification values.

After applying an overlay of the two map products, every pixel from the sample was reclassified according to the different reclassification strategies. The squares represent the possible pixel-level combinations of the two datasets. The rightmost column shows the decision in each of the reclassified map products.

B.2 Accuracies with 95%CI for all original and reclassified products

Table B.2 (page 144) shows all accuracies (UA, PA, OA) for all original and reclassified products.

B.3 Paired T-tests

Table B.3 (page 145) shows the results of the paired T-tests, to test if the accuracies of the original and reclassified products differ significantly.

B.4 Relative Accuracy Changes Per Site

Figure B.4 (pages 146-147) shows the Relative Accuracy Changes of both map products for each individual site.

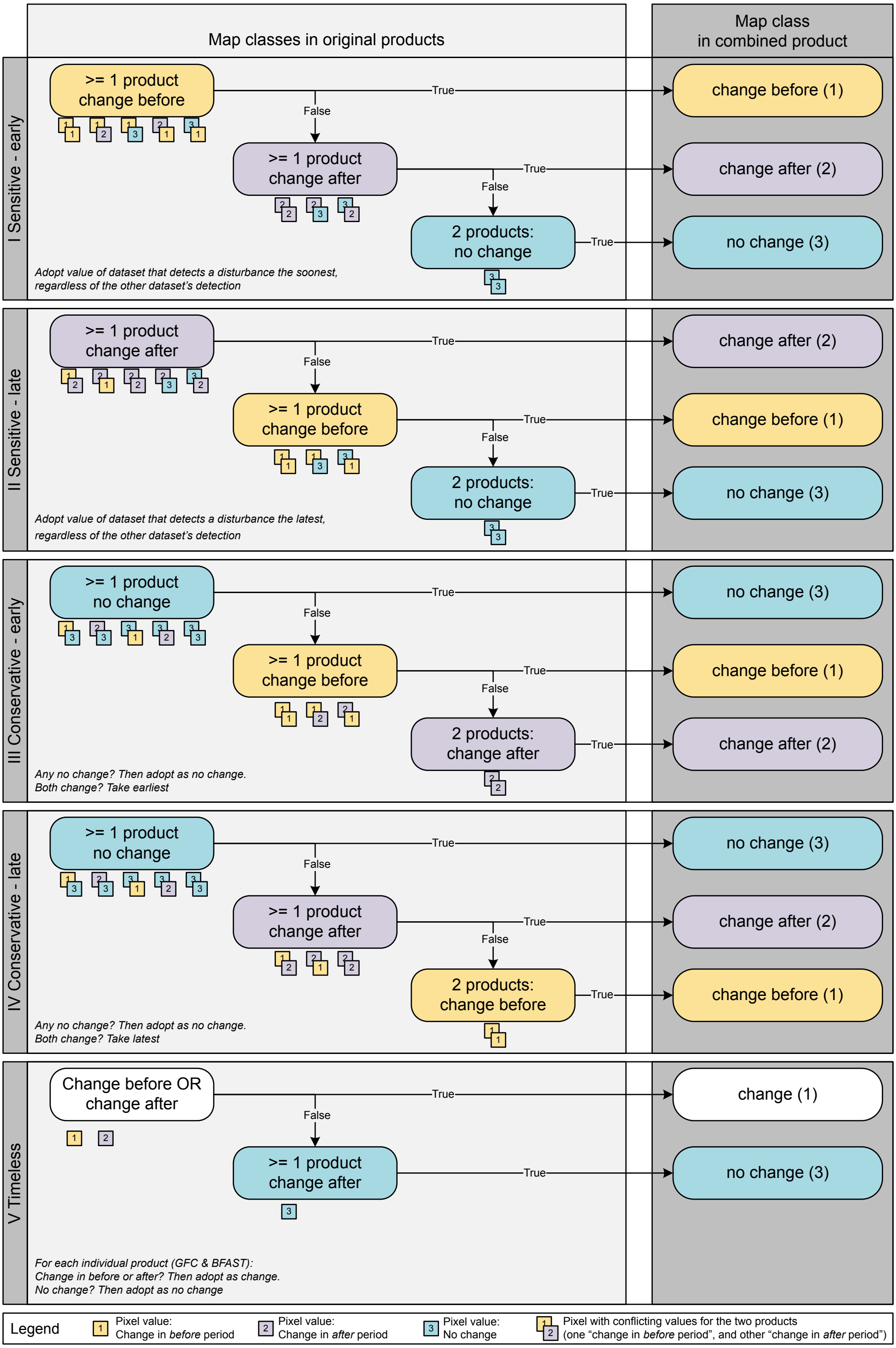


Figure B.1: Decision trees for reclassification strategies

Table B.2: Accuracies with 95%CI for all original and reclassified products

Region	Product	Strata	UA	PA	OA
Peru	GFC	Change before	0.91±0.08	0.40±0.26	0.95±0.04
		Change after	0.59±0.11	0.73±0.13	
		No change	0.96±0.04	0.99±0.00	
	BFAST	Change before	0.86±0.08	0.49±0.32	0.95±0.04
		Change after	0.70±0.12	0.54±0.13	
		No change	0.96±0.04	0.99±0.00	
	I	Change before	0.84±0.07	0.64±0.40	0.96±0.04
		Change after	0.67±0.11	0.94±0.06	
		No change	0.98±0.04	0.99±0.00	
	II	Change before	0.84±0.08	0.57±0.36	0.96±0.04
		Change after	0.58±0.10	0.97±0.06	
		No change	0.98±0.04	0.99±0.00	
	III	Change before	0.97±0.02	0.33±0.21	0.94±0.04
		Change after	0.89±0.09	0.30±0.05	
		No change	0.94±0.04	1.00±0.00	
	IV	Change before	1.00±0.00	0.26±0.16	0.94±0.04
		Change after	0.54±0.07	0.33±0.06	
		No change	0.94±0.04	1.00±0.00	
	V-GFC	Change	0.86±0.06	0.54±0.27	0.95±0.04
		No change	0.96±0.04	0.99±0.00	
	V-BFAST	Change	0.89±0.06	0.55±0.27	0.96±0.04
		No change	0.96±0.04	0.99±0.00	
Tanzania	GFC	Change before	0.41±0.15	0.05±0.05	0.83±0.08
		Change after	0.26±0.15	0.10±0.08	
		No change	0.86±0.08	0.97±0.01	
	BFAST	Change before	0.90±0.10	0.14±0.12	0.89±0.08
		Change after	0.71±0.15	0.50±0.32	
		No change	0.90±0.08	0.99±0.01	
	I	Change before	0.67±0.11	0.18±0.15	0.87±0.08
		Change after	0.54±0.14	0.54±0.34	
		No change	0.91±0.08	0.96±0.01	
	II	Change before	0.67±0.13	0.15±0.13	0.87±0.08
		Change after	0.53±0.14	0.54±0.34	
		No change	0.91±0.08	0.96±0.01	
	III	Change before	0.77±0.10	0.04±0.03	0.85±0.08
		Change after	0.96±0.06	0.06±0.04	
		No change	0.85±0.08	1.00±0.00	
	IV	Change before	1.00±0.00	0.01±0.01	0.85±0.08
		Change after	0.71±0.06	0.06±0.04	
		No change	0.85±0.08	1.00±0.00	
	V-GFC	Change	0.42±0.16	0.11±0.07	0.84±0.08
		No change	0.86±0.08	0.97±0.01	
	V-BFAST	Change	0.91±0.08	0.43±0.22	0.90±0.08
		No change	0.90±0.08	0.99±0.01	
Vietnam	GFC	Change before	0.99±0.03	0.29±0.08	0.80±0.06
		Change after	0.68±0.06	0.43±0.16	
		No change	0.81±0.08	1.00±0.00	
	BFAST	Change before	0.99±0.02	0.73±0.17	0.89±0.06
		Change after	0.88±0.07	0.59±0.21	
		No change	0.88±0.08	1.00±0.01	
	I	Change before	0.98±0.02	0.80±0.18	0.92±0.06
		Change after	0.89±0.07	0.68±0.24	
		No change	0.91±0.08	1.00±0.01	
	II	Change before	0.99±0.03	0.62±0.14	0.89±0.06
		Change after	0.73±0.06	0.68±0.24	
		No change	0.91±0.08	1.00±0.01	
	III	Change before	0.99±0.02	0.40±0.09	0.80±0.06
		Change after	0.93±0.07	0.33±0.12	
		No change	0.78±0.07	1.00±0.00	
	IV	Change before	1.00±0.00	0.22±0.05	0.78±0.06
		Change after	0.64±0.05	0.34±0.12	
		No change	0.78±0.07	1.00±0.00	
	V-GFC	Change	1.00±0.00	0.47±0.10	0.84±0.06
		No change	0.81±0.08	1.00±0.00	
	V-BFAST	Change	0.99±0.02	0.69±0.15	0.90±0.06
		No change	0.88±0.08	1.00±0.01	
Indonesia-A	GFC	Change before	0.85±0.12	0.26±0.11	0.85±0.06
		Change after	0.79±0.07	0.79±0.19	
		No change	0.87±0.08	0.99±0.01	
	BFAST	Change before	0.80±0.10	0.43±0.17	0.83±0.06
		Change after	0.76±0.09	0.58±0.15	
		No change	0.84±0.08	0.96±0.02	
	I	Change before	0.78±0.09	0.56±0.21	0.88±0.06
		Change after	0.76±0.08	0.84±0.20	
		No change	0.91±0.08	0.95±0.02	
	II	Change before	0.77±0.11	0.45±0.17	0.86±0.06
		Change after	0.70±0.07	0.86±0.20	
		No change	0.91±0.08	0.95±0.02	
	III	Change before	0.91±0.06	0.23±0.09	0.82±0.06
		Change after	0.96±0.06	0.52±0.13	
		No change	0.80±0.07	1.00±0.00	
	IV	Change before	1.00±0.00	0.13±0.05	0.81±0.06
		Change after	0.80±0.05	0.54±0.13	
		No change	0.80±0.07	1.00±0.00	
	V-GFC	Change	0.96±0.04	0.62±0.14	0.88±0.06
		No change	0.87±0.08	0.99±0.01	
	V-BFAST	Change	0.84±0.06	0.55±0.13	0.84±0.06
		No change	0.84±0.08	0.96±0.02	
Indonesia-B	GFC	Change before	0.97±0.05	0.60±0.08	0.94±0.01
		Change after	0.74±0.08	0.84±0.09	
		No change	0.96±0.01	0.99±0.01	
	BFAST	Change before	0.79±0.08	0.81±0.08	0.93±0.01
		Change after	0.92±0.08	0.50±0.09	
		No change	0.94±0.01	0.97±0.01	
	I	Change before	0.81±0.08	0.98±0.03	0.97±0.01
		Change after	0.86±0.08	0.98±0.02	
		No change	1.00±0.00	0.96±0.01	
	II	Change before	0.80±0.08	0.87±0.03	0.96±0.01
		Change after	0.76±0.07	1.00±0.00	
		No change	1.00±0.00	0.96±0.01	
	III	Change before	0.98±0.02	0.54±0.05	0.91±0.01
		Change after	0.96±0.06	0.34±0.03	
		No change	0.91±0.01	1.00±0.00	
	IV	Change before	1.00±0.00	0.43±0.04	0.90±0.01
		Change after	0.66±0.05	0.36±0.04	
		No change	0.91±0.01	1.00±0.00	
	V-GFC	Change	0.94±0.04	0.77±0.06	0.95±0.01
		No change	0.96±0.01	0.99±0.01	
	V-BFAST	Change	0.84±0.06	0.70±0.06	0.93±0.01
		No change	0.94±0.01	0.97±0.01	

Table B.3: Paired T-Tests

OVERALL ACCURACY

(reclassified) product	mean	median	sd	Relative to GFC Mean Δ	Relative to GFC p-value	Relative to BFAST Mean Δ	Relative to BFAST p-value
GFC	0.88	0.85	0.07	NA	NA	−0.02	0.82
BFAST	0.90	0.89	0.05	0.02	0.18	NA	NA
I	0.92	0.92	0.05	0.04**	0.04	0.02*	0.08
II	0.91	0.89	0.05	0.03**	0.05	0.01	0.19
III	0.87	0.85	0.06	−0.01	0.81	−0.03	0.95
IV	0.86	0.85	0.07	−0.02	0.92	−0.04	0.96
V – GFC	0.91	0.90	0.04	0.03	0.13	0.01***	0.01
V – BFAST	0.89	0.88	0.06	0.02**	0.02	−0.01	0.59

USER'S ACCURACY

(reclassified) product	BEFORE mean	BEFORE median	BEFORE sd	BEFORE Relative to GFC Mean Δ	BEFORE Relative to BFAST Mean Δ	AFTER mean	AFTER median	AFTER sd	AFTER Relative to GFC Mean Δ	AFTER Relative to BFAST Mean Δ	NO CHANGE mean	NO CHANGE median	NO CHANGE sd	NO CHANGE Relative to GFC Mean Δ	NO CHANGE Relative to BFAST Mean Δ
GFC	0.82	0.91	0.24	NA	−0.04	0.61	0.68	0.21	NA	−0.18	0.89	0.87	0.07	NA	−0.02
BFAST	0.87	0.86	0.08	0.04	NA	0.79	0.76	0.10	0.18**	NA	0.91	0.90	0.05	0.02	NA
I	0.82	0.81	0.11	−0.01	−0.05	0.74	0.76	0.14	0.13**	−0.05	0.94	0.91	0.04	0.05***	0.04**
II	0.81	0.80	0.12	−0.01	−0.06	0.66	0.70	0.10	0.05	−0.14	0.94	0.91	0.04	0.05***	0.04**
III	0.92	0.97	0.09	0.10	0.06	0.94	0.96	0.03	0.33**	0.14**	0.86	0.85	0.07	−0.03	−0.05
IV	1.00	1.00	0.00	0.18*	0.13**	0.67	0.66	0.10	0.06	−0.12	0.86	0.85	0.07	−0.03	−0.05
V - GFC	0.83	0.94	0.24	0.01	−0.03	NA	NA	NA	NA	NA	0.89	0.87	0.07	NA	−0.02
V - BFAST	0.89	0.89	0.06	0.07	0.03**	NA	NA	NA	NA	NA	0.91	0.90	0.05	0.02	NA

PRODUCER'S ACCURACY

(reclassified) product	BEFORE mean	BEFORE median	BEFORE sd	BEFORE Relative to GFC Mean Δ	BEFORE Relative to BFAST Mean Δ	AFTER mean	AFTER median	AFTER sd	AFTER Relative to GFC Mean Δ	AFTER Relative to BFAST Mean Δ	NO CHANGE mean	NO CHANGE median	NO CHANGE sd	NO CHANGE Relative to GFC Mean Δ	NO CHANGE Relative to BFAST Mean Δ
GFC	0.32	0.29	0.20	NA	−0.20	0.58	0.73	0.31	NA	0.03	0.99	0.99	0.01	NA	0.01
BFAST	0.52	0.49	0.26	0.20**	0.00	0.54	0.54	0.05	−0.03	NA	0.98	0.99	0.02	−0.01	NA
I	0.63	0.64	0.30	0.31***	0.11***	0.79	0.84	0.18	0.22**	0.25**	0.97	0.96	0.02	−0.02	−0.01
II	0.53	0.57	0.26	0.21***	0.01	0.81	0.86	0.19	0.24***	0.27**	0.97	0.96	0.02	−0.02	−0.01
III	0.31	0.33	0.19	−0.01	−0.21	0.31	0.33	0.17	−0.27	−0.24	1.00	1.00	0.00	0.01**	0.02**
IV	0.21	0.22	0.15	−0.11	−0.31	0.33	0.34	0.17	−0.25	−0.22	1.00	1.00	0.00	0.01**	0.02**
V - GFC	0.50	0.54	0.24	0.18**	−0.02	NA	NA	NA	NA	NA	0.99	0.99	0.01	NA	0.01
V - BFAST	0.58	0.55	0.11	0.26***	0.07	NA	NA	NA	NA	NA	0.98	0.99	0.02	−0.01	NA

* significance level 0.90.
** significance level 0.95.
*** significance level 0.99.

H0: means do not differ.
Ha: means of reclassification product > input product.
N = 5 (sites).

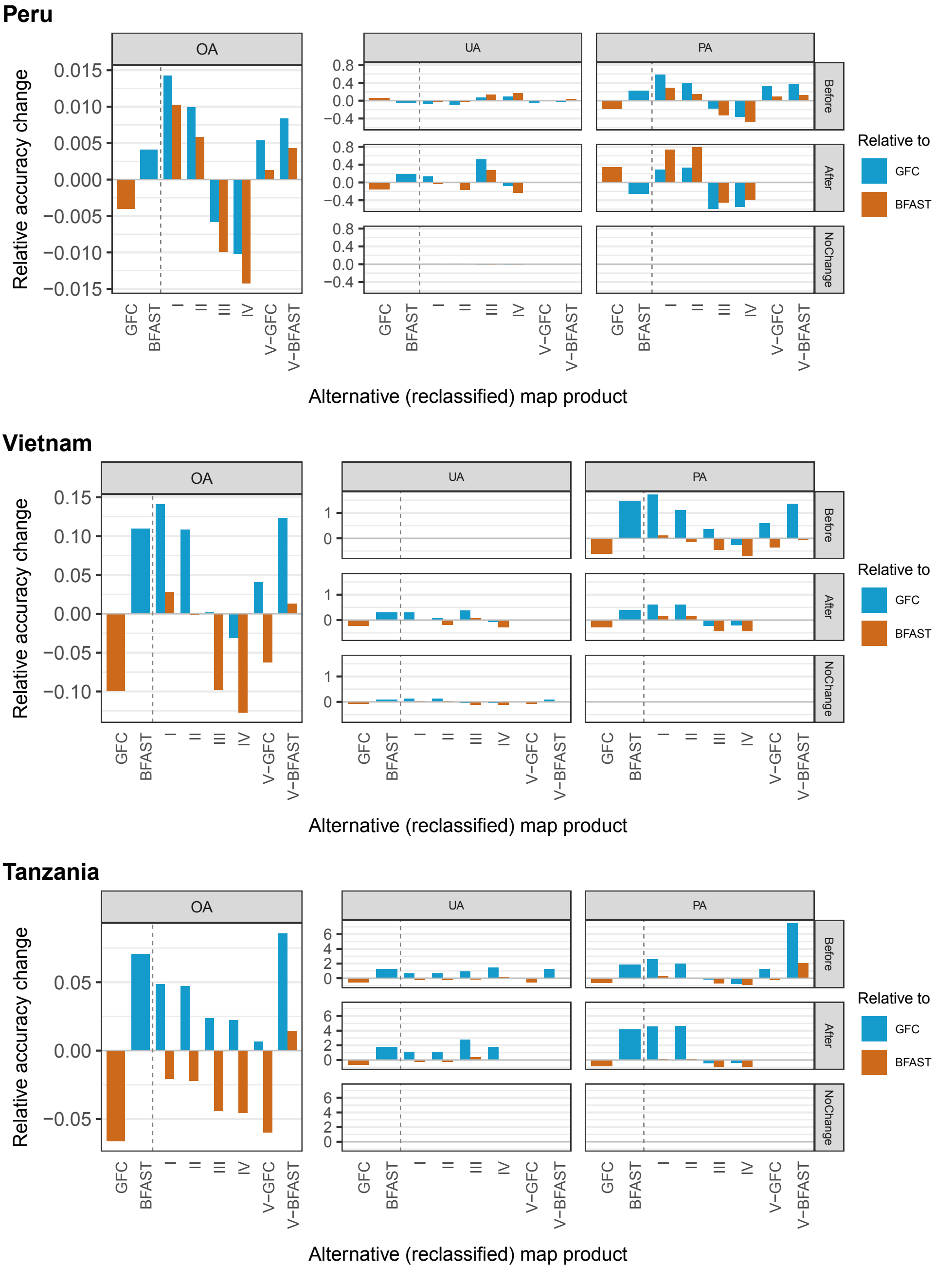


Figure B.4: Relative Accuracy Changes Per Site

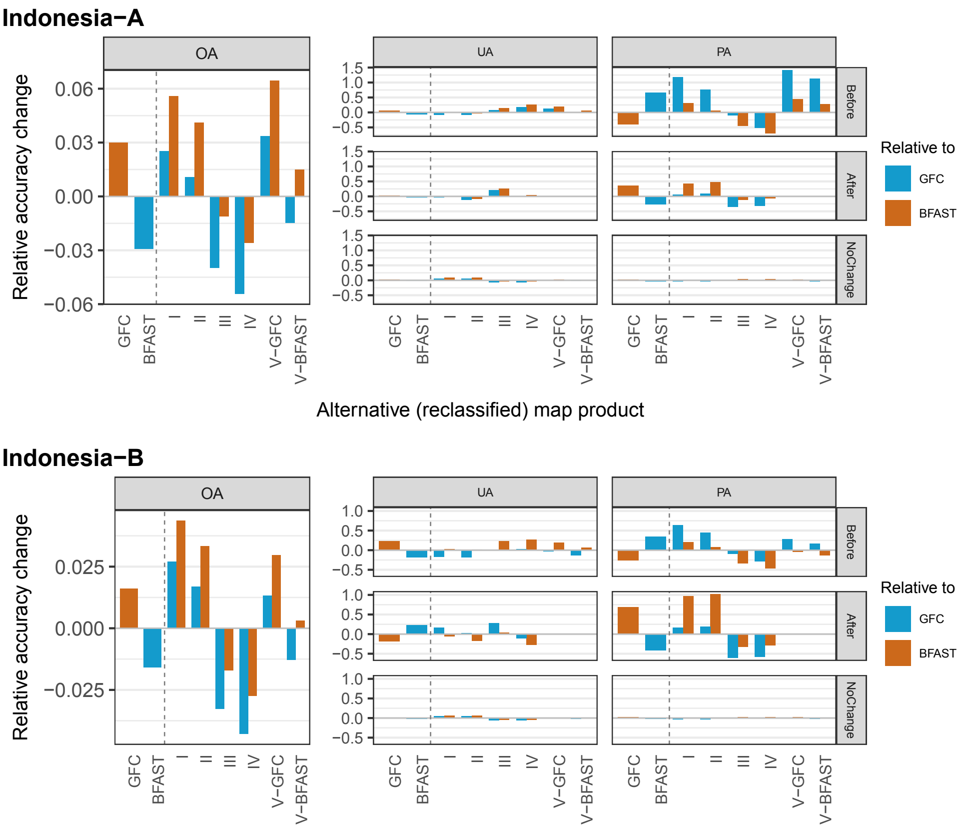


Figure B.4: Relative Accuracy Changes Per Site (continued)

Appendix C Supplementary material for chapter 4

C.1 Socio-economic survey questions

Perceived forest area and forest quality change at village level (mixed gender and women focus group discussions)

Section: Change in forest area, quality and use

1. Overall, has the net area of forest cover in this village increased, stayed the same, or decreased since two years ago? *Codes: 1 = increased; 2 = stayed the same; 3 = decreased; -8 = does not apply; -9 = respondents do not know*

2. Overall, has forest quality increased, stayed the same, or decreased since two years ago? By forest quality we mean the availability of goods and services of the forest related to density of woody material, forest health, and biological productivity and diversity. *Codes: 1 = increased; 2 = stayed the same; 3 = decreased; -8 =does not apply; -9 = respondents do not know*

Perceived forest pressure sources at village level (mixed gender and women focus group discussions)

Section: Change in forest area, quality and use

1. If there has been change in the area of forest cover in the village since two years ago, please tell us in what land tenure areas these changes occurred and the three main on-site causes of those changes. *(coded per land tenure type, type (increase/decrease) and ranking of causes of forest change)*

2. If there are particular driving forces related to these forest cover changes (e.g. change of policies, rules, prices, REDD+ project), please explain what those forces are.

3. If there has been change in forest quality since two years ago, please tell us in what land tenure areas these changes occurred, and the three main on-site causes of those changes. *(coded per land tenure type, type (increase/decrease) and ranking of causes of forest change)*

4. If there are particular driving forces related to these changes in forest quality (e.g. change of policies, rules, prices), please explain what those forces are.

Section: Change in forest cover and forest income in the last two years

We want to know how your forest-based income has changed in the last two years (24 months) and the reasons for that change.

1. Has your household cleared any forest during the past two years? *1 = yes; 0 = no*

2. If yes, how much forest was cleared in total in the last 2 years? *Indicate total area cleared in hectares, in up to 3 parcels total.*

3. If yes, what was the main purpose of clearing the forest land? *Codes: 1=cropping; 2=pasture; 3=tree plantation; 4=non-agricultural uses*

4. If for cropping, which principal crop was grown? *See code book: Product*

C.2 Reported confidence levels for forest conversion reporting using high resolution imagery

Brazil – Transamazon

For the samples where a forest conversion was found (n=197 out of n=270), 82% (n=161) were marked with high confidence, 18% (n=35) with middle confidence and 0.5% (n=1) with low confidence. Most of the samples with middle or low confidence (n=31) were in the agriculture (here: pasture) conversion class. In those cases, the resolution of the available imagery was not high enough to confidently mark these as pastures.

Peru – MDD

71% (n=145) of the forest conversion samples (n=203) were marked with high confidence, 28% (n=57) with middle, and 1 sample with low confidence. Most of the samples with a middle confidence were in the agriculture and degradation class (n=28 and n=25 respectively). Here, it often encompassed small-scale disturbances with some degree of regrowth and recurrent degradation over time, sometimes containing bushy cropland.

Indonesia-KCCP

76% (n=156) of the forest conversion samples (n=206) were marked with high confidence, 24% (n=50) with middle confidence, and no samples were marked as low confidence. The samples with middle confidence occurred mostly in the agriculture (n=15), degradation (n=18) and tree plantation (n=16) classes. In most of these cases the distinction between degraded forest and shrub-mix dryland farming was rather difficult to make. For the tree plantation samples with middle confidence, it was unsure if it was oil palm or some other tree crop (when it concerned plantations in an early stage).

Indonesia-Katingan

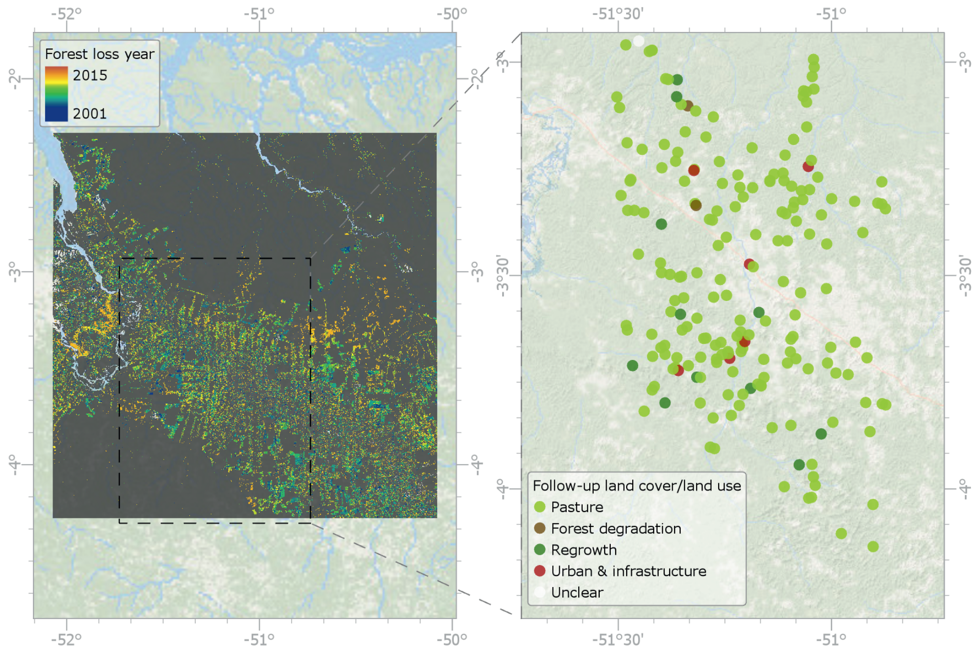
Of the forest conversion samples (n=203), 62% (n=126) and 38% (n=77) were marked with high and middle confidence respectively. Most of the samples with middle confidence were in the degradation class. Due to cloud coverage in the high-resolution imagery, it was often not clear if those samples marked as degraded were later converted to oil palm or not.

Vietnam – Cat Tien

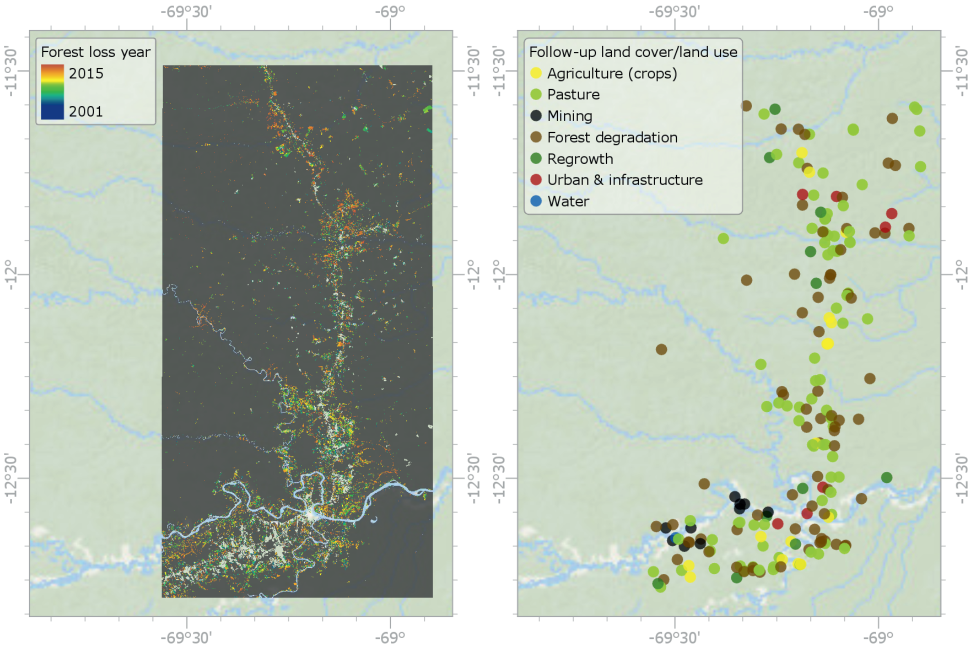
Of the forest conversion samples (n=227), 71% (n=162) and 29% (n=65) were marked with high and middle confidence respectively. The samples with middle confidence occurred mainly in the agriculture (n=39) and degradation (n=21) class. In the agriculture class, this was caused by uncertainty whether some pixels where covered with an orchard or cashew plantation (agriculture) or some other type of plantation. Uncertainty in the degradation class was caused by disturbed forest with partial regrowth.

C.3 Remotely sensed forest loss and follow-up land use/cover

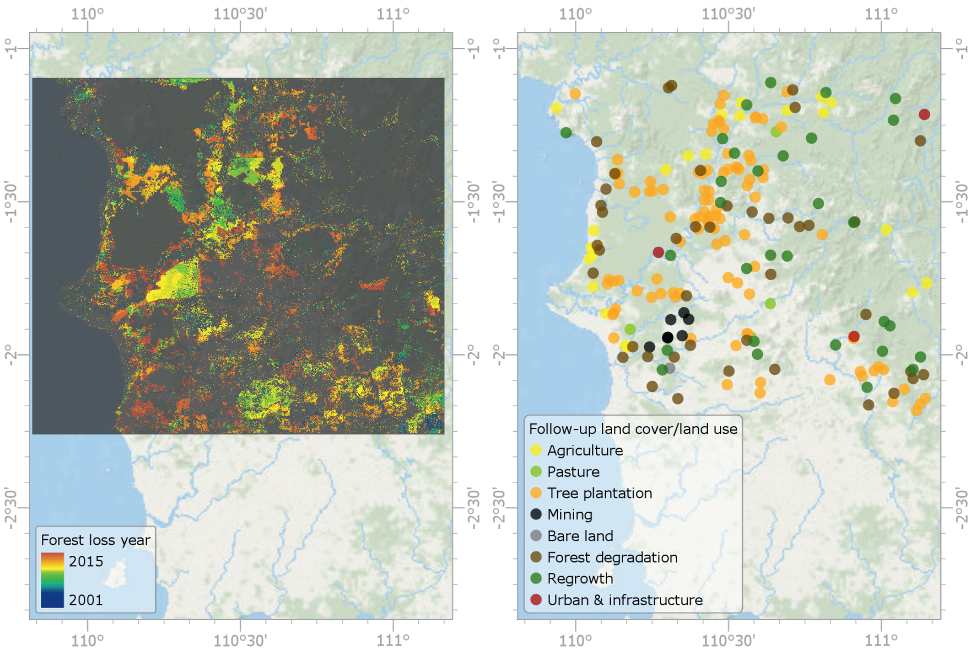
Left maps show forest cover loss in 2001-2015, source Global Forest Watch (Hansen et al., 2013) and forest definition thresholds applied (section 4.2.4). Right maps show the forest conversion samples and corresponding follow-up land cover and land use types as observed with high resolution imagery.



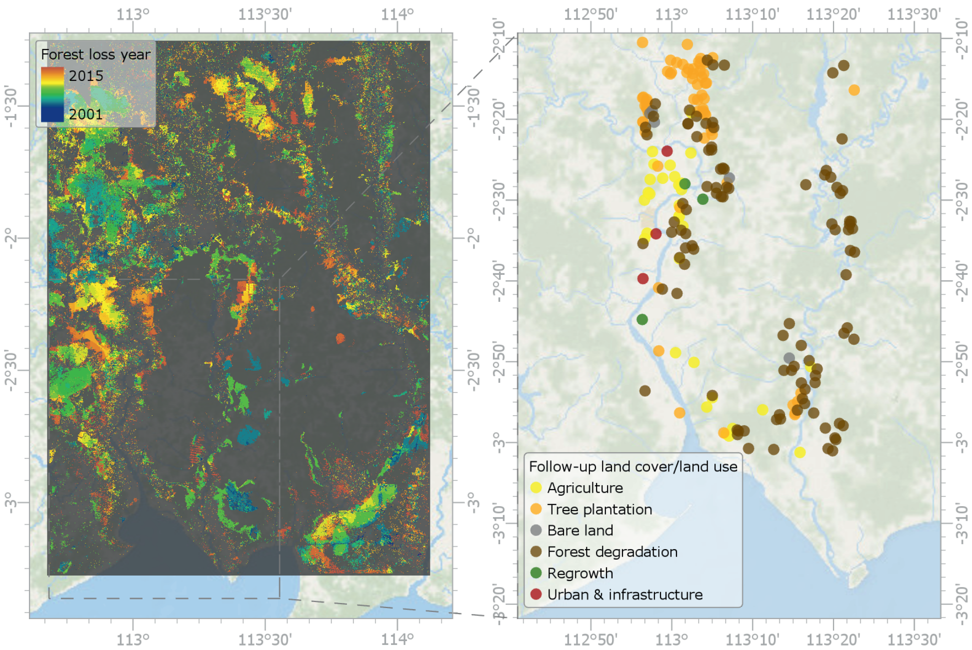
(a) Brazil-Transamazon



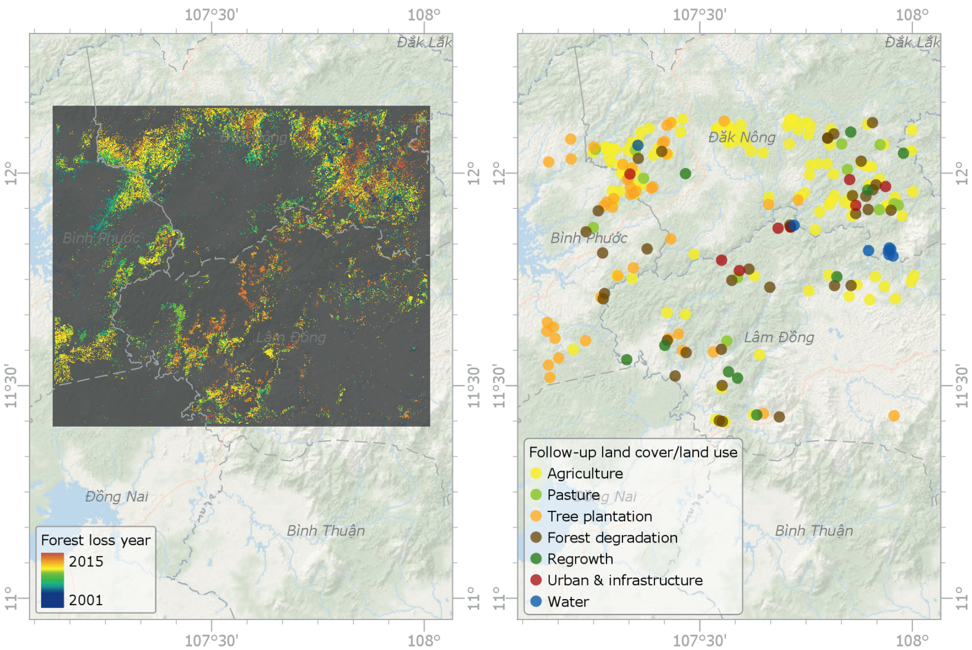
(b) Peru-Madre de Dios



(c) Indonesia-KCCP



(d) Indonesia-Katingan

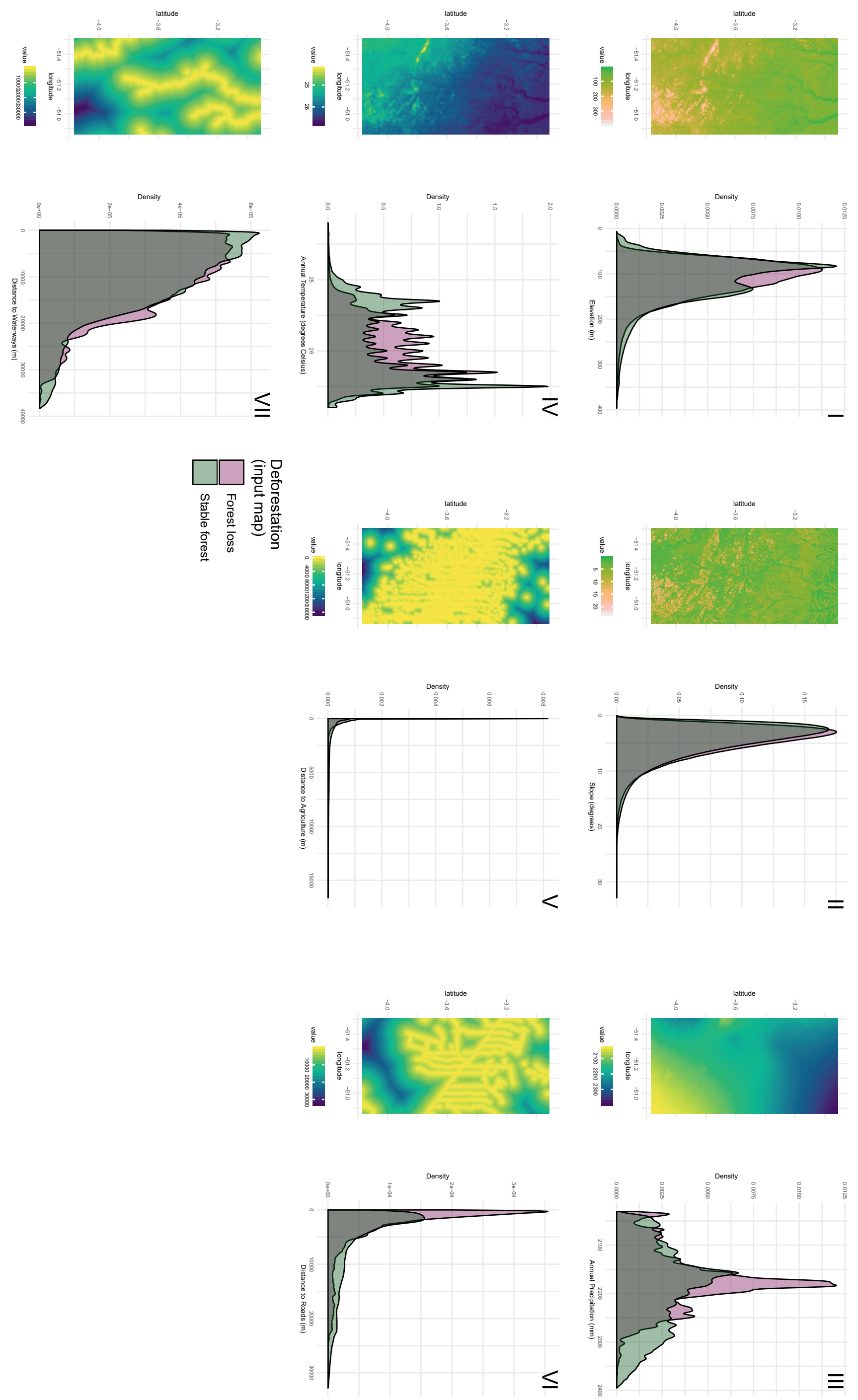


(e) Indonesia-Vietnam-Cat Tien

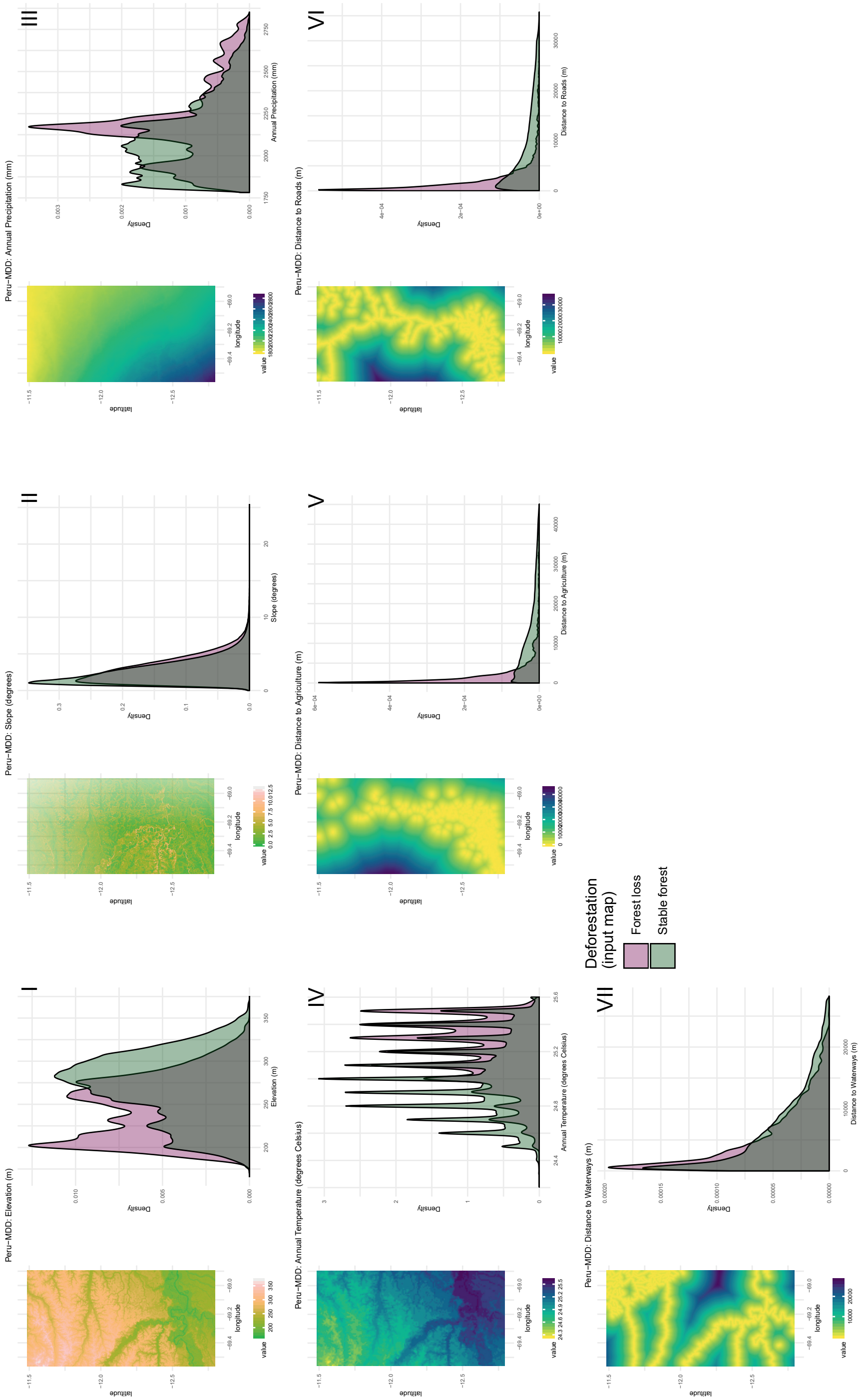
Figure C.3: Remotely sensed forest loss and follow-up land uses and land covers

C.4 Random Forest prediction variables

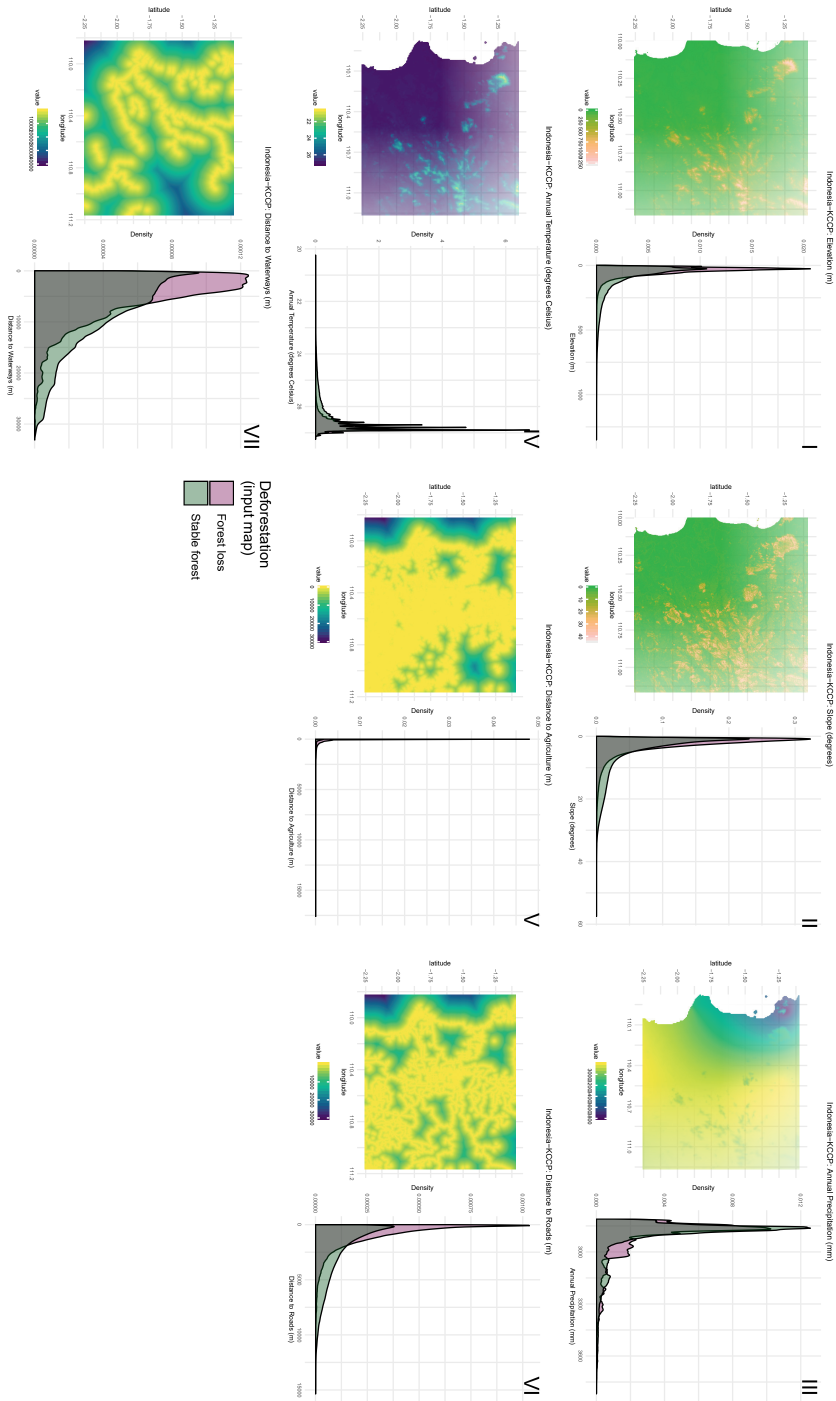
Figure C.4 shows the spatial distribution and density plots of all the Random Forest prediction variables per site. From left to right, top to bottom: elevation (I), slope (II), annual precipitation (III), annual temperature (IV), distance to agriculture (V), distance to roads (VI), and distance to waterways (VII).



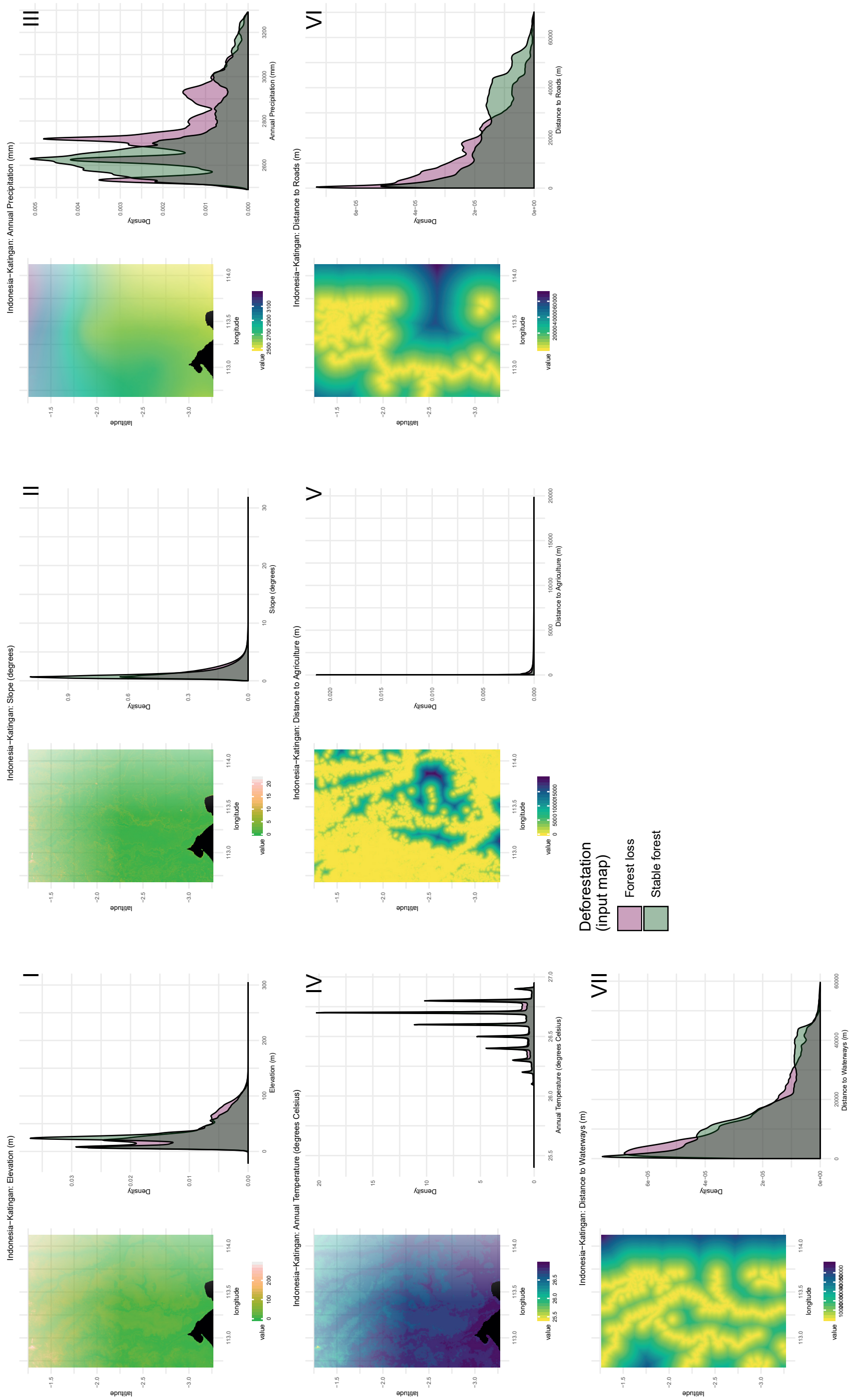
(a) Brazil-Transamazon


I
Peru−MDD: Elevation (m)
Elevation (m)
II
Peru−MDD: Slope (degrees)
Slope (degrees)
III
Peru−MDD: Annual Precipitation (mm)
Annual Precipitation (mm)
IV
Peru−MDD: Annual Temperature (degrees Celsius)
Annual Temperature (degrees Celsius)
V
Peru−MDD: Distance to Agriculture (m)
Distance to Agriculture (m)
VI
Peru−MDD: Distance to Roads (m)
Distance to Roads (m)
VII
Peru−MDD: Distance to Waterways (m)
Distance to Waterways (m)
Density
latitude
longitude
value
Deforestation
(input map)
Forest loss
Stable forest

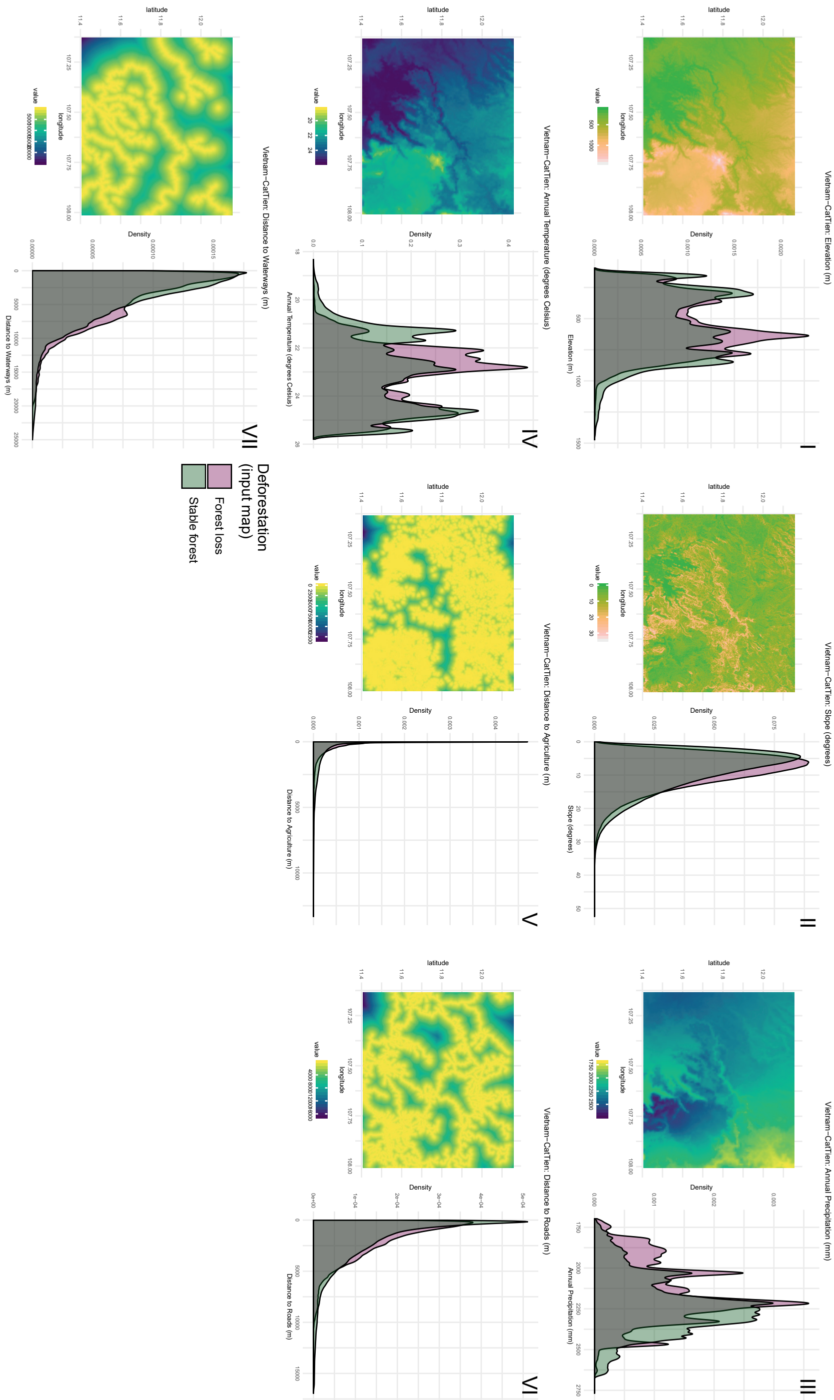
(b) Peru-MDD



(c) Indonesia-KCCP



(d) Indonesia-Katingan



(e) Vietnam-Cat Tien

Figure C.4: Random forest prediction variables – spatial distribution, and density plots of forest loss and stable forest pixels

C.5 Error matrices Random Forest model predictions

Brazil -Transamazon

		Error matrix (pixel counts)					Error matrix (estimated area proportions)			Accuracies		
		Reference (input)					Reference (input)					
		Loss	Stable	Total	W_i		Loss	Stable	Total (W_i)		Loss	Stable
RF prediction	Loss	248,142	17,389	8,576,936	0.808	Loss	0.755	0.053	0.808	User's	0.935	0.848
	Stable	40,294	225,237	2,044,336	0.192	Stable	0.029	0.163	0.192	Producer's	0.963	0.755
	Total	288,436	242,626	10,621,272	1.000	Total	0.784	0.216	1.000	Overall	0.918	

Peru-MDD

		Error matrix (pixel counts)					Error matrix (estimated area proportions)			Accuracies		
		Reference (input)					Reference (input)					
		Loss	Stable	Total	W_i		Loss	Stable	Total (W_i)		Loss	Stable
RF prediction	Loss	550,447	1,062,626	1,613,073	0.146	Loss	0.050	0.096	0.146	User's	0.341	0.999
	Stable	8,242	9,400,193	9,408,435	0.854	Stable	0.001	0.853	0.854	Producer's	0.985	0.898
	Total	558,689	10,462,819	11,021,508	1.000	Total	0.051	0.949	1.000	Overall	0.903	

Indonesia-KCCP

		Error matrix (pixel counts)					Error matrix (estimated area proportions)			Accuracies		
		Reference (input)					Reference (input)					
		Loss	Stable	Total	W_i		Loss	Stable	Total (W_i)		Loss	Stable
RF prediction	Loss	2,526,043	1,881,180	4,407,223	0.287	Loss	0.165	0.123	0.287	User's	0.573	0.976
	Stable	258,710	10,670,168	10,928,878	0.713	Stable	0.017	0.696	0.713	Producer's	0.907	0.850
	Total	2,784,753	12,551,348	15,336,101	1.000	Total	0.182	0.818	1.000	Overall	0.860	

Indonesia-Katingan

		Error matrix (pixel counts)					Error matrix (estimated area proportions)			Accuracies		
		Reference (input)					Reference (input)					
		Loss	Stable	Total	W_i		Loss	Stable	Total (W_i)		Loss	Stable
RF prediction	Loss	9,772,681	3,302,353	13,075,034	0.371	Loss	0.277	0.094	0.371	User's	0.747	0.946
	Stable	1,190,706	21,012,218	22,202,924	0.629	Stable	0.034	0.596	0.629	Producer's	0.891	0.864
	Total	10,963,387	24,314,571	35,277,958	1.000	Total	0.311	0.689	1.000	Overall	0.873	

Vietnam-Cat Tien

		Error matrix (pixel counts)					Error matrix (estimated area proportions)			Accuracies		
		Reference (input)					Reference (input)					
		Loss	Stable	Total	W_i		Loss	Stable	Total (W_i)		Loss	Stable
RF prediction	Loss	1,538,655	782,643	2,321,298	0.334	Loss	0.221	0.113	0.334	User's	0.663	0.965
	Stable	159,888	4,467,936	4,627,824	0.666	Stable	0.023	0.643	0.666	Producer's	0.906	0.851
	Total	1,698,543	5,250,579	6,949,122	1.000	Total	0.244	0.756	1.000	Overall	0.864	

W_i: area weight

Figure C.5: Error matrices, area proportions and map accuracies

Appendix D Supplementary material for chapter 5

D.1 Pre- and post-intervention differences between countries

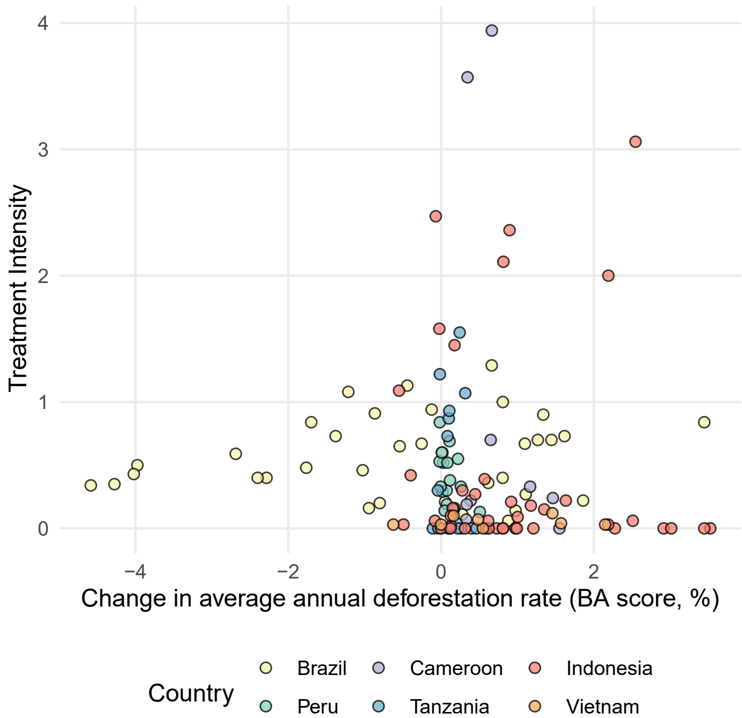


Figure D.1.1: Treatment intensity and BA score per country

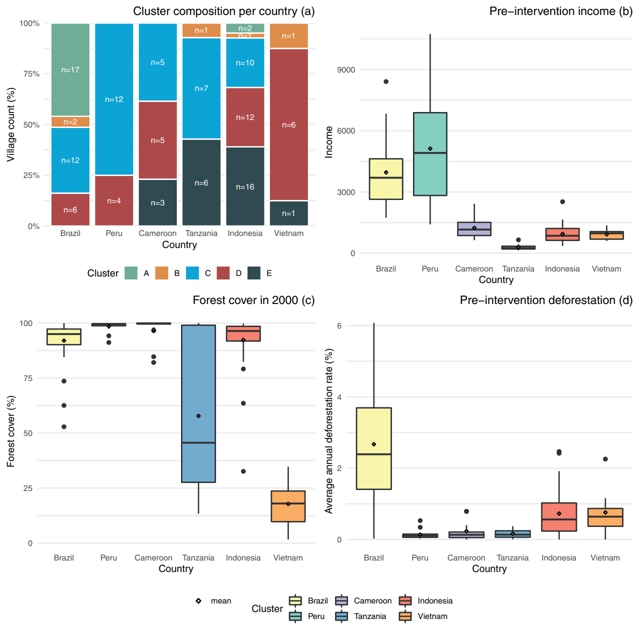


Figure D.1.2: Pre-intervention differences between countries

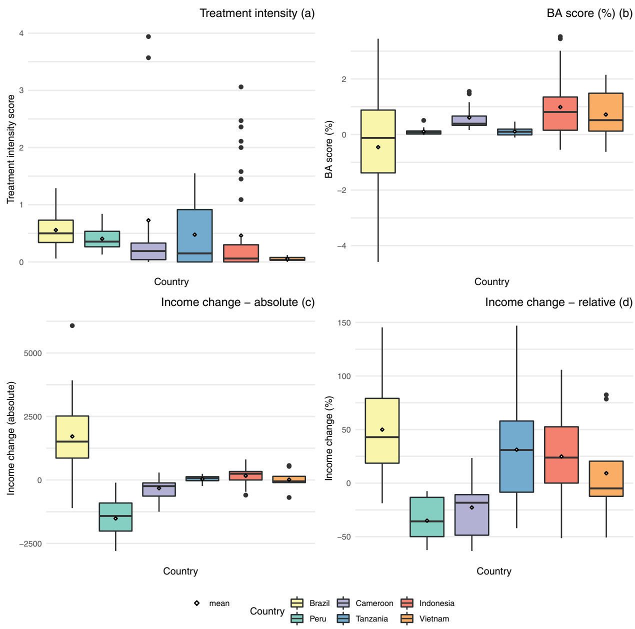


Figure D.1.3: Post-intervention differences between countries

Table D.1: Test results of pre- and post-intervention differences between countries, with p-values <0.05 in bold

Pre-intervention: income

Kruskal-Wallis			Wilcoxon rank sum test, p-values corrected using 'holm' method					
chi-squared	df	p-value		Brazil	Peru	Cameroon	Tanzania	Indonesia
102.2906	5	**1.738338e-20**	Peru	0.503				
			Cameroon	**0.000**	**0.000**			
			Tanzania	**0.000**	**0.000**	**0.000**		
			Indonesia	**0.000**	**0.000**	0.249	**0.000**	
			Vietnam	**0.000**	**0.000**	0.537	**0.000**	0.884

Pre-intervention: forest cover in 2000

Kruskal-Wallis			Wilcoxon rank sum test, p-values corrected using 'holm' method					
chi-squared	df	p-value		Brazil	Peru	Cameroon	Tanzania	Indonesia
45.65748	5	**1.066358e-08**	Peru	**0.001**				
			Cameroon	**0.009**	0.334			
			Tanzania	0.258	0.172	0.065		
			Indonesia	0.386	**0.001**	**0.007**	0.258	
			Vietnam	**0.000**	**0.000**	**0.002**	**0.011**	**0.000**

Pre-intervention: average annual deforestation rates

Kruskal-Wallis			Wilcoxon rank sum test, p-values corrected using 'holm' method					
chi-squared	df	p-value		Brazil	Peru	Cameroon	Tanzania	Indonesia
26.24086	5	**8.01E-05**	Peru	1.000				
			Cameroon	0.502	**0.000**			
			Tanzania	1.000	1.000	**0.001**		
			Indonesia	**0.007**	**0.004**	1.000	**0.006**	
			Vietnam	0.658	0.502	1.000	0.502	1.000

Post-intervention: treatment intensity

Kruskal-Wallis			Wilcoxon rank sum test, p-values corrected using 'holm' method					
chi-squared	df	p-value		A	B	C	D	
25.50602	5	**1.11E-04**	Peru	0.941				
			Cameroon	0.191	0.994			
			Tanzania	1.000	1.000	1.000		
			Indonesia	**0.001**	0.064	1.000	1.000	
			Vietnam	**0.000**	**0.001**	0.941	1.000	1.000

Post-intervention: change in average annual deforestation rates (BA score)

Kruskal-Wallis			Wilcoxon rank sum test, p-values corrected using 'holm' method					
chi-squared	df	p-value		Brazil	Peru	Cameroon	Tanzania	Indonesia
26.24086	5	**8.013152e-05**	Peru	1.000				
			Cameroon	0.502	**0.000**			
			Tanzania	1.000	1.000	**0.001**		
			Indonesia	**0.007**	**0.004**	1.000	**0.006**	
			Vietnam	0.658	0.502	1.000	0.502	1.000

Post-intervention: income change(absolute)

Kruskal-Wallis			Wilcoxon rank sum test, p-values corrected using 'holm' method					
chi-squared	df	p-value		Brazil	Peru	Cameroon	Tanzania	Indonesia
86.02071	5	**4.60E-17**	Peru	**0.000**				
			Cameroon	**0.000**	**0.000**			
			Tanzania	**0.000**	**0.000**	0.127		
			Indonesia	**0.000**	**0.000**	**0.002**	0.170	
			Vietnam	**0.000**	**0.000**	0.170	0.467	0.467

Post-intervention: income change(relative)

Kruskal-Wallis			Wilcoxon rank sum test, p-values corrected using 'holm' method					
chi-squared	df	p-value		Brazil	Peru	Cameroon	Tanzania	Indonesia
52.40483	5	**4.46E-10**	Peru	**0.000**				
			Cameroon	**0.000**	1.000			
			Tanzania	1.000	**0.002**	0.060		
			Indonesia	0.098	**0.000**	**0.001**	1.000	
			Vietnam	0.073	**0.045**	0.534	1.000	1.000

D.2 Test results for assessment of pre-intervention cluster differences

NB: With p-values <0.05 in bold

Table D.2: Test results for assessment of pre-intervention cluster differences

Initial income (pre-intervention)							
Kruskal-Wallis			Wilcoxon rank sum test, p-values corrected using ‘holm’method				
chi-squared	df	p-value		A	B	C	D
33.64487	4	**8.81E-07**	B	6.96E-01	NA	NA	NA
			C	**1.59E-02**	1.00E+00	NA	NA
			D	**9.92E-04**	1.00E+00	1.00E+00	NA
			E	**3.69E-08**	1.00E+00	**6.36E-04**	**9.09E-03**
Forest cover in 2000							
Kruskal-Wallis			Wilcoxon rank sum test, p-values corrected using ‘holm’method				
chi-squared	df	p-value		A	B	C	D
11.47567	4	**2.17E-02**	B	1.00E+00	NA	NA	NA
			C	2.00E-01	2.11E-01	NA	NA
			D	1.00E+00	1.00E+00	2.11E-01	NA
			E	1.00E+00	1.00E+00	1.96E-01	1.00E+00
Average annual deforestation rates (pre-intervention)							
Kruskal-Wallis			Wilcoxon rank sum test, p-values corrected using ‘holm’method				
chi-squared	df	p-value		A	B	C	D
51.5132	4	**1.74E-10**	B	6.10E-02	NA	NA	NA
			C	**1.49E-12**	**2.38E-02**	NA	NA
			D	**1.40E-11**	**4.43E-02**	1.00E+00	NA
			E	**7.38E-12**	**2.09E-02**	1.00E+00	1.00E+00

D.3 Intervention type composition – clusters compared

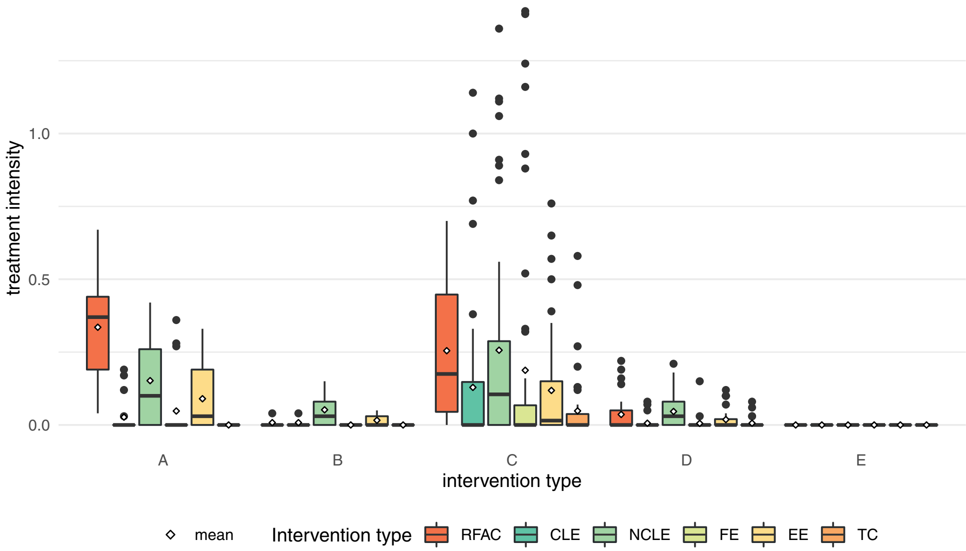


Figure D.3: Boxplots of treatment intensity of intervention types per cluster. RFAC= Restrictions on forest access & conversion; CLE = conditional livelihood enhancements; NCLE = non-conditional livelihood enhancements; FE = forest enhancements; EE = environmental education; TC = tenure clarification. Upper and lower extremes of whiskers represent $Q3 + 1.5*$interquartile range (IQR) and $Q1 - 1.5 * IQR$ respectively, where $IQR = Q3 - Q1$.

D.4 Tests results for differences in treatment intensity per intervention type between cluster A and C

Table D.4: Levene's test and T-Tests results

Intervention type	Levene's test F value	Levene's test p-value	T-Test p-value[a]	Potential difference[b]
RFAC	1.383161	0.243989	0.173087	FALSE
CLE	2.744392	0.102569	0.102569	FALSE
NCLE	2.205201	0.142533	0.233251	FALSE
FE	2.155323	0.147052	0.147052	FALSE
EE	0.488797	0.487038	0.538822	FALSE
TC	3.091226	0.083572	0.083572	FALSE

a Equal variances assumed, as all Levene's tests'p-values were >0.05
b True if p-value of T-test > 0.05

D.5 Test results for assessment of post-intervention cluster differences

NB: With p-values <0.05 in bold

Table D.5: Test results for assessment of post-intervention cluster differences

Change in income (pre-vs post-intervention)							
Kruskal-Wallis			Wilcoxon rank sum test, p-values corrected using 'holm' method				
chi-squared	df	p-value		A	B	C	D
24.72397	4	**5.72E-05**	B	3.70E-01	NA	NA	NA
			C	**5.55E-05**	9.66E-01	NA	NA
			D	**2.18E-04**	9.66E-01	9.66E-01	NA
			E	**8.25E-04**	9.66E-01	7.78E-01	9.66E-01
Relative change in income (pre-vs post-intervention)							
Kruskal-Wallis			Wilcoxon rank sum test, p-values corrected using 'holm' method				
chi-squared	df	p-value		A	B	C	D
14.39925	4	**6.12E-03**	B	1.00E+00	NA	NA	NA
			C	**3.40E-02**	1.00E+00	NA	NA
			D	**1.75E-02**	1.00E+00	1.00E+00	NA
			E	1.00E+00	1.00E+00	3.53E-01	1.23E-01

D.6 Intervention types effects on perceived well-being

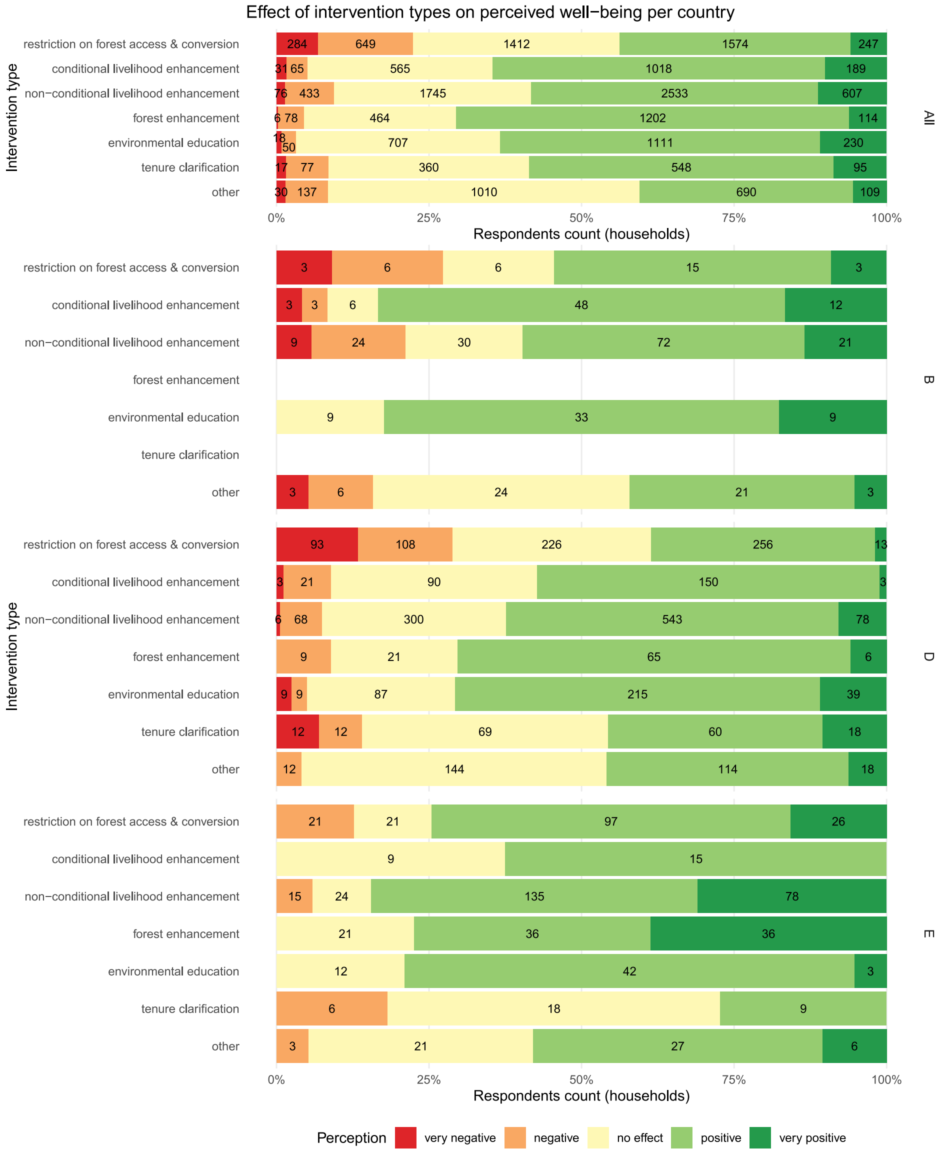


Figure D.6: Effects of intervention types on perceived well-being (All, cluster B, D, E)

D.7 Intervention types effects on perceived well-being (country)

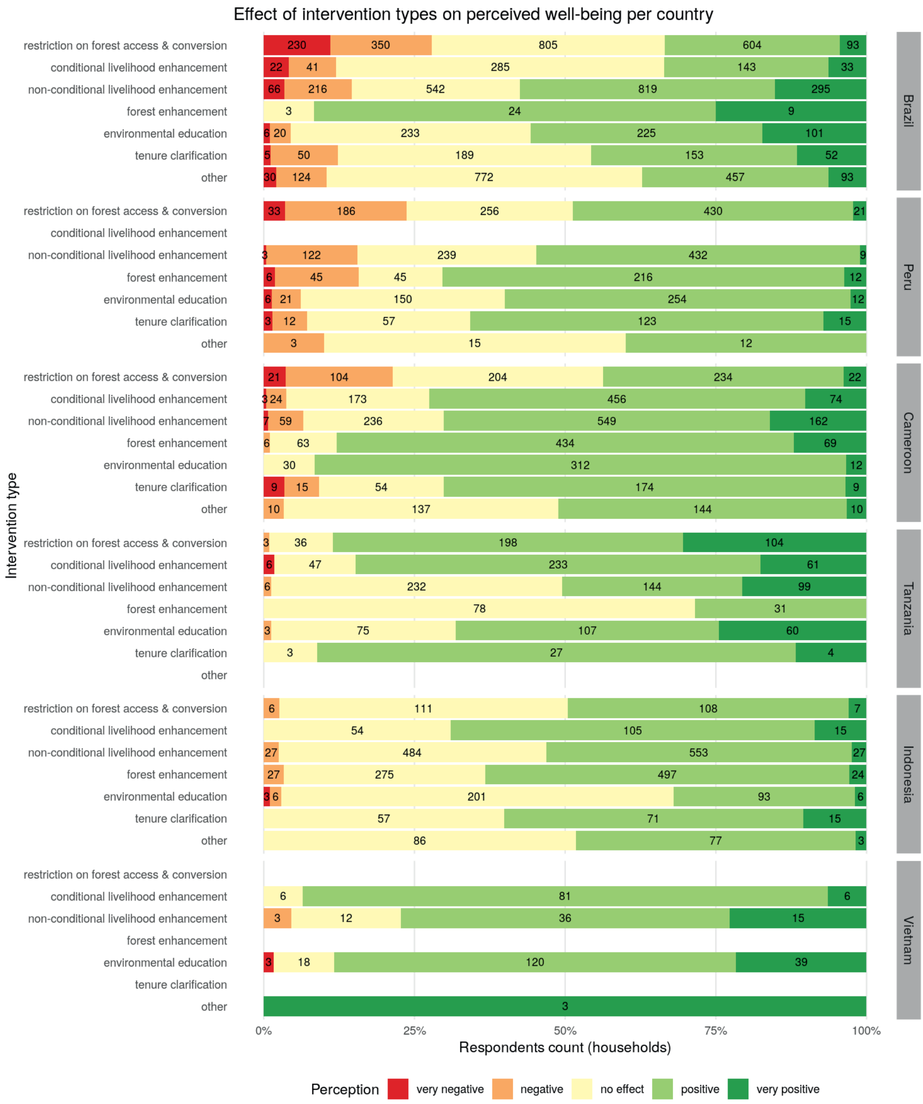


Figure D.7: Effects of intervention types on perceived well-being per country

Acknowledgements

When it was time to start writing the final thesis, the first "chapter" I worked on was this one. Believed to be the best-read part of a thesis, I would like to seize this opportunity to express my gratitude to all who helped me to start, persevere and finally complete this thesis.

I would like to thank my promotor, Martin Herold. For providing me this great opportunity, for building the bridge with CIFOR, for opening (international) doors, and for your trust (and patience!) during the past years. I really enjoyed all our brainstorming sessions.

Like we were once taught in Wageningen, if you want your collaboration to be truly successful, make sure to get an American woman on board. Well, I cannot think of a better suited person for the job as energizer than you, Amy Duchelle! I owe you a big thank you for all your help and for your supervision, for being my co-promotor and for this opportunity to become part of the CIFOR family.

Joyous meetings, in and outside the office... Niki De Sy, thank you for your supervision, for being my co-promotor, for your mentorship and friendship. Your humour and pragmatism were simply indispensable. Your *been there-done that* insights and attitude in the final phase of the PhD helped me to keep the end goal in mind at all times!

Even though you were only involved in the first years of my PhD, I am still grateful today for your supervision during the MSc thesis and start of the PhD, Valerio Avitabile. I really enjoyed our chats and appreciated your practical insights.

My PhD opponents are gratefully acknowledged for their time and effort to review this thesis and to participate in the defence.

Everyone inside the international CIFOR family is gratefully acknowledged. Just weeks after my job interview, you welcomed me at the M2 week in Istanbul. Several meetings, dinners, hikes and karaoke parties in Bogor and across Europe would follow. A special thanks to Arild, Augusta, Auria, Christine, Christopher, Claudio, Daju, Desta, Erin, Gabriela, Jessica, Josil, Made, Marina, Nia, Nugroho, Sven, Tia and William. I would like to thank all participants in and enumerators of the surveys that were conducted in the field, I am truly grateful that I could make use of the data for this thesis.

To Claudius, Lennart & Amber from the PE&RC graduate school, thank you for all your support, advice and motivating words during the PhD weekends, courses and workshops in Kenya and the Netherlands.

My current and former colleagues from the Laboratory of Geo-Information Science and Remote Sensing are gratefully acknowledged for being such a nice international group and for providing a pleasant (working) environment, also after working hours. As one of the few Dutchies I almost felt like the foreigner myself sometimes, but that unique vibe in the group is definitely something to be proud of. Keep it up! Therefore, I would like to thank Adugna, Agnieszka, Aldo, Alex, Andrei, Anne, Arend, Arnan, Arnold, Arun, Ben DV., Benjamin B., Benjamin K., Brice, Chenglong, Christelle, Corné, Cristina, Dainius, Danaë, Daniela, David, Devis, Diego, Eliakim, Erika R., Erika S., Eskender, Federico, Frans, Giulia, Gustavo, Harm, Ilan, Ioannis, Jaap-Willem, Jalal, Jan C., Jan V., Joao, Johannes B., Johannes R., John, Jose, Juha, Kalkidan, Karimon, Karina, Kathleen, Konstantin, Lala, Lammert, Linlin, Loïc, Lukasz, Marcello, Maria, Marian, Marston, Mathieu, Michael, Milutin, Na, Nandika, Natalia, Panpan, Patric, Paulo, Peter, Qijun, Quanxing, Richard, Robert, Roberto, Roland, Ron, Rosa Maria, Sabina, Samantha, Sarah, Shivangi, Simon, Sylvain, Sytze, Tom, Tsoefiet, Wanda, Willy, Ximena, Yang, Yunyu, and everyone whom I have forgotten to mention here. Juultje, I have enjoyed working together during your MSc thesis research. A special thanks to my paranymph Alvaro with whom I have shared most of this PhD journey, and to Antoinette and Truus for their endless support.

Every Tuesday night you were there for me and my fellow “slightly crippled” group members. Thank you Ingi, for keeping my joints and muscles (kind of) flexible. Your enthusiastic attitude works truly energizing!

The joy of putting all your remaining energy (and frustration...) in hitting a ball and playing a nice game of volleyball with you after a long day of computer screen staring...! Thanks to all my (current/former) team mates, trainers and coaches at Invicta: Albert, Anne, Anneleen, Annet, Bert, Carolien, Ellen, Geerte, Guido, Hans de B, Hans J, Henk, Herma, Jenneke, Joan, Karin, Lenneke, Lobke, Marieke, Marike, Maringa, Marit, Marlies, Martina, Nonja and Paula.

Roomie from the PE&RC course in Kenya and friend across three continents. Thank you Ni’ma, for showing me around in Bogor, and for teaching me how to cook Indonesian food.

Our conversations may have changed a bit over the years, but I am glad that good-old student vibe is still there (although I am afraid I really cannot call myself a student any longer after this...!). Thank you Marcel & Suzanne and the rest of the KZW-group, for the countless “borrels”, parties, dinners, board game nights, winter holidays and summer weekends.

Ulrike, I am so thankful that our friendship survives despite the distance which could not have been much larger; 16,653 kilometres as the crow flies... I am wondering, should I now repaint the print on that t-shirt you sent me?

We started our student careers together in Wageningen more than fourteen years ago, and I am glad that, although we all went our own ways, in the end all roads seem to lead back to Wageningen somehow, so thank you for your friendship Iris & Linda! And thank you Iris for standing by me as my paranymph at the day of my defence.

En dan aan mijn lieve familie: papa, mama, Janine, Jasper, Sarah, Piet, Joke, Peter, Marieke, Alexander en kleine Vincent, dank voor jullie morele steun en voor jullie geduld! Papa en mama, fijn dat jullie er altijd voor me zijn, en me al die jaren gesteund en aangemoedigd hebben. Vroeger wilde ik niets liever dan schrijfster worden. Dat mijn eerste boekje mèt ISBN nummer een PhD thesis zou worden, had ik toen niet kunnen bedenken, maar ik weet nu wel dat dit zonder jullie support niet zou zijn gelukt.

No one deserves a bigger thanks than you, Mark. Thank you for joining me in -sometimes spontaneous- adventures ("let's hop off this boat!"), whether big or small. And for being my beacon at difficult or slightly less adventurous times. I cannot wait to start a new adventure with you.

About the author

Astrid Bos was born in Hichtum, Wûnserdiel, the Netherlands on 8 May 1987. After receiving the Atheneum degree at Marne College in Bolsward, she moved to Wageningen, to study at Wageningen University & Research. Astrid obtained her Bachelor's degree in International Development Studies with a specialisation in Sociology of Rural Development in 2008.

After travelling and working in New Zealand, Astrid continued studying at Utrecht University in 2009, doing the Research Master of Science in Sustainable Development. She did a research internship on payments for environmental services and buffer zone management of a national park in Vietnam. For the thesis Astrid travelled to Ghana, where she studied the production and use of bamboo biomass energy and its role in forest landscape restoration. She received a Master's degree in Environmental Sciences in 2012.

During her Master's studies, her interest in human-landscape interactions and in geo-sciences in general grew, leading to a second Master of Science in Geographical Information Management and Applications, a joint Master from University of Twente/ITC Enschede, TU Delft, Utrecht University and Wageningen University & Research. As part of her studies she worked as an intern at ARCADIS, where she studied (semi-automated) change detection techniques for urban contexts. The MSc thesis focussed on analysing deforestation patterns and modelling forest change dynamics in Vietnam and was done in cooperation with researchers at Wageningen University & Research. Astrid obtained her Master's degree in Geographical Sciences in 2013.

Shortly after finishing her studies, Astrid continued as a PhD candidate at the Laboratory of Geo-Information Science and Remote Sensing at Wageningen University & Research where she started her research in 2014 as part of the *Global Comparative Study on REDD+* of the Center for International Forestry Research (CIFOR). Her work focussed on performance assessment of sub-national forest-based climate change mitigation efforts through integrating remote sensing based forest change data with socio-economic survey data. Her research resulted in this thesis and several peer-reviewed publications.

Peer-reviewed journal publications

Bos, A. B., De Sy, V., Duchelle, A.E., Herold, M., Martius, C., Tsendbazar, N.-E., 2019. Global data and tools for local forest cover loss and REDD+ performance assessment: Accuracy, uncertainty, complementarity and impact. *International Journal of Applied Earth Observation and Geoinformation*, 80 (pp. 295–311).

Bos, A. B., Duchelle, A. E., Angelsen, A., Avitabile, V., De Sy, V., Herold, M., Joseph, S., de Sassi, C., Sills, E. O., Sunderlin, W. D., and Wunder, S. (2017). Comparing methods for assessing the effectiveness of subnational REDD+ initiatives. *Environmental Research Letters*, 12(7):074007

de Sassi, C., Joseph, S., **Bos, A.B.**, Duchelle, A.E., Ravikumar, A., Herold, M., 2015. Towards integrated monitoring of REDD+. *Current Opinion in Environmental Sustainability*, 14 (pp. 93–100)

Peer-reviewed book chapters

Simonet, G., **Bos, A. B.**, Duchelle, A. E., Pradnja Resosudarmo, I. A., Subervie, J. & Wunder, S. (2018). Forests and carbon: The impacts of local REDD+ initiatives. In A. Angelsen, C. Martius, V. De Sy, A. E. Duchelle, A. M. Larson, & T. T. Pham (Eds.), *Transforming REDD+* (pp. 117–130). Bogor, Indonesia: Center for International Forestry Research (CIFOR)

Avitabile, V., Schultz, M., Salvini, G., Pratihast, A. K., **Bos, A.**, Herold, N., … Herold, M. (2017). Forest Change and REDD+ Strategies. In A. Nauditt & L. Ribbe (Eds.), *Land Use and Climate Change Interactions in Central Vietnam* (pp. 33–68). Singapore: Springer Singapore

Sunderlin, W. D., Pratama, C. D., **Bos, A. B.**, Avitabile, V., Sills, E. O., de Sassi, C., … Anandadas, A. (2014). REDD+ on the ground: The need for scientific evidence. In E. O. Sills, S. Atmadja, C. de Sassi, A. E. Duchelle, D. Kweka, I. A. P. Resosudarmo, & W. D. Sunderlin (Eds.), *REDD+ on the ground: A case book of subnational initiatives across the globe* (pp. 2–21). Bogor, Indonesia: Center for International Forestry Research (CIFOR)

Other publications

Bos, A.B. (2019). Can you trust the numbers? *Forest News.* URL https://forestsnews.cifor.org/61007/can-you-trust-the-numbers

Bos, A. B., De Sy, V., Duchelle, A. E., & Herold, M. (2018). Are REDD+ interventions targeting the right drivers on the ground? Insights from space and below. Copenhagen, Denmark:

Forests & Livelihoods - Assessment, Research, and Engagement (FLARE)

Bos, A. B., Avitabile, V., De Sy, V., Duchelle, A. E., Herold, M., & Martius, C. (2017). Unravelling uncertainty - Combining forest cover change products and biomass datasets in the context of REDD+. In *Book of abstracts* (p. 119). Freiburg, Germany: International Union of Forest Research Organizations (IUFRO)

Bos, A. B., Avitabile, V., Herold, M., Duchelle, A. E., Joseph, S., Sassi, C. de, … Wunder, S. (2016). Assessing the effectiveness of subnational REDD+ initiatives by tree cover change analysis. Montpellier, France: Annual Meeting of the Association for Tropical Biology and Conservation

PE&RC Training and Education Statement

With the training and education activities listed below the PhD candidate has complied with the requirements set by the C.T. de Wit Graduate School for Production Ecology and Resource Conservation (PE&RC) which comprises of a minimum total of 32 ECTS (= 22 weeks of activities)

The C.T. De Wit Graduate School PE&RC
PRODUCTION ECOLOGY
& RESOURCE CONSERVATION

Review of literature (4.5 ECTS)

- REDD+ monitoring across scales and disciplines

Writing of project proposal (4.5 ECTS)

- Multi-scale monitoring of forest carbon changes for REDD+

Post-graduate courses (5.5 ECTS)

- Uncertainty propagation in spatial and environmental modelling; PE&RC (2014)
- Land dynamics – Getting to the bottom of Mount Kenya; PE&RC (2015)

Laboratory training and working visits (4.5 ECTS)

- Global comparative study on REDD+ (GCS) module 2: project collaboration & progress updates; CIFOR Headquarters, Bogor, Indonesia (2015)
- GCS REDD+ coordination meeting; CIFOR, Bonn, Germany (2016)
- CIFOR GCS Module 2 Meeting, project collaboration; CIFOR, Montpellier, France (2016)
- CIFOR GCS Module 3 Meeting, project collaboration; CIFOR Headquarters, Bogor, Indonesia (2017)

- Annual meetings, Climate Change & Energy meetings, GCS meetings; CIFOR Headquarters, Bogor, Indonesia (2017)

Invited review of (unpublished) journal manuscript (2 ECTS)

- Environmental Research Letters: Global deforestation modelling (2018)
- Environmental Science & Policy: Deforestation-free supply chains (2019)

Deficiency, Refresh, Brush-up courses (3 ECTS)

- Geo Scripting (GRS-33806); GRS-WUR (2015)

Competence strengthening / skills courses (3.7 ECTS)

- Writing workshop; CIFOR (2016)
- Workshop Guidos Toolbox: Digital Image Analysis of Pattern, Connectivity, Fragmentation and more; European Commission Joint Research Centre (2017)
- Supervising BSc and MSc thesis students; Educational Staff Development, WUR (2018)
- Communication with the Media and the General Public, Wageningen Graduate Schools (2018)
- Infographics and Iconography, Wageningen Graduate Schools (2018)

PE&RC Annual meetings, seminars and the PE&RC weekend (3 ECTS)

- PE&RC introduction weekend (2014)
- Wageningen Graduate Schools (WGS) Workshop Carousel (2015, 2016, 2018)
- PE&RC mid-term weekend (2016)
- PE&RC last year weekend (2018)

Discussion groups / local seminars / other scientific meetings (7.5 ECTS)

- Remotely sensed spatio-temporal dynamics of vegetation (2014)
- Remote sensing thematic group at WUR (2014-2019)
- PhD discussion group on Agriculture–Climate–Forest–Food (2014-2019)
- Evaluating anthropogenic changes in tropical forests (2015)

International symposia, workshops and conferences (9 ECTS)

- Annual meeting CIFOR workshop; Bogor, Indonesia (2015)
- Side event 44th UNFCCC Subsidiary Bodies (SBI/SBSTA); Bonn, Germany (2016)
- Association for Tropical Biology and Conservation (ATBC); Montpellier, France (2016)
- International Union of Forest Research Organizations (IUFRO) 125th Anniversary Congress; Freiburg, Germany (2017)
- Forests & Livelihoods - Assessment, Research, and Engagement (FLARE) 4th Annual Meeting; Copenhagen, Denmark (2018)

Lecturing / supervision of practical's / tutorials (0.9 ECTS)

- GRS33806 Geo scripting (2017)
- GRS10806 Geo-Information Science for Planning and Design (2017, 2019)

Supervision of MSc students (3 ECTS)

- Linking people and landscapes: Analysing deforestation trends and their drivers in REDD+ subnational initiatives (2018)

This research is part of CIFOR's Global Comparative Study on REDD+ (www.cifor.org/gcs). The funding partners that have supported this research include the Norwegian Agency for Development Cooperation (Norad), the International Climate Initiative (IKI) of the German Federal Ministry for the Environment, Nature Conservation, Building and Nuclear Safety (BMUB), and the CGIAR Research Program on Forests, Trees and Agroforestry (CRP-FTA) which is financially supported by the donors to the CGIAR Fund.

Financial support from Wageningen University for printing this thesis is gratefully acknowledged.

Cover design by Astrid Bos