Governance, Management, and Conservation Success of Protected Areas in Brazil and Colombia

by

Christoph Nolte

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Natural Resources and Environment)
in the University of Michigan
2014

Doctoral Committee:

Professor Arun Agrawal, Chair
Professor Daniel G. Brown
Professor Paul J. Ferraro, Georgia State University
Professor Elisabeth R. Gerber

DEDICATION

To the diversity of this planet

ACKNOWLEDGEMENTS

It is a joyful experience to stand on the shoulders of giants. From the moment I set out upon my PhD adventure at the University of Michigan's School of Natural Resources and Environment, I received an incredible diversity of generous support from many, many wonderful people. If, through my research, I was able to look just a little bit farther, I owe it to them.

Germans call their thesis advisor "*Doktorvater*" (doctoral father). In Arun Agrawal, I could not have wished for a better one. From my first baby steps in U.S. academia, to my first words in the language of grants, and on to the playground of research and publishing, I could count on him for thoughtful advice, unfailing support, and joyous banter. Arun's laughter fills rooms. I hope I will hear it often.

Paul Ferraro was the first scholar I invited to join my committee. I would have begged. His advice felt like receiving jewels from a treasure chest of coherent instrumental knowledge. I hope we will be chasing causal claims together for many years to come. Dan Brown sharpened my remote sensing view on Latin America's forests. Liz Gerber scrutinized my work through the lens of public policy. I am indebted to all of you.

Fieldwork needs partners and I was blessed with many. Johannes Scholl and Matthias von Bechtolsheim (knowledge managers at the German KfW Development Bank) were the first to signal interest in a rigorous evaluation of protected area concepts. Their endorsement opened the doors of park agencies, conservation trust funds, and non-governmental organizations (NGOs) across the Andes-Amazon region. Thank you. I am grateful to directors and staff of Colombia's national park service, Brazil's federal protected area agency ICMBio, and the Brazilian conservation NGOs WWF-Brasil and IMAZON for the remarkably open exchange of connections, data, and ideas. I am looking forward to the next step of our collaborative work.

Quasi-experimental models tend to be data hungry, and won't run properly without good fodder. I thank Lauren Coad, Jonas Geldmann, Fiona Leverington, and others for taming the unwieldy

METT database; Lilian Hangae and Marcelo Kinouchi for access to the Brazilian RAPPAM surveys; Matthew Hansen and Charlene M. Dimiceli for sharing remote sensing products; Britaldo Soares-Filho for his advanced road dataset; and Desirae Hoffmann for computing travel time estimates. Their ingredients were essential. For instrumental knowledge, I am grateful to Jeff Smith, Michael Moore, Kerby Shedden and his colleagues at the Center for Statistic Consultation and Research (CSCAR) at the University of Michigan. I wish researchers of every university in the world had free access to the services of such incredible analysts.

Sharing is beautiful in the world of thought. I was grateful for thoughtful feedback from Kobe de Pourq, Merlin Hanauer, Gommert Mes, Dan Miller, Baru Mshale, Gustavo Kattan, Pete Newton, Alexander Pfaff, Mark Zimsky, and eight anonymous reviewers of the *Proceedings of the National Academy of Sciences (PNAS), Conservation Biology*, *Environmental Research Letters*, and *Oryx*. They and many others helped sharpen my arguments.

This research would not have been possible without the generous funding from numerous organizations, including the National Science Foundation, the Gordon and Betty Moore Foundation, the Rights and Resources Initiative, Marshall Weinberg, and four University of Michigan institutions: the School of Natural Resources and Environment, the Rackham Graduate School, the Graham Environmental Sustainability Institute, and the International Institute.

I thank my colleagues at the International Forestry Resources and Institutions (IFRI) lab for the shared moments. I will miss Dan, Rebecca, Pete, Joan, Johan, Baru, Jodi, Catherine, Ahmed, and many others. Wonderful people ran my school: Rosina, Sondra, Diana, Jennifer, Judy, Dave, Lisa, Shannon, Moe, Mike – it was a pleasure to have met you.

Ann Arbor made me grow and dream in many more ways than I expected. I am eternally grateful for these four years and four months.

May all of us be able to spend more time in trees.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

ARPA	Amazon Region Protected Areas program
ATT	Average Treatment Effect on the Treated
CBD	Convention on Biological Diversity
CNUC	Cadastro Nacional de Unidades de Conservação (National Cadaster of Conservation Units)
GBMF	Gordon and Betty Moore Foundation
GEF	Global Environment Facility
GFCL	Global Forest Cover Loss
GLC	Global Land Cover
IUCN	International Union for the Conservation of Nature
LOESS	Locally Weighted Scatterplot Smoother
METT	Management Effectiveness Tracking Tool
MODIS	Moderate Resolution Imaging Spectroradiometer
NGO	Non-Governmental Organization
NIMA	National Imagery and Mapping Agency
NNM	Nearest Neighbor Matching
PAME	Protected Area Management Effectiveness
PiP SCS	Parks in Peril Site Consolidation Scorecard

PROARCA/CAPAS	Programa Ambiental Regional para Centroamerica / Central American Protected Areas System
PRODES	Programa de Cálculo do Desflorestamento da Amazônia (Program for the Calculation of Deforestation in the Amazon)
RAPPAM	Rapid Assessment and Prioritization of Protected Area Management
REDD	Reduced Emissions from Deforestation and Forest Degradation
TNC	The Nature Conservancy
UNEP	United Nations Environment Program
UNEP-WCMC	UNEP World Conservation Monitoring Centre
USAID	United States Agency for International Development
VCF	Vegetation Continuous Fields
VMAP	Vector Smart Map
WCPA	World Commission of Protected Areas
WDPA	World Database of Protected Areas
WWF	World Wide Fund for Nature

ABSTRACT

Protected areas are the most widespread policy instrument for the conservation of tropical biodiversity. Over the past decade, the effectiveness of their management has come under increased scrutiny, with donors, governments, non-governmental organizations, and other conservation stakeholders evaluating the management of over 9000 protected areas in 140 countries. Designed to inform the planning, implementation and evaluation of conservation interventions, such evaluations of protected area management effectiveness (PAME) tend to collect data on a wide range of management aspects considered important for protected area success. However, many widespread PAME evaluation methods do not directly measure whether protected areas achieve their conservation goals, and why. Moreover, there is a lack of evidence whether such methods can be used to predict conservation outcomes, and thus obviate their direct measurement. Using the example of tropical protected areas in the Brazilian Amazon and Colombia's Northeast, this dissertation 1) examines whether the two most widespread PAME evaluation scorecards provide good surrogates for the capacity of protected areas to achieve conservation objectives and 2) identifies factors which are closely associated with protected area performance, and may thus be particularly suitable as performance indicators and, potentially, targets of interventions. Combining remote sensing imagery, management indicator datasets and data from my own field research with a new take on quasi-experimental econometric estimation techniques, my analyses challenge the notion that widespread PAME evaluation methods measure the capacity of tropical forest protected areas to achieve conservation objectives: Indicator scores of the two most widespread PAME evaluation methods failed to predict the success of protected areas in reducing forest fires and deforestation in the Brazilian Amazon. Furthermore, my results also suggest that that land use rights – their design, clarification and enforcement – are key to understanding and predicting conservation outcomes in tropical protected areas, but tend to be neglected by the most widespread PAME evaluation methods. I conclude that the gap between PAME evaluations and the improvement of conservation performance needs to be narrowed, and propose five directions for research.

CHAPTER 1: INTRODUCTION

Protected areas have long been hailed as a cornerstone of tropical biodiversity conservation. Over the past decade, however, their effectiveness and management has come under increased scrutiny. In 2004, 194 countries party to the Convention on Biological Diversity (CBD) decided to "implement management effectiveness evaluations of at least 30% of [their] protected areas by 2010" (CBD 2004). As the deadline approached, 67 countries had reached this target, and 32 more had reached half of it, evaluating at least 15% of their protected territories (Leverington et al. 2010). With the support of conservation donors, non-governmental organizations (NGOs), and other stakeholders, at least 9000 protected area management effectiveness (PAME) evaluations had been carried out in more than 140 countries. Parties to the CBD did not stop there: The same year, they agreed to "work towards assessing 60 per cent of the total area of protected areas by 2015" (CBD 2010).

Tropical forest ecosystems have been at the core of this recent evaluation movement. Forest initiatives such as the World Bank/World Wide Fund for Nature (WWF) Alliance for Forest Conservation and Sustainable Use and WWF's Forest for Life Program developed what would later become the world's most widespread PAME assessment methodologies (Ervin 2003b; Stolton et al. 2007). PAME evaluations are the most widespread in regions containing large tropical forests: Latin America, Africa, and Asia (Coad et al. 2013). Countries in the biodiversity and forest-rich Andes-Amazon[1] region seem particularly committed to evaluation, applying an average of 5.6 distinct methodologies to evaluate their protected areas (Leverington et al. 2010). Although biome-specific analyses of the coverage of PAME assessments have not been carried out, the current evidence suggests that protected tropical forest ecosystems have been, and remain, a preferential target of PAME evaluations.

How did the international conservation community become so preoccupied with the management of tropical forest protected areas? Several factors appear to converge. First, protected areas have

[1] Bolivia, Brazil, Colombia, Ecuador, Peru, Venezuela

long been the most widely used tool for biodiversity conservation worldwide, especially in tropical forests. In 2012, national governments had declared more than 175.000 protected areas across the globe, covering 17 million km^2 of land (Bertzky et al. 2012). In 2010, countries party to the CBD set themselves the even more ambitious goal of protecting 17% of the area of each of the world's 823 ecoregions by 2020 (CBD & UNEP 2010). Among the 273 ecoregions (33%) that are already reaching this target, tropical forests are particularly well represented (Schmitt et al. 2009; Bertzky et al. 2012).

Second, in spite of the significant growth of the global protected area network, the status of most global biodiversity indicators continues to deteriorate (Butchart et al. 2010; Pereira et al. 2012). Tropical forests are no exception, with large-scale deforestation continuing almost unabated in many tropical countries (Hansen et al. 2010, 2013). Although global studies demonstrate that protected areas tend to reduce anthropogenic land use change and help conserve biodiversity (e.g. Bruner et al. 2001; Andam et al. 2008; Joppa & Pfaff 2011; Geldmann et al. 2013), there is also a consensus that designation does not guarantee protection. Many protected areas are failing to entirely eliminate anthropogenic pressures within their boundaries (Dudley et al. 2004; Laurance et al. 2012), and illegal deforestation is not uncommon within protected tropical forests (Liu et al. 2001; Barber et al. 2012). While not all critics go as far as to call for alternative pathways to protect biodiversity (Mora & Sale 2011), it is now understood that the mere extent of protected areas performs poorly as an indicator of conservation status (Chape et al. 2005).

Third, forest protected areas have long been – and still are – a preferential funding target for conservation donors. In the beginning of the 21st century, international and national funding agencies continued to invest billions of dollars into expanding and consolidating protected area networks in the tropics (Gordon and Betty Moore Foundation 2006; GEF 2009; Kasparek et al. 2010). Studies suggest that international conservation funding would need to increase by an order of magnitude to designate and effectively manage all protected areas required to stem global extinction risks (Bruner et al. 2004; McCarthy et al. 2012). Hopes for new protected area funding are linked to the emergence of international carbon finance (Nauclér & Enkvist 2009; Parker et al. 2009) and prospects for the establishment of an international financing mechanisms to Reduce Emissions from Deforestation and Forest Degradation (REDD) (Angelsen 2008; Corbera & Schroeder 2011). However, as new studies began to highlight the significant climate mitigation potential of tropical protected areas (Nelson & Chomitz 2009; Scharlemann et al. 2010; Soares-Filho et al. 2010), the REDD debate also drew attention to the fact that protected areas are not the only strategy to protect tropical forests (Angelsen et al. 2009; Angelsen 2010) – and perhaps not even the most cost-effective.

Given the scale of the global protected area network, its mixed performance record, and its continued international financial backing, it is not surprising that the management quality and performance of its constituents have come under increased scrutiny (Salafsky et al. 2002; Ferraro & Pattanayak 2006; Keene & Pullin 2011). Donors, government agencies, field staff, and other project implementers are increasingly expected to demonstrate that protected areas generate desired results, and that investments are being allocated in ways that generate the most positive impacts. The resulting demands for information were grouped under the umbrella term of "protected area management effectiveness" and equipped with an official analytical framework by the International Union for the Conservation of Nature's (IUCN) World Commission of Protected Areas (WCPA) (Hockings et al. 2006).

While diverse and distinct in their goals and approaches, most PAME evaluations tend to be driven by a set of similar questions of interest to decision makers:

1) How effective are (given) protected areas at achieving their objectives?

2) What kind of interventions will make (given) protected areas more effective?
   And, occasionally,

3) Did a (given) intervention improve the effectiveness of (given) protected areas?

The apparent simplicity of these questions belies the fact that they usually cannot be answered accurately without extensive, complex, and costly analyses. For instance, all protected areas, as per IUCN's definition, aim to "achieve the long-term conservation of nature with associated ecosystem services […]" (Bertzky et al. 2012:3). Even a descriptive statement on the current performance of a protected area seems therefore deficient without a comprehensive field-based collection of biodiversity indicators: a time-consuming activity whose cost-effectiveness has often been placed in doubt (Sheil 2001, 2002; Gardner et al. 2008). Predictions of intervention impact require an even more extensive body of knowledge on the relative impact of different intervention types on diverse biodiversity outcomes of different types of protected areas in different contexts; such a body of knowledge, however, does not exist in the literature (Ferraro et al. 2011). Finally, even a single estimate of the conservation impact of a single intervention in a single protected area can be difficult to achieve if study design and case do not permit the attribution of observed outcomes to the intervention (Ferraro 2009). In reality, conservation interventions rarely meet the preconditions needed for rigorous retrospective impact estimation (Margoluis et al. 2009).

Given the substantial challenges involved in the rigorous assessment of management effectiveness and the prediction and verification of intervention impacts, it comes as no surprise that many designers and implementers of PAME evaluations content themselves with simpler, more affordable, and less rigorous assessment methodologies than those desired by academics. The vast majority of PAME evaluations consists of little more than a compilation of readily available management indicators, occasionally supported by one or two days of stakeholder workshops and, sometimes, field visits (Leverington et al. 2007, 2010; Nolte et al. 2010). All of the four internationally most widespread PAME evaluation methods[2] come in the format of a simple standardized scorecard, whose indicators cover a wide range of management aspects, including protected area law, planning, finances, staff, equipment, infrastructure, stakeholder participation, and research, among others. The implicit underlying assumption seems to be that those indicators, individually or collectively, are reasonably good proxies for the actual effectiveness of protected areas in achieving its objectives – and thus sufficiently accurate for monitoring protected area performance, prioritizing interventions, and evaluating project success.

The main caveat with this approach is that this important assumption is rarely explicitly tested. Instead, the credibility of most PAME evaluation methods rests on the procedure by which they have been developed – usually by consortia of protected area experts, often in conjunction with stakeholder consultations and pilot tests – and their endorsement by many of the world's largest conservation donors and organizations[3]. Furthermore, most widespread PAME evaluation

[2] Rapid Assessment and Prioritization of Protected Area Management (RAPPAM), Management Effectiveness Tracking Tool (METT), Parks in Peril Site Consolidation Scorecard (PiP SCIS), Scorecard used by the Programa Ambiental Regional para Centroamerica / Central American Protected Areas System (PROARCA/CAPAS)

[3] Such as the Global Environment Facility (GEF), the World Bank, the World Wide Fund for Nature (WWF), the U.S. Agency for International Development, The Nature Conservancy (TNC), as well as many national protected area agencies around the tropics.

methods tend to measure similar management aspects (see above), implying a tacit consensus that those aspects are important for the capacity of a protected area to achieve its objectives.

However, there are also reasons to question the informative value of such PAME evaluation scores. First, it remains usually unclear how the collected scores are to be interpreted. Most PAME scorecards measure a large number of management aspects (up to 90[4]). Some of these aspects are likely to be more important than others in shaping site performance; others can be expected to substitute or complement each other – or interact in even more complex ways. However, most widespread PAME evaluation methods either treat all indicators or do not provide any information on the presumed relationship between scores and performance. Second, data quality is often a concern. Most widespread PAME evaluation methods use qualitative indicators of management – often with context-specific qualifiers such as "adequacy", "appropriateness" and "sufficiency" – and rely heavily on protected area managers as information sources (Cook & Hockings 2011). Eliciting what are essentially the opinions of individuals with a potentially strong interest in specific assessment results, this approach not only seems vulnerable to multiple sources of bias; it can also undermine the reliability of comparisons across sites and time periods.

In sum, more and better empirical research along at least two new strands is needed to support the evaluation and improvement of protected area management in tropical forests. One is about examining whether PAME evaluation methodologies that are currently being extensively applied across the globe provide sufficiently good surrogates for a protected area's performance in achieving its conservation objectives. The second is about identifying those factors which are, in fact, closely associated with protected area performance, and may thus not only be suitable as performance indicators, but also, potentially, as targets for conservation interventions[5].

This dissertation comprises four empirical analyses designed to enhance the current state of knowledge in both of these research directions. Together, they examine the relations between governance, management indicators, and conservation outcomes of selected protected areas in

---

[4] In the Rapid Assessment and Prioritization of Protected Area Management (RAPPAM) method

[5] Note: Confirming the suitability of a management aspect as a target of conservation interventions would require additional empirical research establishing at least a positive causal link between intervention, management change, and desired conservation outcome, if not also an estimate of the relative size of that effect (as compared to other interventions), and the relative cost involved in achieving it. However, this third research strand is outside the scope of this dissertation.

Colombia and Brazil, two of the most bio-diverse nations on the planet (GEF 2008). Although the four analyses are related in aim and scope, each represents an independent piece of research, and has been published or submitted as a stand-alone journal article (see chapters for references).

Following this introduction, **Chapter 2** is the first to shed empirical light on the relation between commonly used management indicators and the conservation success of tropical protected areas. Specifically, it examines whether scores of the Management Effectiveness Tracking Tool (METT) – the most widespread method to track PAME of individual protected areas over time, using 30 management indicators – are associated with observed effects on fire occurrence in the Brazilian Amazon rainforest. Employing two different econometric estimation strategies, I do not find strong relations between METT scores and protected area impact on forest fires. I conclude that METT may be falling short of its potential as an indicator for the capacity of a protected area to reduce undesired land-use changes – a result that could raise eyebrows at the World Bank, the Global Environment Facility (GEF), and other major donors requiring METT reporting as a component of their protected area support projects.

Failing to use METT scores for the prediction of conservation outcomes, I examine in **Chapter 3** the extent to which protected area success can be explained by three key variables that could potentially overshadow the importance of local management, and for which we could find reliable indicators: Location, governance type, and external government enforcement. Using a rich spatial dataset, including fine-resolution (~30m) multi-temporal satellite imagery, I find that both location and governance are highly significant predictors of avoided deforestation for 292 protected areas and indigenous lands in the Brazilian Amazon. We observe that strict protected areas consistently avoided more deforestation than sustainable use areas; and indigenous lands were particularly effective at avoiding deforestation in locations exposed to high deforestation pressures. We conclude that location and governance are two variables that seem to hold strong explanatory power for the conservation success of protected areas.

In the light of these insights, I revisit and significantly enhance the prior indicator analysis. **Chapter 4** not only incorporates governance and location in determining avoided deforestation success of tropical protected areas, and fine-tunes the statistical methodology developed in both previous chapters. It also makes use of a rich panel dataset of management indicators from the Rapid Assessment and Prioritization of Protected Area Management (RAPPAM) tool, the PAME assessment method used most widely to support the prioritization of protected area interventions within countries. Using 90 management indicators for 66 protected areas in the Brazilian Amazon and four different model parameterizations, I find no statistically significant associations between avoided deforestation and indicators that reflect preferential targets of conservation investments (e.g. budget, staff, equipment, management plans, and stakeholder collaboration). Instead, I observe the absence of unsettled land tenure conflicts to be strongly and consistently associated with avoided deforestation – as the only of 90 indicators. I conclude that unresolved land rights appear to overshadow the potential importance of most other management factors in shaping deforestation in protected areas and, together with location and governance type, deserve special attention as determinants of the success of forest protected areas.

In the quest for management indicators that better predict conservation outcomes, **Chapter 5** proposes and applies a new methodology for measuring what I assume, a priori, to be a crucial determinant of conservation success: the capacity of a protected area to enforce its regulations. Analyzing verifiable incident data through field visits and park staff workshops in 15 Colombian national parks, I observe that the capacity to enforce regulations against priority threats is very low in most visited parks. When it comes to threats linked to deforestation (clearing, farming, grazing, and construction), the analyzed parks are more likely to face challenges in sanctioning rule violators than in detecting violators. Furthermore, park guards most often relate those sanctioning challenges to unsettled land rights issues prevailing within the boundaries of their park. This potential link between unsettled land rights and failures to enforce regulations offers a hypothetical causal pathway for the close association between land rights issues and avoided deforestation observed in the Brazilian Amazon.

Chapter 6 concludes and identifies new research directions with the potential to improve our knowledge on the effectiveness, evaluation, and improvement of protected area management in tropical forests.

CHAPTER 2: LINKING MANAGEMENT EFFECTIVENESS INDICATORS OF PROTECTED AREAS TO EFFECTS ON FIRE OCCURRENCE IN THE AMAZON RAINFOREST[6]

Abstract

We examined whether management effectiveness indicators of protected areas are associated with the effectiveness of protected areas in reducing fire occurrence in the Amazon rainforest. Management effectiveness scorecards are widely used by donors and implementers to prioritize, track and evaluate protected area investments. However, there is little evidence whether these scorecards measure what they are assumed to measure: the capacity of protected areas to deliver conservation outcomes. We use data collected with the Management Effectiveness Tracking Tool (METT) scorecard, adopted by some of the world's largest conservation donors to track management indicators believed to be crucial for protected area effectiveness. Our outcome of interest is the occurrence of forest fires during the period 2000-2010 as a proxy for deforestation, the major driver of land-cover change and carbon emissions in the Amazon Basin. We use matching to compare the estimated effect of protected areas with low vs. high METT scores on fire occurrence. We also estimate effects on fire occurrence of individual protected areas and explore their associations with METT scores. Results indicate that the associations between METT scores and effects of protected areas on fire occurrence are weak. Across our sample, protected areas with higher 2005 METT scores do not seem to have performed better at reducing fire occurrence within their boundaries over the last decade. Further research into the associations between management effectiveness indicators and effects on conservation outcomes seems necessary, and our analysis offers new insights on the applicability of matching methods for that purpose.

---

[6] This chapter has been published as: Nolte C, Agrawal A (2013) Linking Management Effectiveness Indicators of Protected Areas to Effects on Fire Occurrence in the Amazon Rainforest. **Conservation Biology** 27(1):155-65

Introduction

Protected areas are one of the most prominent tools for conserving biodiversity worldwide. By 2011, national governments and international organizations had reported more than 160,000 sites "recognized, dedicated and managed […] to achieve the long term conservation of nature" (UNEP-WCMC 2011). In the 21st century, major conservation donors continue to dedicate substantial levels of financial resources to the establishment and consolidation of protected area networks (Gordon and Betty Moore Foundation 2006; GEF 2009; Kasparek et al. 2010). Put forward as a potentially effective means to reduce emissions from deforestation and forest degradation (Trumper et al. 2009; Soares-Filho et al. 2010), protected areas can be expected to attract conservation funds for decades to come.

Given their substantial and growing financial commitments, conservation donors and implementers have come under pressure to demonstrate that investments into protected areas are worth the money spent (Ferraro & Pattanayak 2006; Fuller et al. 2010; Mascia & Pailler 2011). Reacting to such concerns, parties to the Convention on Biological Diversity (CBD) committed to evaluating the management effectiveness of 30% of their protected areas by 2010 (CBD 2004). Guiding this effort, the World Commission of Protected Areas (WCPA) defined protected area management effectiveness (PAME) as being about more than conservation outcomes. Instead, PAME evaluations incorporate a wide range of management themes, including the adequacy of protected area design, planning, resources and processes (Hockings et al. 2006). Major donors such as the Global Environment Facility (GEF), the World Bank, the Gordon and Betty Moore Foundation (GBMF), and the United States Agency for International Development (USAID), and implementers such as The Nature Conservancy (TNC), and the World Wide Fund for Nature (WWF) have used PAME to prioritize, track, or evaluate their investments in protected areas. Recent global and regional surveys have recorded more than 9000 PAME assessments in 140 countries (Leverington et al. 2010; Nolte et al. 2010).

Although PAME evaluations have been driven by an ultimate interest in the effects of protected areas, i.e. changes in outcomes that can be attributed to the existence of a protected area and its management (Ferraro 2009; Joppa & Pfaff 2011), the focus seems to have shifted in practice. Among the three most widely used PAME scorecards, only one contains an outcome indicator (Table 1). None provides a framework to interpret changes in outcomes as an effect of the examined protected area or its management strategies, e.g. by estimating what outcomes would have been observed in the absence of protection or management (counterfactual) (Ferraro 2009). Instead of looking at effects, most PAME scorecards collect a multitude of indicators on the management capacity of a protected area, such as the adequacy of its budget, staffing, planning processes, participation, and enforcement. The worldwide adoption of this approach seems to rest on the assumption that these indicators provide a reasonably good proxy for the extent to which a protected area is effective in delivering desired outcomes.

Table 1: Most widely used scorecards to assess protected area management effectiveness (applied in > 10 countries)

Name	Mainly used by	Coverage [a]	Outcome indicators	Ref.
Management Effectiveness Tracking Tool	Global Environment Facility, World Bank	100 countries [b] (~ 2000 assessments) [b]	One indicator (see Table 3)	Stolton et al. 2007
Rapid Assessment and Prioritization of Protected Area Management	World Wide Fund for Nature	49 countries (> 1600 assessments)	Absent	Ervin 2003
Parks in Peril Site Consolidation Scorecard	U.S. Agency for International Development, The Nature Conservancy	15 countries (323 assessments)	Absent	Martin & Rieger 2003

[a] Data from Leverington et al. (2010) unless stated otherwise.

[b] Data from Coad et al. (2011)

However, this assumption has barely been explicitly tested. There are studies that analyze PAME data at the global level and identify global patterns in protected area management, e.g. the chronic inadequacy of protected area budget, staff, infrastructure, and community relations (Dudley et al. 2007; Leverington et al. 2010). However, the absence of outcome indicators has limited the ability of these studies to illuminate the relation between PAME scores and the effectiveness of protected areas in delivering conservation outcomes. One recently published study does investigate this relationship: a report by the GEF evaluation team. Juxtaposing scores of the Management Effectiveness Tracking Tool (METT), the world's most widely used PAME scorecard, with species data and expert opinion on 11 wildlife reserves in Zambia, the authors find a positive correlation between an increase in METT scores and increases in species populations (Zimsky et al. 2010). However, the study does not control for confounding factors that could have caused the observed changes in outcomes.

Our analysis empirically examines the associations between METT scores and the effect of protected areas on the occurrence of forest fires in the Amazon rainforest. Recognized as a global priority area for the conservation of both biodiversity (Brooks et al. 2006) and carbon stocks (Saatchi et al. 2007), the Amazon Basin has attracted considerable amounts of conservation funding over the last decade. Forest fires in the Amazon are predominantly of anthropogenic origin (Cochrane 2011; Pivello 2011), and exhibit close spatial associations with deforestation (Aragão et al. 2008; Nelson & Chomitz 2011). They therefore provide a good proxy for the type of land cover change that protected areas seek to reduce. For the purpose of this study, we define "effectiveness" of a protected area as the extent to which the occurrence of fires on forest parcels within its boundaries is lower than that on similar unprotected forest parcels. We use matching, a statistical technique widely used to estimate protected area effectiveness (Andam et al. 2008; Joppa & Pfaff 2011; Nelson & Chomitz 2011), to compare the estimated effect of protected area groups with low vs. high METT scores on fire occurrence (see Methods). We then estimate effects on fire occurrence of individual protected areas and explore the associations of these estimates with METT scores.

Methods

Study Area

Our analysis focuses on the tropical and subtropical moist broadleaf forests in the Amazon Basin as defined by WWF's Terrestrial Ecoregions of the World (Olson et al. 2001). Within this region, we selected all ~1km² parcels that contained at least 25% forest cover as estimated by the Vegetation Continuous Fields (VCF) algorithm (Hansen et al. 2003), and had been classified as forest or forest mosaic by the ~1km resolution Global Land Cover (GLC) dataset produced by the European System for Earth Observation (Bartholome & Belward 2005). The intersection of VCF and GLC produces a conservative estimate of tropical forest area, chosen to limit the risk of including observation of fire occurrence unrelated to deforestation, e.g. in tropical savannas or on land that was already cleared of forest or used predominantly for agriculture (Nelson & Chomitz 2011). GLC is only available for the year 2000, which we consequently chose as the starting year of our period of analysis (2000-2010).

Outcome Data: Fires In Forests

We extracted our fire data from the Active Fires product of the Moderate Resolution Imaging Spectroradiometer (MODIS), which provides globally consistent daily estimates of the location and intensity of active fires at ~1km resolution since October 2000 (Justice 2002). Following Morton et al. (2008), we only extracted fires occurring at night and daytime fires with >330K brightness in the 4μm channel ("high-confidence fires") that were observed between 2000 and 2010 on forest parcels in our study area. Our outcome variable is binary: did MODIS detect at least one high-confidence fire on a given forest parcel between 2000 and 2010?

Effectiveness Data

The Management Effectiveness Tracking Tool (METT) is an assessment scorecard initially developed by the WWF / World Bank Forest Alliance to assess PAME of forest protected areas (Stolton et al. 2007). METT has been applied across different ecosystems as a reporting requirement of the World Bank and the Global Environmental Facility (GEF). Respondents are usually accountable to these donors, and include protected area managers, project staff, consultants, and management councils. A recent effort to compile all existing METT data (Coad et al. 2011) recorded >2000 METT assessments in >100 countries, making it one of the most widespread PAME assessment methods in the world.

METT consists of a 30-item scorecard that measures management aspects believed to be crucial for effective protected area management, including legal status and regulations, adequacy of budget, staff and resources, research and monitoring, and stakeholder relations. Respondents assign each indicator a score from 0 to 3, with qualitative statements providing indicator-specific guidance about the meaning of each number and low values generally reflecting lower performance (see Supplementary Information). Individual indicator scores are summed up to create the composite METT score which has been suggested as a possible proxy for overall management effectiveness (Dudley et al. 2007). METT also allows for the exclusion of indicators that are not applicable and for subsequent rescaling of the composite score. To account for such missing indicators, we calculated METT composite scores as the average of all indicators for which a value was reported (0-3).

We extracted all assessments carried out in 2005 or earlier from the most recent version of the Global METT Database maintained at the University of Oxford (Coad et al. 2011), keeping only the most recent scores for each protected area in the case of repeat assessments. Because many GEF/World Bank projects in the region were launched before METT became a reporting requirement, our sample contains scores for 41 protected areas:, 2 located in Bolivia, 6 in Peru, and 33 in Brazil. Although this sample is not random, it is geographically well distributed over the Amazon Basin.

Covariates and Treatment

The probability of protection and the probability of forest fires both depend on a number of variables that must be controlled when estimating the effectiveness of protected areas in reducing forest fires. Drawing on related assessments of the effectiveness of protected area networks in reducing tropical deforestation and fires (Andam et al. 2008; Joppa & Pfaff 2011; Nelson & Chomitz 2011), we included five covariates in our analyses: elevation, slope, travel time, distance to forest edge, and rainfall.

- Slope and elevation strongly influence whether a given location is suitable for different land uses, and can thus be expected to be associated with the probability that a given forest parcel will be converted to agriculture. We controlled for median elevation and average slope, extracted from spatial data layers of the Global Agro-Ecological Zones Assessment (Fischer et al. 2007).

- The probability of timber extraction and agricultural use is strongly associated by access to markets, a function of a parcel's distance to roads, rivers, and major cities. To account for market access, we used travel time estimates to major cities (>50,000 inhabitants) computed by the Joint Research Centre of the European Commission (Nelson 2008).

- Deforestation is more likely to occur close to the forest edge and to locations that previously were deforested. We used the above intersection of GLC and 25% VCF to define forest extent. We added a 1km buffer to smooth out small non-forest patches that were surrounded by forest and therefore less likely to be part of the agricultural frontier. Distance to forest edge was then computed as the Euclidian distance from each forest parcel to the closest non-forest parcel.

- Available moisture and precipitation influence the probability of both fire occurrence and fire detection. First, average annual rainfall can influence the suitability of a given parcel for agricultural production. Second, higher precipitation rates may be associated with higher frequencies of cloud coverage, which can inhibit the ability of remote sensors to detect fires beneath those clouds. We therefore used average annual precipitation rates provided by WorldClim to control for this covariate (Hijmans et al. 2005).

Geographical limits of protected areas were extracted from the World Database of Protected Areas (WDPA) (UNEP-WCMC 2011). Boundaries of countries and states were based on the Vector Smart Map (VMAP) Level 0 dataset as provided by the United States National Imagery and Mapping Agency (NIMA 2000). We used ArcGIS 10.0 to re-project and resample spatial data layers for all five covariates, protected areas and countries/states into the format used by MODIS Active Fires product (equal area sinusoidal projection, ~ 1km resolution), and extracted all variables into one table with 5.26 million forest parcels for the ensuing analysis.

Effects of Protected Area Groups with High and Low METT Scores

We divided our sample of protected areas with METT data into two groups with low vs. high composite METT scores in order to compare their respective effects on fire occurrence. We conducted separate analyses for the full sample and for the Brazilian subsample. We included all protected areas that had been designated in or prior to 2002, and chose thresholds that produced groups with a roughly similar number of protected areas (METT score threshold: 1.33 and 1.22 for the full and Brazilian sample, respectively). Because the total area of forest cover varies among protected areas, the number of forest parcels in each group varies. To test the sensitivity of our analysis to the choice of these threshold parameters, we also explored alternative group definitions, e.g. using different cutoff years, creating groups with similar forest extent, and limiting our analysis to the upper and lower quartiles of METT scores (see Appendix A).

We used nearest neighbor matching (NNM) to estimate effects on fire occurrence. Widely used to estimate protected area effectiveness (Andam et al. 2008; Joppa & Pfaff 2011; Nelson & Chomitz 2011), NNM is a statistical technique that mimics random assignment of treatment in observational data by matching each treated unit (forest parcel) to a unit from a pool of candidate control units whose covariates are similar to selected treatment unit. The difference in outcomes between treatment and artificial control group is assumed to reflect the average treatment effect on the treated (ATT). We measure our ATT of interest, the effect on fire occurrence, as the difference in the percentage of forest parcels with observed fires between treatment and control group.

We conducted four group comparisons. First, we estimated effects on fire occurrence of high- and low-METT protected areas as compared to the counterfactual of no protection, by matching forest parcels from either group to a third group of parcels that had never been protected. This estimation strategy is similar to that of Nelson & Chomitz (2011), who compared the effectiveness of different protected area categories (strict protection, sustainable use, indigenous lands) in reducing fire occurrence. However, we found that distributions of key covariates of forest parcels differed considerably between high- and low-METT protected areas (see Appendix A), reducing the extent to which differences in estimated effects on fire occurrences could be ascribed to differences in METT scores as opposed to differences in other characteristics. We therefore conducted two additional comparisons to achieve better covariate balance between forest parcels from high-METT and low-METT protected areas. We first matched forest parcels from high-METT protected areas to forest parcels from low-METT areas and compared differences in fire occurrences between the two. We then repeated this process for low- vs. high-METT protected areas. As matching was with replacement, we expected these two comparisons to generate different groups of forest parcels, and impact estimates (see Appendix A for further elaboration).

Our matching-based estimates of the effects of protected areas on fire occurrence require the assumption that the distribution of forest fires is well-behaved. Recent findings cast some doubt on this assumption. Areas burned by wildfires have been found to be power law distributed in a variety of different eco-regions, including the Amazon (Malamud et al. 2005; Pueyo et al. 2010), introducing potential spatial dependence in the likelihood of fire occurrence on neighboring forest parcels. While our method does not allow us to explicitly control for spatial autocorrelation, we followed earlier studies (Andam et al. 2008; Joppa & Pfaff 2011; Nelson & Chomitz 2011) in reducing the risk of spatial dependence by randomly sampling a small percentage of forest parcels (2%) from the entire population of forest parcels, and conducting our analysis on these samples.

We matched forest parcels in R using Sekhon's (2007) Matching library. We used Mahalanobis distance NNM with replacement and bias adjustment with average slope, average elevation, rainfall, distance to forest edge, and travel time as covariates. We required control parcels to be situated in the same country (Peru, Bolivia) or state (in Brazil) as the treatment parcels (exact matching). We dropped treatment parcels for which no nearest neighbor could be found within one standard deviation of each covariate (calipers). We repeated the process of random sampling and matching 30 times, averaging the estimated differences between treatment and artificial control groups. In line with earlier matching studies (Andam et al. 2008; Joppa & Pfaff 2011), we used the Abadie-Imbens variance formula, and reported the average standard error as the square root of the mean variance across 30 runs. We also tested whether our results were sensitive to the size of our calipers (see Appendix A).

Effects and METT Scores of Individual Protected Areas

Our second analysis examined the associations between METT scores and individual effects of protected areas on fire occurrence. We first developed protected area level estimates of the probability of fire occurrence in the absence of protection by matching forest parcels from each protected area to unprotected forest parcels and averaging fire occurrence on the latter. We used the same data, covariates, and matching parameters as in our first analysis, but applied calipers of 0.5 standard deviations. We sampled 5% of the forest parcels from each protected area and matched them to a 5% sample of unprotected forest parcels, repeating the sampling and matching process 30 times, and averaging the resulting fire probability estimates. Estimates were computed for each WDPA-reported protected area in Bolivia, Brazil and Peru that had been designated in or prior to 2002, contained at least 500 forest parcels and did not overlap with other protected areas. After excluding protected areas for which matching was considered unrepresentative (protected areas for which > 40% parcels had been dropped due to calipers), our final sample contained 182 protected areas, 29 of which had METT scores.

The distribution of our estimates of the probability of fire occurrence in the absence of protection was strongly skewed: While 9.8% of protected areas were estimated to have more than 50% of their forest parcels exposed to fires in the absence of protection, 44% of protected areas were estimated to have less than 1% of forest parcels affected by fire if unprotected (Fig. 1). Such variation poses challenges when it comes to defining which of these protected areas are more effective in reducing fire occurrence. Where fire probabilities in the absence of protection are high, protected areas may have had a considerable effect on fire occurrence, although fires may still be frequent as compared to protected areas with low fire probabilities in the absence of protection. The latter, in turn, may not have been exposed to forest fires at all, but cannot claim to have had major effects on fire occurrence either.

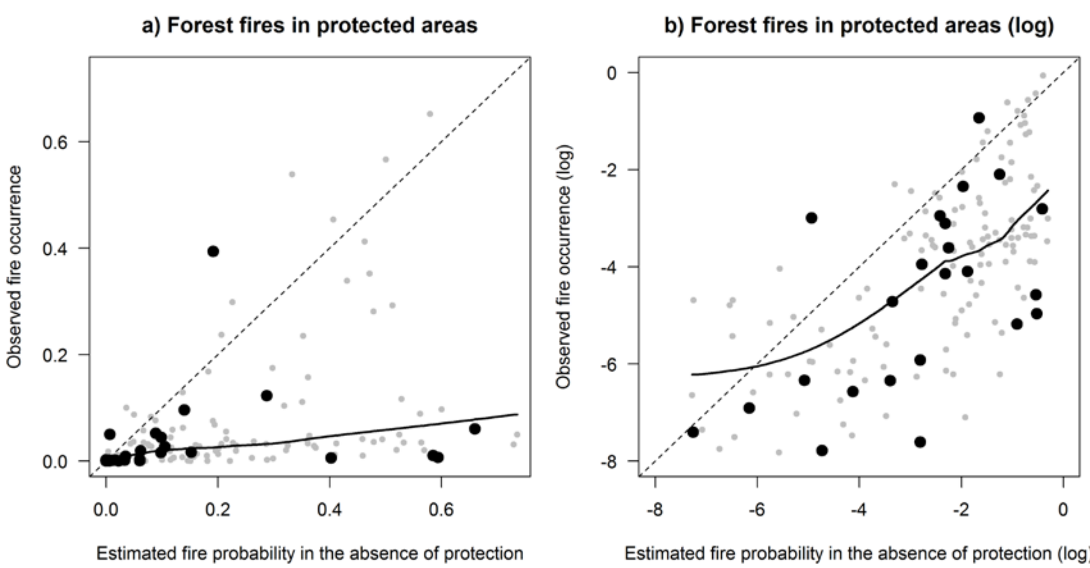

Figure 1: Estimated probability of fire occurrence in the absence of protection and observed fire occurrence of 182 forest protected areas in Peru, Bolivia and Brazil with (black dots) and without (grey dots) Management Effectiveness Tracking Tool (METT) data. Black line: Estimated fire occurrence within protected areas conditional on estimated probability of fire occurrence in the absence of protection (LOESS estimation).

We defined protected areas as being “more effective” if they had reduced fire occurrence to a greater extent than other protected areas with similar probabilities of fire occurrence in the absence of protection. Put differently, we asked: “Given the estimated probability of fire occurrence in the absence of protection, what actual fire occurrence would we expect to observe in a protected area?". We estimated expected fire occurrence in protected areas conditional on estimated fire probabilities non-parametrically by fitting a locally weighed scatter plot smoothing function (LOESS, span=0.75) to the full sample of 182 protected areas. We divided the sample into two groups: Protected areas with fire occurrences below and above the threshold defined by the LOESS function (black line in Fig. 1) were defined as belonging to the "more effective" and "less effective" category, respectively. In addition, we considered three alternative definitions of the relative effectiveness of protected areas in reducing fire occurrence (Table 2).

Table 2: Four alternative definitions of the relative effectiveness of protected areas in reducing the occurrence of forest fires

Definition	**Definition of "more effective"**	**Split variable**	**Split at (n_{low}:n_{high})**
Fire occurrence	Fire occurrence is lower than in other protected areas	Fire occurrence	Median of full sample (11:18)
Fire reduction	Absolute effect on fire occurrence is higher than in other protected areas	Fire probability in the absence of protection – fire occurrence	Median of full sample (18:11)
% Fire reduction	Relative effect on fire occurrence is higher than in other protected areas	Fire occurrence / fire probability in the absence of protection	Median of full sample (14:14)
LOESS	Fire occurrence is lower than that of protected areas with comparable fire probability in the absence of protection	Fire occurrence	LOESS curve (9:20)

Given our small sample size, we used a simple two-step approach to explore associations between METT scores and effects of individual protected areas on fire occurrence: First, we compared the differences in composite METT scores between "more effective" and "less effective" groups of protected areas. Second, we compared the differences of selected indicators from the METT scorecard between "more effective" vs. "less effective" protected areas. For each indicator, we discarded protected areas for which no score was provided. We used two-tailed t-tests to test for significance between the differences in score means between groups. With 18 score indicators and 29 observations, our objective in doing so was not to claim that differences in scores were causally related to relative effectiveness, but to explore patterns within the data that could direct further research.

Results

Group matching results suggest that both high and low-METT protected areas reduce the occurrence of fires within their boundaries relative to similar unprotected areas (Fig. 2). However, results did not offer clear evidence that fire occurrence was lower in high-METT protected areas than in low-METT protected areas. In our full sample, the estimated effect on fire occurrence of high-METT protected areas was considerably *smaller* than for low-METT protected areas (-2.6% vs. -7.3%). The opposite pattern obtained for the Brazilian sample (-7.6% vs. 3.5%). The discrepancy seems to be due to three Brazilian protected areas with strong effects on fire occurrence (i.e. high estimated fire probabilities in the absence of protection, but low actual fire occurrence) switching groups as the METT threshold changes from 1.33 to 1.22. Comparing only the upper vs. lower METT quartiles of protected areas resulted in roughly similar effect estimates of high and low-METT groups (-2.7% vs. -2% in the full sample, -3.3% vs. -3.6% in the Brazilian sample, see Appendix A).

Matching forest parcels from the high-METT group to the low-METT group retained only 64% (full sample) and 83% (Brazilian sample) of the observations, as matched pairs were dropped if not sufficiently similar (Fig. 2). After matching, differences in fire occurrence on high vs. low-METT parcels fell below 1%. Matching low-METT to high-METT parcels suggested that the former may even had a greater effects on fire occurrence than the latter (2.0% and 2.1%), although this estimate was based on less than half of the low-METT parcels.

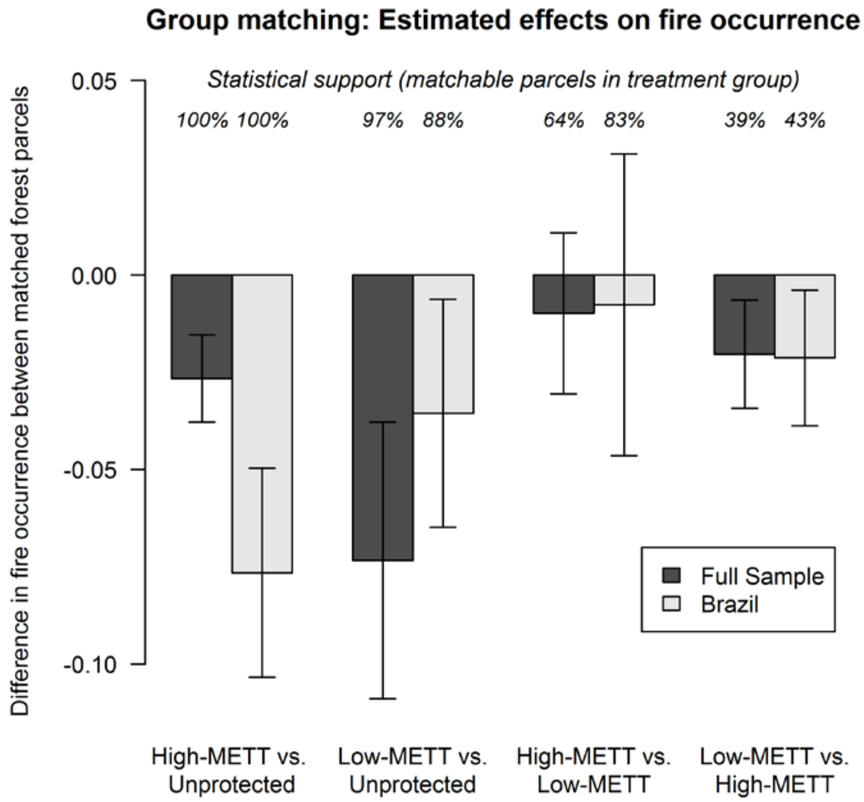

Figure 2: Estimated effects on fire occurrence of protected area groups with high and low composite scores of the Management Effectiveness Tracking Tool (METT). Groups were split at a composite METT score of 1.33 (full sample) and 1.22 (Brazilian sample). Error bars proxy confidence intervals (average SE * 1.96).

Our second analysis assigned the same weight to each protected area, and thus, METT score. Results, however, are similar: Although METT composite scores varied substantially among protected areas (mean: 1.33, SD: 0.41), differences in METT composite scores between "more effective" and "less effective" protected areas were small (Fig. 3). The highest absolute difference in average composite METT scores (1.21 vs. 1.39) was observed using our preferred definition of the relative effectiveness of protected areas in reducing fire occurrence (LOESS). However, the difference was not found to be statistically significant even at the 0.25 confidence level.

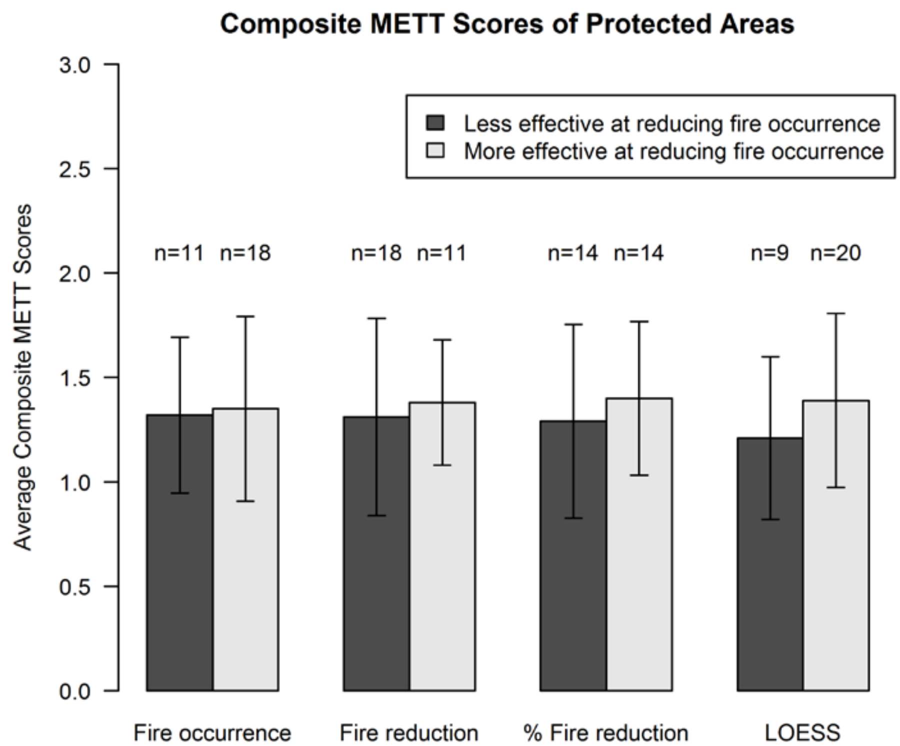

Figure 3: Composite scores of the Management Effectiveness Tracking Tool (METT) for protected area groups considered "more effective" and "less effective" according to our four definitions of relative effectiveness in reducing fire occurrence. Error bars show standard deviations of scores for each group.

Although composite METT scores did not exhibit strong associations with the effectiveness of protected areas in reducing fire occurrence, several individual indicators of the METT scorecard did (Table 3). Given the ratio between indicators and observations, much of this variance could be the result of random variation. However, "more effective" areas in our sample tended to have higher subjective ecological-condition scores (Indicator 27). Decision makers in such protected areas are more likely to cooperate closely with neighboring official and commercial land users (Indicator 21), but not allow for input into management decisions from local communities (Indicator 23). "More effective" protected areas are also likely to have more research activities (Indicator 10) and better access control (Indicator 28).

Table 3: Associations of indicators of the Management Effectiveness Tracking Tool with relative effectiveness in reducing fire occurrence*

			Definition of relative effectiveness			
Category	**Scorecard indicator**	**Question**	**Fire occurrence**	**Fire reduction**	**% Fire reduction**	**LOESS**
Resources	12: Staff numbers	Are there enough people employed to manage the protected area?	(-)			
	15: Current budget	Is the current budget sufficient?				
	16: Security of budget	Is the budget secure?	(+)	-		(+)
	18: Equipment	Are there adequate equipment and facilities?			+	
Institutions	2: Regulations	Are inappropriate land uses and activities controlled?	(-)			
	3: Law enforcement	Can staff enforce protected area rules well enough?		(+)		
	6: Boundary demarcation	Is the boundary known and demarcated?				
	28: Access assessment	Is access / resource use sufficiently controlled?		(+)	+	(+)
Information	30: Monitoring and evaluation	Are management activities monitored against performance?				
	10: Research	Is there a program of management-oriented survey and research work?			(+)	++
	9: Resource inventory	Do you have enough information to manage the area?				(+)

Category	Scorecard indicator	Question	Fire occurrence	Fire reduction	% Fire reduction	LOESS
			Definition of relative effectiveness			
Planning	7: Management plan	Is there a management plan and is it being implemented?				
	8: Regular work plan	Is there an annual work plan?				
	17: Budget management	Is the budget managed to meet critical management needs?		++	+	(+)
Relationships	21: State and commercial neighbors	Is there co-operation with adjacent land users?	++		(+)	++
	22: Indigenous people	Do indigenous and traditional peoples resident or regularly using the protected area have input into management decisions?				
	23: Local communities	Do local communities resident or regularly using the protected area have input into management decisions?			--	(-)
Condition	27: Condition assessment	What is the condition of the important values of the protected area as compared to when it was first designated?	(+)	(+)	+	+

* Significance of differences between individual scores between "more effective" and "less effective" groups (two-tailed t-test). Positive associations: +++ ($p < 0.01$), ++ ($p < 0.05$), + ($p < 0.1$), (+) ($p < 0.25$). Negative associations: -- ($p < 0.05$), - ($p < 0.1$), (-) ($p<0.25$).

Second, a number of individual METT indicators do not exhibit observable or consistent differences between “more effective” and “less effective” groups. This was particularly true for management aspects traditionally assumed to be closely related to protected area effectiveness, and thus classical targets of conservation investments, including adequacy (Indicator 15) and security of budget (Indicator 16), staff numbers (Indicator 12), management plans (Indicator 7), and boundary demarcation (Indicator 6).

Third, the behavior of institutional variables was inconsistent: Although controlling access or use of the protected area (Indicator 28) was positively associated with effectiveness, other variables such as mechanisms for controlling inappropriate land use and activities (Indicator 2), as well as capacities and resources of staff to enforce regulations and legislation (Indicator 3) had weak or negative associations with the relative effectiveness of protected areas in reducing fire occurrence.

Discussion

Our analysis of the associations between PAME scores and the effectiveness of protected areas in reducing fire occurrence was motivated by the goal of refining and improving existing strategies to measure and track management effectiveness. However, using fire data as a proxy for deforestation in the Amazon Basin, we did not find strong associations between the composite METT scores used for reporting and our multiple definitions of the relative effectiveness of protected areas in reducing fire occurrence. At least for the protected areas in our sample, METT seemed to fall short of its potential as an indicator for the capacity of a protected area to reduce the extent of undesired land use changes.

Certainly, such correlation is not causality. Our failure to observe significant differences in 2005 METT scores among protected areas which have shown to be more *vs.* less effective in reducing fire occurrence between 2000 and 2010 could be a result of conservation actors adapting their support strategies as a function of protected area success. For example, it is plausible that support for protected areas has systematically targeted underperforming protected areas within our time period of interest.

Although our findings do not allow us to establish claims of causality, the lack of observed associations between management effectiveness indicators of protected areas and their effectiveness in reducing forest fires is illuminating. Developed by long-standing protected area experts, METT has been endorsed by major conservation donors as a mandatory evaluation tool, making it a de-facto standard for assessing PAME. If METT scores do not actually serve as a good proxy for the capacity of a protected area to reduce undesired land use change, results of project evaluations that rely on METT scores may be biased. A greater concern is that METT and similar management characteristics-based evaluations may create incentives for project implementers to invest in activities that improve effectiveness scores without necessarily making a protected area more effective in terms of conservation outcomes.

Given the widespread use of PAME scores in conservation projects and policy worldwide, it seems necessary to direct further efforts into understanding the relation between protected area management, protected area effectiveness, and the indicators used to measure both. Future studies should examine the strength of associations between PAME indicators and effectiveness estimates of protected areas in other eco-regions, and using data from other widespread PAME methodologies (e.g. RAPPAM and PiP SCS, see Table 3). Insights would allow evaluators to learn which indicators are more closely associated with effectiveness, and adapt existing evaluation methods accordingly. The widespread use of PAME scores for accountability purposes also justifies a renewed quest for indicators that are cheap-to-verify, costly-to-fake (Ferraro 2008) and possibly more objective than the existing judgments of “adequacy” (see Table 3) which can differ considerably across respondents, protected areas, and time.

If we want to understand why some areas are effective and what type of support makes them effective, however, our analyses will need to move from correlation to causation. Did protected areas that received a specific type of support reduce undesired land use changes to a larger extent than those that did not receive the same support – even if support allocation may be influenced by expected effects? The large number of protected areas and support projects around the world makes it increasingly possible to construct such counterfactual evidence for a number of management interventions, an approach that promises to provide strong evidence for the relative effectiveness of such investments.

Finally, our analysis offers new methodological insights that can help improving the utility of matching methods in estimating the relative effectiveness of protected areas. We show that studies comparing effect estimates of different protected area groups vs. untreated units (e.g. Nelson & Chomitz 2011) can conflate potential differences in the probability of undesired land use change in the absence of protection with the effectiveness of protected areas in reducing them. Between-group matching allows the analyst to single out each of these two estimates, thus providing a better estimate of differences in the relative effectiveness between groups of interest. In addition, our approach to compute effectiveness estimates at the protected area level allows for comparisons that assign the same weight to each protected area (and METT score), and are thus less vulnerable to differences in the size of protected areas. Our analyses suggest that while matching is certainly not a methodological panacea, it can, if carefully designed, become a useful tool to examine whether protected areas are effective at delivering conservation outcomes, and why.

CHAPTER 3: GOVERNANCE REGIME AND LOCATION INFLUENCE AVOIDED DEFORESTATION SUCCESS OF PROTECTED AREAS IN THE BRAZILIAN AMAZON[7]

Abstract

Protected areas in tropical countries are managed under different governance regimes, whose relative effectiveness in avoiding deforestation has been the subject of recent debates. Participants in these debates answer appeals for more strict protection with the argument that sustainable use areas and indigenous lands can balance deforestation pressures by leveraging local support to create and enforce protective regulations. Which protection strategy is more effective can also depend on 1) the level of deforestation pressures to which an area is exposed, and 2) the intensity of government enforcement. We examine this relationship empirically, using data from 292 protected areas in the Brazilian Amazon. We show that for any given level of deforestation pressure, strictly protected areas consistently avoided more deforestation than sustainable use areas. Indigenous lands were particularly effective at avoiding deforestation in locations with high deforestation pressure. Findings were stable across two time periods featuring major shifts in the intensity of government enforcement. We also observed shifting trends in the location of protected areas, documenting that between 2000 and 2005 strictly protected areas were more likely to be established in high-pressure locations than sustainable use areas and indigenous lands. Our findings confirm that all protection regimes helped reduce deforestation in the Brazilian Amazon.

---

[7] This chapter has been published as: Nolte C, Agrawal A, Silvius KM, Soares-Filho BS (2013) Governance regime and location influence avoided deforestation success of protected areas in the Brazilian Amazon. **Proceedings of the National Academy of Sciences** 110(13):4956-4961

Introduction

Terrestrial protected areas, an integral component of biodiversity conservation policy, have also become a centerpiece of global efforts to reduce carbon emissions from tropical deforestation (Scharlemann et al. 2010). In the past decade, governments across the tropical biome have continued to expand their protected area networks (Bertzky et al. 2012), while international donors pledge billions of dollars for forest-based climate change mitigation (Collaborative Partnership on Forests 2012; Parker et al. 2012). Situated at the overlap between multiple global and local interests (Sunderland et al. 2008; Hirsch et al. 2010), protected areas are managed under a wide range of governance regimes to achieve better ecological and social outcomes. Although all these regimes establish some form of spatially explicit restrictions on land use and resource extraction, such restrictions can vary substantially (Dudley 2008).

A common distinction between governance regimes is that between strictly protected areas that discourage consumptive resource use or even physical access, and sustainable use areas that allow for controlled resource extraction, land use change, and in many instances human settlements (Nelson & Chomitz 2011). Indigenous lands, established primarily to safeguard the rights and livelihoods of indigenous people, are put forward as a third type of protected areas with considerable potential to contribute to climate change mitigation (Nepstad et al. 2006). Recent prospects of international carbon payments tied to avoided deforestation have reignited the interest of donors and governments to understand the extent to which each of these governance arrangements are effective in helping conserve tropical forest carbon (Angelsen 2010; Ferraro et al. 2011).

Keen theoretical debates surround the extent to which controlled resource use in protected areas can reduce deforestation. Proponents of strict conservation have long argued that ruling out resource extraction coupled with enforcement by protected area guards is more likely to be effective at achieving conservation than more inclusionary approaches (Oates 1999; Terborgh 2004; Hilborn et al. 2006; Laurance et al. 2012). Other contributors highlight that such enforcement has often proved insufficient to inhibit extraction in tropical parks (Infield & Namara 2001; Robinson et al. 2010; Petursson et al. 2012), and that forest-dependent communities, including indigenous people, can have stronger incentives than disinterested or understaffed government agencies to protect their livelihood base against externally-driven deforestation pressures (Gibson et al. 2005; Hayes 2006; Chhatre & Agrawal 2008). From this latter perspective, allowing controlled resource use in protected areas can help leverage local support for creating and enforcing regulations against externally driven deforestation pressures (Waylen et al. 2010; Porter-Bolland et al. 2012). Supporting indigenous communities in their efforts to demarcate and manage their territories promises similar synergies (Schwartzman & Zimmerman 2005).

Although these lines of argument differ, authors commonly identify two contextual factors as influencing the advantages of one protection regime over the other: 1) the willingness and capacity of government agencies to enforce conservation regulations and 2) the intensity of deforestation pressures to which a given area is exposed. Whether and how the relative effectiveness of protection regimes varies along these contextual dimensions, however, remains poorly understood. High-pressure locations, for instance, may prove particularly challenging for strict protected areas that lack local constituencies (Pedlowski et al. 2005), but could facilitate external enforcement because of greater accessibility and lower travel costs (Börner et al. 2011). Indigenous actors have been characterized both as weak (Vuohelainen et al. 2012) and strong (Nepstad et al. 2006; Adeney et al. 2009; Porter-Bolland et al. 2012) in avoiding deforestation in high-pressure areas. Similarly, strengthening government enforcement and other regulatory policies could improve the performance of strictly protected areas. However, positive effects could be offset if enforcement displaced deforestation into less accessible parks (Davalos et al. 2009) or increased subsistence deforestation in sustainable use areas and indigenous lands.

Empirical evidence also continues to be inconclusive. Recent studies find evidence that sustainable use areas and indigenous lands tend to be situated in locations with higher deforestation pressure compared to strictly protected areas (Joppa & Pfaff 2009; Nelson & Chomitz 2011; Pfaff et al. 2013; World Bank 2013), giving the former a greater potential to avoid deforestation (Fig. 4). In line with this observation, three studies have found that sustainable use areas and indigenous lands, in the aggregate, have avoided more deforestation and forest fires than strictly protected areas in the Brazilian Amazon and globally (Nelson & Chomitz 2011; Pfaff et al. 2013; World Bank 2013). Another study from Brazil suggests that strictly protected areas, in the aggregate, blocked deforestation pressures more successfully than did sustainable use areas, while indigenous lands were more effective yet (Soares-Filho et al.

2010). Taken together, these studies seem to suggest that sustainable use areas and indigenous lands are more successful by virtue of location, while strict protected areas and indigenous lands are more successful by virtue of successfully enforced regulations. However, more systematic empirical examination is necessary to understand the joint functional relationships between avoided deforestation, governance regimes, deforestation pressures, and government enforcement in tropical protected areas.

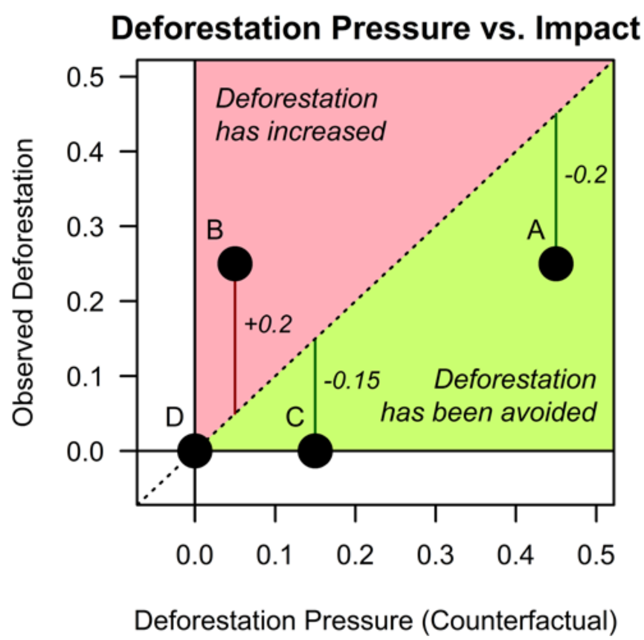

Figure 4: Relationship between deforestation pressure (deforestation rate in the absence of protection) and impact of four imaginary protected areas: A has high deforestation rates, but is estimated to have avoided deforestation compared to what would have been expected in the absence of protection. B has identical deforestation rates as A, but due to its location in a low-pressure area is estimated to have increased deforestation[8]. C, although perfectly untouched by deforestation, is estimated to have lower absolute impact than A. Located in an area of extremely low deforestation pressure, D is "passively protected" and will thus never be able to claim avoided deforestation, regardless of its observed deforestation rates.

[8] Global protected area assessments have identified countries whose protected areas exhibit higher rates of land use change than the counterfactual of no protection (Joppa & Pfaff 2011). While this phenomenon is poorly understood, and may point to methodological weaknesses, protected areas can have undesired negative effects, e.g. if resources users engage in environmentally degrading activities as a form of protest against protection (Kull 2002; Holmes 2007).

We examined whether and how the effectiveness of 292 strictly protected areas, sustainable use areas, and indigenous lands in the Brazilian Amazon co-varied with differences in deforestation pressure and federal government enforcement. Covering an area of more than 5 million km2, the Brazilian Amazon exhibits significant spatial differences in terms of agricultural potential, transport infrastructure, and market access; as a result, deforestation pressures vary widely across the region. In addition, Brazil's federal enforcement efforts underwent a major shift in recent history: Having made international headlines for a historical high in Amazon deforestation rates between 2000 and 2005, Brazil achieved radical reductions in deforestation rates in the second half of the past decade (Tollefson 2012). While part of these reductions were attributed to price declines of agricultural commodities, recent analyses also show that regulatory government policies – including a drastic increase in enforcement activities, embargoes on soy and beef markets in selected municipalities, and the expansion and strengthening of protected area networks – all contributed significantly to the observed reductions (Kis-Katos & Gonçalves da Silva 2010; Soares-Filho et al. 2010; Assunção et al. 2012). By examining the relationships between avoided deforestation, protection type, and deforestation pressure in both the first and the second half of the past decade, our analysis sheds analytical and empirical light on how governance regime, location and government enforcement jointly influence conservation outcomes in protected areas.

Results

We considered all forested protected areas in the Brazilian Amazon that had been declared in or prior to 2005 and contained at least 200 km^2 of humid tropical rainforest (Fig. B1). Strictly protected areas include state and national biological stations, biological reserves, national and state parks; sustainable use areas include state and national forests, extractive reserves and sustainable development reserves. We included indigenous lands as a third protection type of interest; although governed through different regulatory frameworks than other protected areas, indigenous lands in Brazil are subject to restrictions on development and resource use that are devised through joint planning processes involving governments and indigenous communities.

We defined deforestation pressure as the rate of deforestation that would have been expected within the boundaries of a protected area had it not been protected (counterfactual). Following earlier quasi-experimental assessments of protected area impacts (Andam et al. 2008; Gaveau et al. 2009; Joppa & Pfaff 2011; Nelson & Chomitz 2011), we non-parametrically estimated deforestation pressure as the rate of deforestation observed on artificial control groups of forest parcels. Unlike previous matching studies, we estimated deforestation pressure for each protected area individually which later allowed us to include pressure as an explanatory variable in regression-based explanations of protected area effectiveness. We identified control groups by repeatedly sampling forested parcels from within the boundaries of each individual protected area, and matching them to forested parcels that had never been protected up to 2010 but were similar in terms of key covariates associated with likelihood of protection and deforestation. We dropped forest parcels for which no sufficiently similar control parcels could be found. Estimates of deforestation rates came from two datasets: Brazil's official PRODES dataset, based on ~30m resolution LandSat imagery (Câmara et al. 2006), and the coarser Gross Forest Cover Loss (GFCL) dataset based on ~500m MODIS imagery (Hansen et al. 2010). We report deforestation rates as the total ratio of deforestation observed within a given time period on control and treatment parcels, averaged across 30 repetitions (see Methods and Materials).

To verify whether results are consistent with earlier matching studies (Nelson & Chomitz 2011; Pfaff et al. 2013; World Bank 2013), we first aggregated estimates of pressure and impact by protection type, weighting estimates for each protected area by its number of matched forest parcels (Table 4). For protected areas declared in 2000 or before, results allowed conclusions similar to earlier analyses: First, protected areas of all types exhibited less deforestation on average than similar unprotected areas. Second, sustainable use areas were, on average, situated in locations with higher deforestation pressure than strictly protected areas. Third, sustainable use areas were estimated to have avoided more aggregate deforestation than strictly protected areas in spite of higher aggregated deforestation rates in the former. Fourth, indigenous lands were consistently estimated to face the highest levels of deforestation pressures and to have achieved the greatest avoided deforestation.

Table 4: Estimates of deforestation pressure and impact, aggregated by protection type.

		Strict Protection	Sustainable Use	Indigenous Lands
Protected areas established ≤ 2000				
PRODES Deforestation 2001-05 (%)	Pressure (est.)	2.40	3.04	4.47
	Observed	0.39	0.91	0.21
	Impact (est.)	-2.00	-2.13	-4.26
Gross Forest Cover Loss 2000-05 (%)	Pressure (est.)	2.16	2.44	4.29
	Observed	0.28	0.62	0.11
	Impact (est.)	-1.88	-1.82	-4.18
[# protected areas]		[34]	[42]	[92]
[# pairs of matched forest parcels]		[5852]	[7541]	[24432]
Protected areas established ≤ 2000				
PRODES Deforestation 2006-10 (%)	Pressure (est.)	0.87	1.51	1.61
	Observed	0.16	0.64	0.10
	Impact (est.)	-0.71	-0.87	-1.51
Gross Forest Cover Loss 2005-10 (%)	Pressure (est.)	0.63	1.23	1.51
	Observed	0.08	0.50	0.13
	Impact (est.)	-0.54	-0.73	-1.38
[# protected areas]		[34]	[42]	[92]
[# pairs of matched forest parcels]		[5846]	[7538]	[23566]
Protected areas established ≤ 2005 (includes ≤ 2000)				
PRODES Deforestation 2006-10 (%)	Pressure (est.)	1.85	0.96	1.32
	Observed	0.17	0.37	0.13
	Impact (est.)	-1.68	-0.58	-1.19
Gross Forest Cover Loss 2005-10 (%)	Pressure (est.)	1.8	0.73	1.24
	Observed	0.15	0.27	0.12
	Impact (est.)	-1.65	-0.46	-1.11
[# protected areas]		[47]	[81]	[164]
[# pairs of matched forest parcels]		[9187]	[15017]	[39415]

Comparisons across time periods revealed new patterns. As expected, estimated deforestation pressure dropped considerably between 2000-05 and 2006-10 as a result of a decrease in deforestation rates on unprotected forest parcels in the Amazon. Despite this reduction, the relative ordering of protection types in terms of pressure and impact remained similar in both time periods for protected areas declared in 2000 or earlier. However, when the sample for the second time period includes protected areas established in and before 2005, the ordering of protection types changes. Strictly protected areas in the extended sample were estimated to be exposed to higher average pressure than either sustainable use areas or indigenous lands (Table 4 and Fig. B2). Closer examination revealed that these changes in average pressure estimates were driven by the creation of only a small number of large strictly protected areas in locations with high deforestation pressure (e.g. Terra do Meio, Serra do Pardo, Nascentes da Serra do Cachimbo), and the declaration of large numbers of sustainable use areas and indigenous lands in areas with very low deforestation pressure (mostly located in the state of Amazonas, see Fig. B1). These shifts in average pressure induced similar shifts in impact estimates: In spite of protection types retaining their relative ordering in terms of observed deforestation rates in the second period, strictly protected areas were estimated to have avoided *more* deforestation on average than indigenous lands and sustainable use areas.

Table 4 highlights the importance of differences in deforestation pressure as a driver of the average impact of protection types. It also demonstrates how aggregate estimates of average impact can be vulnerable to the addition of only a small number of protected areas in high-pressure locations. However, it does not provide insights into the effectiveness of protection types in inhibiting *given* levels of deforestation pressure, nor whether such effectiveness varies with high or low pressure. To illuminate these more complex relationships, we used scatterplot smoothers to non-parametrically examine observed deforestation as a function of deforestation pressure, and conducted this analysis separately for each protection type and for both time periods. We then tested the significance of the observed differences using multiple linear regressions. As most protected areas were found to be located in low-pressure locations and to exhibit low deforestation rates (Fig. B2 and Fig. B3), we transformed both variables to allow for a more detailed examination of differences in low-pressure contexts (Fig. 5).

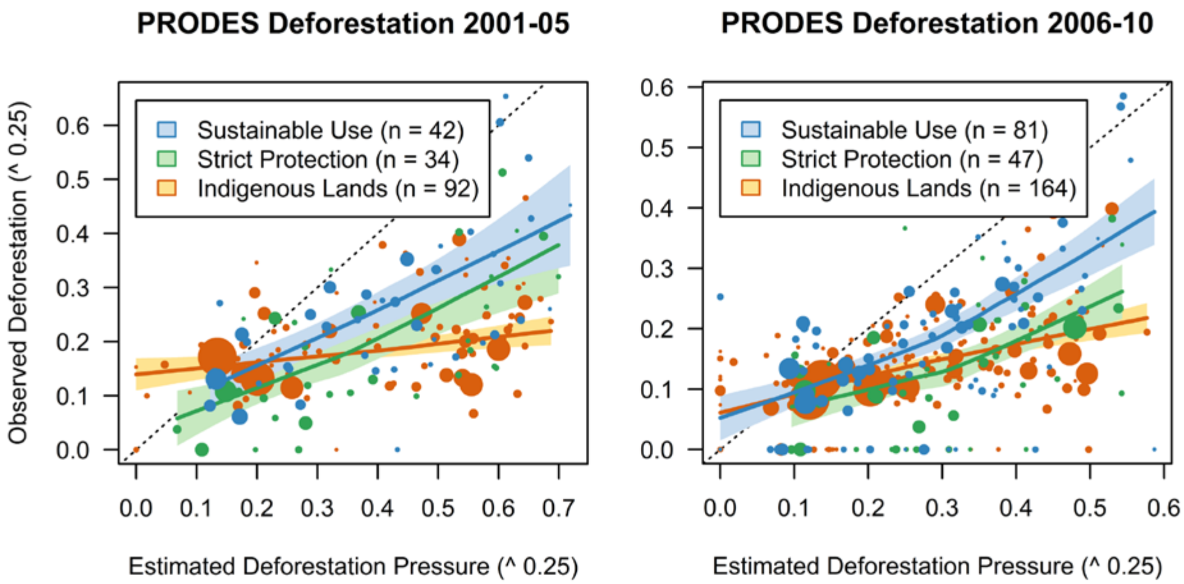

Figure 5: Observed deforestation in different types of protected areas as a function of estimated deforestation pressure (solid lines) based on protected areas established ≤ 2000 for 2001-05 impacts (left) and ≤ 2005 for 2006-10 impacts (right). Points represent protected areas, with the area of each point corresponding to the number of matched forest parcels. Shaded areas indicate 95% confidence intervals of the non-parametric estimator. All protected areas below the diagonal (black dotted line) are estimated to have avoided deforestation

Results suggest that strictly protected areas had been more effective than sustainable use areas at avoiding deforestation, regardless of the level of deforestation pressure. Across the gradient of estimated deforestation pressures, deforestation in strictly protected areas was consistently observed to be lower than in sustainable use areas, for the most part well below the 95% confidence interval around the mean (Fig. 5). We observed similar patterns in both time periods, whether we used PRODES or GFCL as measure of deforestation (Fig. B4), whether or not we applied areal weighing (Fig. B5), or excluded protected areas declared between 2000 and 2005 from the second time period (Fig. B6). Linear regressions confirmed the significance of these differences (Table B1).

Indigenous lands followed a less consistent pattern (Fig. 5). At lower levels of deforestation pressure, they exhibited deforestation rates similar to those of sustainable use areas and, between 2001 and 2005, higher than strictly protected areas. However, they appeared at least as effective as strictly protected areas at moderate levels of pressure and more effective than any other protection type at high levels of pressure. Indeed, the comparatively flat slopes of the estimated functions suggest that deforestation rates in indigenous lands seemed to be less influenced by external deforestation pressure than in other types of protected areas. Linear regressions with interactions confirmed that indigenous lands differed from strict protection and sustainable use areas in their response to deforestation pressure (Table B1). The relationship seemed less pronounced when using the coarse-resolution GFCL as the measure of deforestation (Fig. B4), providing indication that deforestation rates in low-pressure indigenous lands may largely reflect small-scale subsistence deforestation.

Discussion

Our analysis confirms that all types of protected areas have contributed to avoiding deforestation in the Brazilian Amazon regardless of their specific conservation objectives. Results also reaffirm the important role of strictly protected areas relative to sustainable use areas as a component of national strategies to mitigate climate change. First, we find that both in low and high pressure locations, strictly protected areas in the Brazilian Amazon have consistently avoided more deforestation than sustainable use areas. Second, the observed difference between strict and sustainable use areas was robust both before and after the Brazilian government stepped up efforts to curb deforestation, indicating that strict protection was not ineffective even under conditions of limited government enforcement. Third, we observe that between 2000 and 2005, a number of strictly protected areas were established in locations with high deforestation pressure, while sustainable use areas seemed more likely to be declared in low-pressure locations. Reversing earlier trends of designation patterns in Brazil, this observation suggests that both strictly protected and sustainable use areas can make substantial contributions to avoiding deforestation by virtue of their location.

Indigenous lands appeared particularly effective at curbing high deforestation pressure, relative to both strictly protected and sustainable use areas. Where we estimated deforestation pressure to be low, indigenous lands exhibited slightly more deforestation than other protection types between 2001 and 2005. This finding was not stable over time and across our robustness checks, but may suggest that deforestation in indigenous lands is less likely to be driven by the external, market-driven pressures for which our covariates controlled, and more likely a result of internal, subsistence-oriented resource use.

No governance regime guarantees protection. In spite of the consistency of average patterns, we also observed individual cases with high and low deforestation rates for all protection types, pressure levels and time periods. Assessments that seek to explain such remaining variance by looking at other policy variables – e.g. government vs. state designation (Vitel et al. 2009; World Bank 2013) or the availability of protected area management resources (Nolte & Agrawal 2013) – could benefit from applying our analytical approach to disentangle the many factors that influence success. Furthermore, our analysis does not make a distinction between illegal deforestation, which all protection types seek to reduce, and subsistence deforestation driven by the livelihood needs of indigenous and traditional people, which is legally sanctioned in sustainable use areas and indigenous lands. Incorporating protected area zonation and land rights in future parcel-based analyses has the potential to further improve our understanding of the role of enforcement and sustainable resource use in reducing deforestation in protected areas.

Although our results suggests that strictly protected areas on average are more successful at counteracting location-specific deforestation pressures than sustainable use areas, this finding cannot be read as a devaluation of the latter. Indeed, the focus of our analysis on one outcome of interest – change in forest cover – precludes statements on the relative effectiveness of protected areas in reducing other anthropogenic pressures on biodiversity and carbon, such as forest degradation, hunting, fishing, mining, and infrastructure development. Our analysis neither accounts for potential positive or negative impacts on local economies and the livelihoods of forest users nor does it consider the political and ethical dimensions of demarcating protected areas in regions with existing communities of indigenous or traditional people. Future rigorous assessments that incorporate such diverse outcomes and carefully contrast the effectiveness of different strategies in achieving the multiple objectives of protected areas will certainly be welcomed by the global conservation community as an input for effective, efficient and equitable strategies to mitigate global climate change.

Materials and Methods

Data. We obtained protected area boundaries and characteristics from the World Database of Protected Areas (UNEP-WCMC 2011) and the National Cadaster of Conservation Units (CNUC) of the Brazilian Ministry for the Environment (www.mma.gov.br). Deforestation estimates were based on 1) a fine-scale dataset (PRODES) based on LandSat imagery and published by the Brazilian Institute for Space Research (Câmara et al. 2006) and 2) the coarse-resolution Gross Forest Cover Loss (GFCL) dataset based on Moderate Resolution Imaging Spectroradiometer (MODIS) imagery and published by South Dakota State University (Hansen et al. 2010). Baseline forest cover in 2000 and 2005 came from the Vegetation Continuous Fields (VCF) of the Global Land Cover Facility (Hansen et al. 2003). We computed travel time estimates to major cities based on the algorithm and datasets of (Nelson 2008), supplemented by improved road datasets generated by SimAmazonia (Soares-Filho et al. 2006) and land cover estimates for 2000 obtained from MODIS Land subsets (Oak Ridge National Laboratory Distributed Active Archive Center 2011). Other datasets include slope and terrain from the International Institute for Applied Systems Analysis (Fischer et al. 2007), floodable areas as identified by GlobCover 2005 (Arino et al. 2009), and state boundaries from the Global Administrative Areas database (www.gadm.org). We projected all datasets into MODIS' own sinusoidal projection, resampled them to ~1km resolution, and extracted all humid tropical forest parcels with more than 25% average forest cover (VCF) into one table (see Appendix B).

Estimating Deforestation Pressure. We used matching to create artificial control groups of forest parcels for each protected area. We considered all protected areas established ≤ 2005 that had at least 50% average tree cover in 2000 (Hansen et al. 2003), were located to at least 60% in the Humid Tropical Forest Biome (Olson et al. 2001), and contained at least 200 forest parcels (at ~1km resolution). We excluded Brazil's Environmental Protection Areas from the group of sustainable use areas, as they primarily consist of private lands on which the protected area does not impose significant additional restrictions (Verissimo et al. 2011). We did not consider military areas. We randomly sampled 5% of forested parcels from each of the remaining 292 protected areas, and matched them to 5% samples of forested parcels that 1) had never been protected up to 2010 and 2) were situated further than 10km away from any protected area boundary. Following related studies (Andam et al. 2008; Gaveau et al. 2009; Joppa & Pfaff

2011; Nelson & Chomitz 2011), we controlled for elevation, slope, probability of flooding, baseline forest cover, distance to forest edge, travel time to major cities, and state. Control groups for 2000-05 and 2006-10 were estimated separately, the latter accounting for changes in covariates (baseline forest cover and distance to forest edge) that had occurred within the first time period. Matching was with replacement. We dropped forest parcels for which no nearest neighbor could be found within one standard deviation of each covariate (caliper). We repeated the process of random sampling and matching 30 times for each protected area and averaged the resulting estimates of observed deforestation and deforestation pressure. See Appendix B for supporting information on covariate choice, covariate balance, and leakage.

Comparing Effectiveness. We estimated and contrasted pressure-specific effectiveness of different protection types using both non-parametric and parametric regressions. Locally weighted scatterplot smoothers (LOESS) allowed us to flexibly examine differences in the response of observed deforestation in different protection types as a function of deforestation pressure (Fig. 5). Results from these non-parametric regressions informed the specifications of the linear regressions we used to formally test for the strength of the observed differences (Table B1 and Appendix B). In order to reduce skewness of distributions, reduce issues of heteroskedasticity, and to allow for a more detailed examination of differences in low-pressure locations, we transformed estimates of observed deforestation and deforestation pressure prior to applying regressions (see Appendix B).

CHAPTER 4: SETTING PRIORITIES TO AVOID DEFORESTATION IN AMAZON PROTECTED AREAS. ARE WE CHOOSING THE RIGHT INDICATORS?[9]

Abstract

Cost-effective protected area networks require that decision makers have sufficient information to allocate investments in ways that generate the greatest positive impacts. With applications in more than 50 countries, the Rapid Assessment and Prioritization of Protected Area Management (RAPPAM) method is arguably the tool used most widely to assist such prioritization. The extent to which its indicators provide useful measures of a protected area's capacity to achieve its conservation objectives, however, has seldom been subject to empirical scrutiny. We use a rich spatial dataset and time series data from 66 forest protected areas in the Brazilian Amazon to examine whether RAPPAM scores are associated with success in avoiding deforestation. We find no statistically significant association between avoided deforestation and indicators that reflect preferential targets of conservation investments, including budget, staff, equipment, management plans and stakeholder collaboration. Instead, we find that the absence of unsettled land tenure conflicts is consistently associated strongly with success in reducing deforestation pressures. Our results underscore the importance of tracking and resolving land tenure in protected area management, and lead us to call for more rigorous assessments of existing strategies to assess and prioritize management interventions in protected areas.

[9] This chapter has been published as: Nolte C, Agrawal A, Barreto P (2013) Setting priorities to avoid deforestation in Amazon protected areas: Are we choosing the right indicators? **Environmental Research Letters** 8 (2013):015039

Introduction

Protected areas are the pride of the global conservation movement, but also remain one of its primary concerns. As diverse as the 177,000 units of the global protected area network are in terms of species, ecosystems, threats and management responses (Bertzky et al. 2012), as rich are they in stories of both successes and failures (Bruner et al. 2001; Joppa & Pfaff 2011). Especially in the tropics, home to some of the world's greatest ecological diversity, many sites fail to achieve stated conservation objectives fully (Oates 1999; Verissimo et al. 2011). Given the pervasive and chronic budget constraints under which most protected areas operate (Bruner et al. 2004), supporters and managers of protected area networks are increasingly expected to allocate resources and efforts in ways that yield the most cost-effective outcomes (Ferraro & Pattanayak 2006). With protected areas being put forward as a potentially cost-effective strategy to reduce deforestation and forest degradation (Nauclér & Enkvist 2009; Venter et al. 2009; Soares-Filho et al. 2010), the effective prioritization of conservation funds has become a key task for donors, park agencies and project managers alike.

Decision makers can use an abundance of methods to assess protected area management and prioritize interventions. In the past decade, more than 70 methods have been developed to provide standards for indicator collection, analysis and interpretation (Leverington et al. 2010; Nolte et al. 2010). Their value for prioritization hinges on their ability to help decision makers predict and compare the potential outcomes of alternative interventions. This prerequisite translates into two challenges: First, methods need to be able to provide an accurate assessment of the management *status quo* of a given protected area network (baseline). Second, they need to provide insights into how alternative interventions in different sites will affect management and, ultimately, the likelihood of achieving desired future outcomes (prediction). Selection and interpretation of method indicators thus presuppose a thorough understanding of the causal pathways through which interventions affect management and outcomes. Defining metrics and data collection procedures also involves a trade-off between cost and precision. While ecological monitoring systems and independent experts can provide more reliable metrics than subjective self-assessments of staff, their additional effort is not necessarily commensurate with the potential value of improved accuracy (Hockings et al. 2009).

Recent reviews suggest that popular methods to assess protected area management have important similarities (Cook & Hockings 2011). Most methods collect data on a wide range of management aspects, often selected with reference to an assessment framework developed by the World Commission of Protected Areas (WCPA) (Hockings et al. 2006). Covering a large diversity of management indicators reduces the risk of omission errors and can enhance flexibility in the application of methods in different management contexts. When it comes to metrics and data collection, many methods rely on qualitative indicators elicited directly from protected area managers, suggesting that this approach is generally perceived as striking a satisfactory balance between cost and precision.

One of the most widespread prioritization methods for protected area management is the Rapid Assessment and Prioritization of Protected Area Management (RAPPAM) method (Ervin 2003a, 2003b). Developed by the World Wide Fund for Nature (WWF) for assessments of protected area networks, RAPPAM aims, among other things, to “help develop and prioritize appropriate policy interventions and follow-up steps to improve protected area management effectiveness” (Ervin 2003a:3). RAPPAM questionnaires elicit responses from protected area managers who are asked to rank 90 qualitative statements on a four-point scale based on how well the statement applies to their protected area site (“yes”, “rather yes”, “rather no”, “no”). To ensure “consistent scoring across different parks” (Ervin 2003b:834), the RAPPAM manual encourages the questionnaire to be filled out in national-level workshops, with participants clarifying the meaning of terms such as “adequate”, “appropriate”, and “sufficient” in the national context (Ervin 2003b). By 2010, RAPPAM had been applied in more than 2000 protected areas in more than 50 countries on five continents (Leverington et al. 2010; Kinouchi 2012), making it a *de facto* standard in present-day assessments of protected area management. However, whether or not RAPPAM provides useful measures of a protected area’s capacity to achieve its conservation objectives has seldom been subject to empirical scrutiny.

In this paper, we examine the relationship between RAPPAM scores and the success of protected areas in avoiding deforestation in the Brazilian Amazon. Given that RAPPAM was “developed specifically for forest protected areas” (Ervin 2003a:6), we expected the method to perform particularly well at characterizing success in reducing what constitutes a major threat to forest biodiversity. We chose Brazil because its position as the world’s largest deforester has prompted considerable investments into enlarging and consolidating the country's protected area network. Protected areas and indigenous lands now cover 43.9% of the Brazil's Amazon region (Verissimo et al. 2011). Brazil is also home to one of the world's largest protected area support programs – the Amazon Region Protected Area (ARPA) program – whose investment strategy provides insights into what indicators reflect preferential targets of conservation interventions (ARPA 2010). Finally, Brazil's government has taken RAPPAM seriously, collaborating with WWF to apply the method to more than 250 federal protected areas in 2005 and 2010 (Kinouchi 2012). We use a rich spatial dataset and statistical matching to discriminate between protected areas that have been more and less successful at countering deforestation pressures between 2006 and 2010. We then examine the extent to which both groups differ in terms of RAPPAM scores, paying particular attention to indicators that reflect preferential targets of interventions, and to those showing the strongest associations with success.

Methodology

Estimating Deforestation Pressure

We considered all 152 protected areas in the Brazilian Amazon that had been reported to the World Database of Protected Areas (UNEP-WCMC 2011), had been declared in 2006 or earlier, were located in tropical and subtropical moist broadleaved forests (Olson et al. 2001) and contained at least $200km^2$ of forest cover in 2000 (Hansen et al. 2003). In accordance with Brazilian nomenclature, we considered biological stations, biological reserves, and national and state parks to be "strictly protected areas", and classified national forests, extractive reserves and sustainable development reserves as "sustainable use areas". We excluded Environmental Protection Areas as they consist primarily of private lands without significant additional restrictions (Verissimo et al. 2011). We did not include indigenous lands as they had not been included in RAPPAM analyses in Brazil.

We defined deforestation pressure as the rate of deforestation to which each protected area would have been exposed had it not been declared as protected (counterfactual). We estimated deforestation pressure non-parametrically by repeatedly sampling $1km^2$ forest parcels from each protected area, matching sampled forest parcels to similar parcels that had not been protected until 2010, and measuring deforestation rates on these artificially generated control groups of forest parcels. Control parcels were located in the same state and outside a 10km buffer around protected areas to reduce possible effects of local leakage on our estimates (Andam et al. 2008). In line with related matching studies (Andam et al. 2008; Joppa & Pfaff 2011; Nelson & Chomitz 2011), we used nearest-neighbor matching with replacement, controlling for six important covariates (all resampled to ~1km2 resolution using MODIS' sinusoidal projection):

- Average slope, from Fischer et al. (2007)
- Average elevation, from Fischer et al. (2007)
- % Floodable area, as identified by Arino et al. (2009)
- Travel time to major cities, using own computations based on Nelson (2008) and improved road datasets from Soares-Filho et al. (2006)
- Distance to forest edge, based on own computations that used percent tree cover estimates (Hansen et al. 2003), ESRI hydropolygons, as well as road datasets from Soares-Filho et al. (2006).
- Baseline % forest cover (Hansen et al. 2003).

We measured deforestation rates on protected and matched unprotected parcels as the total ratio of deforestation observed by Brazil's official deforestation monitoring system PRODES (Câmara et al. 2006) between 2006 and 2010. We repeated sampling and matching 30 times for each protected area and averaged resulting estimates of deforestation pressure. We discarded forest parcels for which no suitable control parcel could be found within 1 SD of each covariate (calipers). To assure that matching was sufficiently representative for a given protected area, we discarded sites for which less than 50% of parcels could be matched. A total of 142 protected areas met all quality criteria, 66 of which had been subject to RAPPAM analyses. Average deforestation rates on protected and matched unprotected parcels were 0.51% and 1.89%, respectively, as compared to 1.12% for the entire Amazon.

Defining Success Groups

We used quantile regression to identify groups of protected areas whose deforestation rates in 2005-2010 had been high ("low success") vs. low ("high success") as compared to protected areas of similar category and pressure. Deforestation in protected areas in the Brazilian Amazon has been shown to increase with deforestation pressure, with sustainable use areas exhibiting significantly higher deforestation rates than strict protected areas exposed to similar pressure (Nolte et al. 2013b). We therefore used deforestation pressure estimates and a dummy variable for sustainable use areas as the two independent variables in the quantile regression. Conducting the analysis without the sustainable use dummy variable yielded similar results. As distributions of both observed deforestation and deforestation pressure were highly skewed toward low values, we transformed these two variables to better satisfy regression assumptions. We chose τ (tau) values of 0.25, 0.5, and 0.75 to split our sample into four quartiles ranging from most to least successful in avoiding deforestation.

Following quantile regressions, we employed two distinct strategies to split our sample into more vs. less successful protected areas. For our "full sample", we merged the two upper and two lower quartiles into a "low" (n=38) and a "high success" (n=28) group, respectively. While taking advantage of the full sample size, this strategy did not yield a neat separation of groups, as many moderately successful areas exhibited deforestation rates very similar to their less successful counterparts (Figure 6). Furthermore, many protected areas in the full sample had very low deforestation pressure estimates. We expected such remote areas to be more likely to allocate their management capacity toward other threats (e.g. logging, hunting, fishing, tourism impacts) and to be less relevant as barriers to deforestation; moreover, their categorization into success groups seemed more vulnerable to small variations in pressure estimates. We therefore developed an additional "high confidence" sample, for which we juxtaposed only the least (n=12) and most (n=11) successful quartiles, and ignored observations whose deforestation pressure estimates were below a minimum threshold (<0.1%). Our assumption was that protected areas in the latter sample would differ more strongly in terms of RAPPAM scores that measure management aspects relevant to avoided deforestation success.

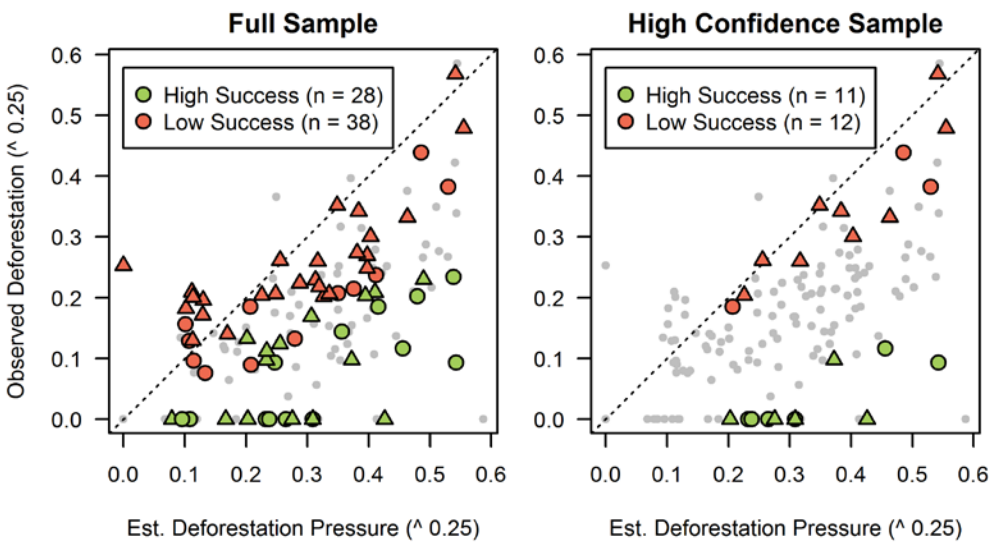

Figure 6: Observed deforestation rates and estimated deforestation pressure (2006-2010) of 142 forest protected areas in the Brazilian Amazon. Circles indicate strict protected areas, triangles sustainable use areas. Grey dots indicate protected areas that were included in regressions, but did not have RAPPAM scores (left) or were excluded from the high-confidence sample (right). All protected areas below the diagonal (y=x, dashed line) are estimated to have avoided deforestation.

Comparing RAPPAM Indicators Between Groups

Once we had defined success groups, we examined how more and less successful areas differed in terms of RAPPAM scores. Given our interest in the predictive potential of RAPPAM, our analytical focus was on associations of 2005 scores with subsequent impacts on deforestation (2006-2010). However, many protected areas in the Brazilian Amazon received investments between 2006 and 2010 through the Amazon Region Protected Area (ARPA) project (ARPA 2010), which supported enforcement missions, acquisition of field equipment, and elaboration of management plans, among other activities. As ARPA investments could have been preferentially allocated toward areas with low or high RAPPAM scores, or low or high expected deforestation rates, we used 2010 RAPPAM data to examine whether observed differences between scores had changed between 2005 and 2010.

We used two-tailed t-tests to test for differences in indicators for each time period and each sample. Error probabilities for the full and high confidence sample are reported as p_f and p_h, respectively. We first examined whether success groups differed with respect to their composite RAPPAM management effectiveness scores, calculated as the average of all management-related RAPPAM scores (sections 6-16) for a given protected area. We then looked at group differences in terms of selected “priority indicators” that reflect typical targets of conservation investments and were thus expected to be positively associated with success in avoiding deforestation. These indicators included adequacy of past and future budget, staff numbers, equipment, management plans, and stakeholder collaboration (see Figure 7 for wording). As a third step, we mined the full set of 90 RAPPAM scores for significant differences between success groups, applying the Benjamini-Hochberg correction (α=0.1) to reduce the risk of false discoveries. We used a similar approach to test whether success groups differed in the extent to which their RAPPAM scores had *changed* from 2005 to 2010 in order to detect possible biases in investments.

Results

RAPPAM responses exhibited several noticeable patterns (Figure 7). For many indicators, respondents used the full range of possible answers, creating sufficient variance to allow for meaningful comparisons. However, distributions of indicator scores pertaining to funding, staff numbers, equipment, and management plans were skewed toward negative responses in 2005, resulting in very low or even zero variance in some cases. The observation of low responses to funding and staff questions is consistent with global patterns (Leverington et al. 2010). In our sample, those indicators improved considerably between 2005 and 2010, possibly reflecting the impact of ARPA investments. Although success groups differed in the extent to which some scores had changed over time, none of these differences was found be significant at the threshold defined by the Benjamini-Hochberg correction.

Composite RAPPAM management effectiveness scores of more successful protected areas were consistently higher than those of less successful protected areas. However, these differences tended to be very small and only significant for 2010 RAPPAM scores of the high confidence sample. Priority indicators seemed to perform even worse as predictors of avoided deforestation success. Across samples and time periods, none produced responses that differed sufficiently between high and low success groups to be significant at the 10% level. For some indicators and time periods, the absence of variance in scores precluded the existence of significant group differences.

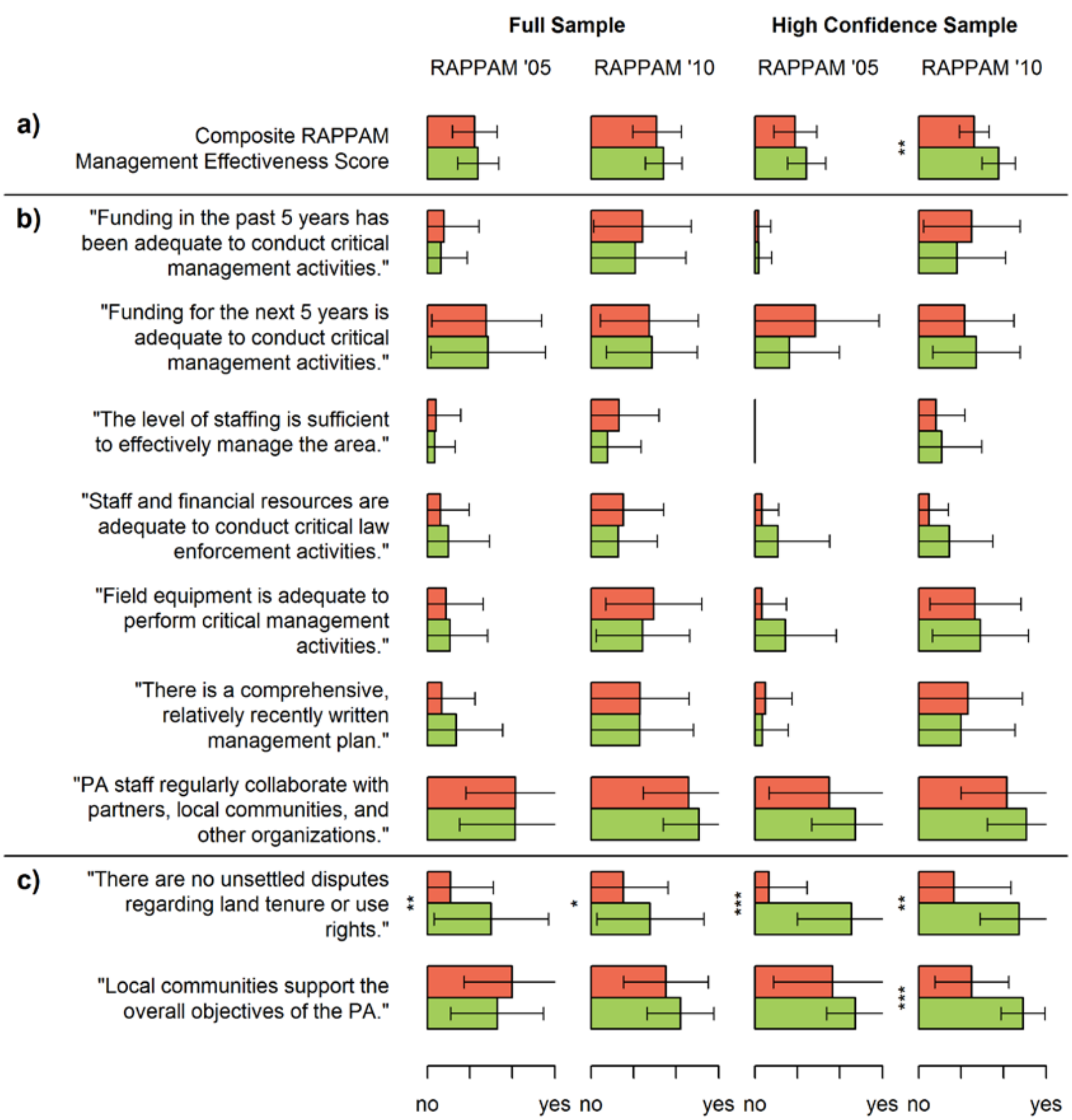

Figure 7: RAPPAM scores of protected areas estimated to have been more (green) vs. less (red) successful at avoiding deforestation between 2006 and 2010. Possible responses include no (0), rather no (1), rather yes (2) and yes (3). We report a) composite RAPPAM scores b) scores of priority indicators and c) scores of indicators for which false discovery rate analysis identified significant group differences in at least one time period. Error bars indicate SD of each success group. Significance levels of individual t-tests: *** p < 0.001, ** p < 0.01, * p < 0.05

Our data mining process identified only two indicators that passed the Benjamini-Hochberg test for at least one time period. Across samples and time periods, success groups differed significantly in whether protected area managers reported “unsettled disputes regarding land tenure or use rights” ($p_f = 0.003/p_h = 0.0005$ in 2005 and $p_f = 0.036/p_h = 0.004$ in 2010). In 2010, success groups in the high confidence sample also differed significantly in the extent to which the protected area manager perceived local communities to support the overall objectives of the protected area (p_h=0.0007). However, this difference was not found to be significant for other combinations of time period and sample ($p_f = 0.203/p_h = 0.259$ in 2005 and $p_f = 0.124$ in 2010). We added both indicators to Figure 2 for reference.

Discussion

Given the widespread use of RAPPAM for prioritizing protected area interventions worldwide, the associations between its scores and avoided deforestation seem surprisingly weak. As RAPPAM is only one of many methods using subjective self-assessments to evaluate management, this observation constitutes both a puzzle and a potential reason for concern. If such a method failed to help discriminate between protected areas that are more or less successful in achieving a key conservation goal – especially in terms of typical targets of conservation investments – how are decision makers to use these scores to prioritize among policy, management and resource allocation responses? And if its scores seem to lack predictive power, does this finding point towards the method not measuring the right aspects – or not measuring them right?

Although the absence of verifiable alternative measures prevents us from providing a definitive answer to these questions, our results indicate that both “what” and “how” to measure have an influence on the observed patterns. Indeed, we observe the strongest associations with success for an indicator whose wording is relatively unambiguous (“no unsettled disputes”). Most priority indicators have qualifiers such as “adequate” and “sufficient” and exhibit far weaker associations. Furthermore, we find overall RAPPAM scores and local community attitudes in 2010 to have stronger associations with success than their respective 2005 scores. This finding seems to indicate that protected area managers are inclined to judge current management more positively if the protected area has successfully curbed deforestation in the past. Indeed, we would expect RAPPAM-measured management improvements to *precede* conservation impact and not vice versa. If subjective evaluations by protected area managers indeed exhibited such systematic biases, this would not only call into question the utility of many existing assessment methods (Cook & Hockings 2011), but might also undermine the findings of earlier studies using manager responses to estimate park success (Bruner et al. 2001; see also Hayes 2006).

If, instead, we allow for the assumption that RAPPAM's workshop format eliminates potential biases in scores between protected areas, the missing links between budget, staff, equipment, management plans, collaboration and avoided deforestation seem all the more striking. Is it possible that the local management capacity of protected areas in the Brazilian Amazon bears only a minor influence on deforestation patterns? With a large number of protected areas located in remote locations, Brazil's federal and state governments have traditionally allocated only small budgets and staff numbers to individual conservation units. Recent attempts to curb Amazon deforestation relied heavily on concerted enforcement with federal and state-level police units, municipality-wide embargoes of agricultural commodities, and other policy instruments whose scope extended beyond the boundaries of individual protected areas. Spatial heterogeneity in the effects of these policies could potentially explain some of the observed differences in avoided deforestation success. However, such heterogeneity is not reflected in current RAPPAM scores.

Disputes regarding land tenure emerged as the one factor to be most consistently associated with the extent to which protected areas succeeded at avoiding deforestation. It was the only score that differed between success groups across time, and overshadowed all other factors in terms of significance. However, legal and financial responsibilities for resolving land tenure issues in Brazil are predominantly vested with central conservation authorities and thus beyond the obligations and budget lines of individual protected area managers. Indeed, of 24 protected area managers reporting land tenure issues in 2010 ("rather yes"/"yes"), 46% considered their budget in the previous five years to have been adequate ("rather yes"/"yes") for critical management activities. At the same time, RAPPAM provides little information about the adequacy of human and financial resources of authorities responsible for the resolution of land tenure issues, which may or may not vary across regions and protected areas and explain part of the differences in avoided deforestation success.

Outlook

If RAPPAM in its current version is only of limited utility for discriminating between protected areas that are more vs. less successful at avoiding deforestation, our analysis provides a starting point for potential improvements. The possibility of a weak link between local management capacity and deforestation suggests that judgments on the adequacy of key management resources and processes may be inadequate if they do not account for the fact that important responsibilities for reducing threats to biodiversity may be distributed among actors at different administrative levels. Discriminating between key threats and corresponding key authorities in the judgments of management capacities may be a potential way forward to improve the extent to which RAPPAM helps identifying key bottlenecks in multi-level protected area management, and thus provide a better basis for the prioritization of conservation resources. Although our analysis did not explore the causal direction of links between scores and deforestation, its results suggest that the existence and resolution of land tenure conflicts warrants particular attention, and that RAPPAM's value as a prioritization method could improve if it helped to identify effective strategies in resolving such conflicts.

Although the focus of our analysis – protected areas and deforestation in the Amazon – is instructive, our study constitutes only a first look at the relationship between management effectiveness scores and conservation success. To improve the value of RAPPAM for conservation decisions, several other questions merit additional analysis: What causal relations can be identified between individual management aspects, especially land use disputes, and their relationship to conservation success? If local management capacity shows only weak associations with avoided deforestation, may it be more closely related to success in preventing more furtive types of threats, such as logging, hunting and fishing? Which management aspects seem more responsive to external conservation investments, and which don't? Can enhancements to the measurement of indicators (e.g. rephrasing, quantification, triangulation) enhance their predictive power? What influence do differences in institutional arrangements and political contexts bear on the usefulness of RAPPAM to characterize successful protected areas? Given the magnitude of international and national funding for protected areas, decision makers will continue to ask for information that helps prioritize investments. Only when available decision tools begin to provide insights into these questions will they begin to pave the way toward cost-effective conservation.

CHAPTER 5: IDENTIFYING CHALLENGES TO ENFORCEMENT IN PROTECTED AREAS: EMPIRICAL INSIGHTS FROM 15 COLOMBIAN PARKS[10]

Abstract

Protected areas aim to conserve biodiversity by restricting human activities within their boundaries. However, many tropical parks struggle to fully enforce such restrictions. Improving regulatory enforcement requires an understanding of prevailing challenges in current detection and sanctioning activities. Drawing from an empirical field research of 15 Colombian parks, I show that current enforcement efforts may create only minimal deterrents to most prevalent priority threats. Long-term infractions, such as agriculture, livestock grazing, and constructions, mostly posed challenges in sanctioning violators, while furtive infractions, such as logging and hunting, mostly posed challenges in detecting violators. Investments into staff, equipment and infrastructure may fail to increase enforcement capacity and conserve biodiversity if they are not accompanied by a resolution of land tenure, clarifications of use rights, improved patrolling strategies, and protection of park guards from conflict.

Introduction

Protected areas are the most widely used spatial policy instrument in efforts to conserve the variety of biological life on Earth. More than 175,000 sites, covering an area twice the size of Brazil, establish spatially explicit regulations to protect species and ecosystems against anthropogenic pressures (Chape et al. 2005; Bertzky et al. 2012). Widespread evidence indicates that those living near protected areas often do not observe these regulations (Oates 1999; Dudley et al. 2004; Terborgh 2004), and that severe and chronic budget constraints limit the ability of

[10] This chapter has been resubmitted to Oryx after a first set of favorable reviews

many tropical protected areas to respond (Bruner et al. 2004; McCarthy et al. 2012). Improving the effectiveness of protected areas has thus become a key objective for the international nature conservation community (CBD & UNEP 2010; Coad et al. 2013), and international donors continue to dedicate billions of dollars to support tropical parks and reserves (GEF 2009; Kasparek et al. 2010; Miller et al. 2013).

Regulatory enforcement plays an important role in the quest for more effectively protected areas. Empirical analyses identify enforcement as a major determinant for the conservation success of parks (Bruner et al. 2001; Hilborn et al. 2006) and community-managed forests (Gibson et al. 2005; Chhatre & Agrawal 2008). Guards and other enforcement costs tend to make up the bulk of annual budgets in many parks (Robinson et al. 2010). However, the vast expanses of remote or inaccessible land that many tropical protected areas cover can turn enforcement into a very costly activity. Analysts have thus begun to acknowledge that it is rarely optimal to prevent all illegal activity, redirecting analytical efforts instead to optimizing enforcement under existing budgetary constraints (Robinson et al. 2010; Albers & Robinson 2013).

Effectively improving a park's enforcement capacity requires a comprehensive understanding of weaknesses in the prevailing enforcement regime. The presence of illegal activities suggests that violators perceive enforcement deterrents to be smaller than their expected benefits from engaging in the illegal activity. To improve enforcement, decisions makers need to estimate how small these deterrents are, and where the weaknesses lie. Enforcement deterrents can be understood as the product of an "enforcement chain" of consecutive steps, such as detection, arrest, prosecution, and conviction (Sutinen 1987; Bruner et al. 2001; Akella & Cannon 2004). Weaknesses in any step can undermine the effectiveness of the entire enforcement regime. For instance, if sanctioning processes are entirely ineffective, a mere intensification of patrolling may not be sufficient to increase enforcement deterrents. To complicate things, most protected areas are exposed to various types of illegal activities (Leverington et al. 2010), each of which can pose distinct challenges for each enforcement step, but requires resources from the same limited budget. The identification of enforcement weaknesses thus requires a comprehensive assessment of all enforcement steps and priority threats within the protected areas of interest.

Figure 8: Spatial boundaries and names of parks included in this study.

Study Area

I demonstrate the value of such a comprehensive approach through an empirical assessment of enforcement regimes in 15 Colombian parks. One of the world's most biologically diverse nations, second only to Brazil in terms of species richness (Groombridge & Jenkins 2002), Colombia has been ranked among the five countries with the highest potential to generate global biodiversity benefits (GEF 2008). Colombia's system of national nature parks is representative of that of many developing countries rich in tropical biodiversity: parks cover vast expanses of the national territory and their managers struggle to enforce regulations against diverse anthropogenic threats. In 2013, the German government made a multi-million dollar commitment to improve the management and enforcement of parks situated in Colombia's

Northeast. Assessing enforcement patterns prior to implementation was considered valuable for project implementation and future impact assessments. The main criterion for park selection was therefore their inclusion among project recipients: all but one park (Serranía De Los Yariguíes) are scheduled to receive new enforcement resources from 2014 onwards (Fig. 8). Results illustrate how the incorporation of all enforcement steps and multiple threats in a comprehensive assessment of enforcement enables policy insights that would not have emerged by looking at individual threats, steps or parks only.

Methods

Several application-oriented criteria informed the research design: First, to be generalizable beyond the context of this study, the method needed to demonstrate its value across a wide range of different enforcement settings. The fifteen parks selected for this study (Fig. 8) vary considerably in size (6.4km^2 to 3830km^2), altitude (0 to approx. 5700m), accessibility, as well as the range and intensity of threats. They span ecosystems as diverse as coral reefs, tropical beaches, saltwater and freshwater lagoons, mangrove forests, tropical rainforests, tropical dry forests, cloud forests, páramo, rock, and ice. Parks also differed in the extent to which they overlapped with historical land use, indigenous lands and private property claims.

Second, to be applicable under conventional project conditions, the method needed to be cost-effective, i.e. produce reliable insights with reasonable time investments of analyst and park staff. All data was collected through on-site workshops with key informants that took place between September and November 2013. At minimum, informants included 1) the park manager, 2) the staff member responsible for overseeing enforcement and 2) a senior park ranger with extensive experience in the area. In some cases, the park manager invited all staff members to attend the workshop. Depending on the complexity of the parks' enforcement context, workshops would take between 2 and 6 hours (average: 3 hours). Where possible, I accompanied park guards during routine patrols on the day prior or following the workshop (10 out of 15 parks). I did not interview other actors, such as rule violators.

Third, to be suitable for project evaluation, the method emphasized indicators that could be verified and tracked across time, wherever possible. Rather than using ratings or rankings, data collection was predominantly based on observable incidences of detection and sanctioning. Whenever indicator values were unobservable and had to be estimated by participants, observable supporting evidence was collected prior to the moment of estimation. For instance, before asking respondents to estimate the total extent of a given threat (both detected and undetected), I elicited supporting evidence on the spatiotemporal patterns of patrols, the estimated duration of an infraction, and the number of detected incidences. This approach of collecting more uncontroversial and verifiable indicators early in the workshop, and using them to cross-check the plausibility of later estimates also served as a strategy to control potential strategic bias. For instance, staff could have under-reported enforcement success to signal greater need for resources, or over-reported success in order to suggest higher management capabilities. To minimize recall bias, the method was applied to the three-year period prior to the workshop. In most parks, the number of detected incidences was small, and informants had little difficulty recalling specific enforcement events and their proceedings. If the number of detected incidences was high (Park Tayrona), supplementary information was obtained from lawyers in the regional park office.

At the beginning of each workshop, I collected data on the park's enforcement resources (staff, equipment, and infrastructure), patrolling intensity, patrolling patterns (including spatial and temporal predictability), and land tenure situation (private tenure claims and overlap with indigenous lands). I then asked participants to list all priority threats for enforcement, i.e. illegal activities that occurred within the area, were perceived as an important threat to conservation values, and fell within the parks' enforcement mandate. Threats caused by human activities outside the park boundaries (e.g. climate change, or siltation) or subject to high-level political decisions (e.g. major mining or infrastructure projects) did not fall under this definition.

To estimate the magnitude of the enforcement deterrent for a given priority threat, participants were first asked to define a unit of infraction (e.g. logged trees, fishing trips). We then used this unit of infraction to quantify the amount of detected incidences, estimated total incidences, arrests, convictions, and average sanctions (fines, value of confiscated goods, etc.). Following data collection, we computed the enforcement deterrent as the total cost of observed sanctions divided by the estimated number of total incidences (Sutinen 1987; Akella & Cannon 2004). I then asked participants to estimate how large this enforcement deterrent was in comparison to the violator’s expected benefits from a single infraction, eliciting this estimate as an ordinal variable with five categories (see columns of Table 5). If estimated deterrents were zero (as in the majority of cases), this step did not require much thought. However, whenever estimated deterrents were larger than zero, we relied on the knowledge and cognitive models of respondents about the economics of the illegal activity and relevant benefits for violators (material or non-material). In spite of the potential cognitive burden, most groups did not have major disagreements in picking indicator values. Subsequent discussions gave participants the opportunity to elaborate on reasons for the effectiveness or ineffectiveness of each step of the enforcement chain, and to suggest perceived solutions.

Following field research, I used a simple hierarchical categorization scheme to identify key challenges to enforcement: A priority threat was defined to face a “detection challenge” in the park in question, if park staff estimated the empirical probability of detecting infractions and identifying the violator to be below 1%. Priority threats without a detection challenge were categorized as facing a “sanctioning challenge”, if less than 1% of all identified violators had been sanctioned through fines, arrest, confiscation, demolition, or similar. Although this simple categorization could mask more complex enforcement processes (such as confiscating valuable equipment to incite self-denunciations), it proved useful for summarizing key enforcement challenges across diverse priority threats and parks.

Results

Interviewed park staff reported between one and six categories of priority threats for enforcement within their protected area, adding up to a total of 54 individual priority threats (mean: 3.6 threats / park). Priority threats linked to agricultural, pastoral and extractive resource uses were mentioned most frequently (80%, see Table 5). Other priority threats included constructions, tourism activity, and fires.

Table 5: Identified priority threats and correspondent enforcement deterrents in 15 Colombian national parks

			Distribution of Estimated Enforcement Deterrents				
Priority Threat	**# of Parks**	**Characteristics**	**Very low (<1%)**	**Low (1-10%)**	**Medium (10-50%)**	**High (50-100%)**	**Very high (<100%)**
Livestock grazing	12	Mostly cattle, also sheep, goats, donkeys, horses; within and outside of fenced properties	11	-	1	-	-
Logging and wood extraction	9	Selective timber extraction, pole cutting for fences, firewood collection, charcoal production	6	2	-	1	-
Agriculture (established)	6	Agricultural production on permanent plots, mostly small-scale farms, rarely larger plantations	6	-	-	-	-
Agricultural frontier expansion	6	Slash-and-burn cultivation of new or overgrown plots, mostly small-scale	6	-	-	-	-
Hunting, extraction of fauna/flora	6	Hunting and trapping for recreational, commercial and subsistence purposes, collection of non-timber forest products	6	-	-	-	-
Constructions	5	Construction or improvements of family dwellings and tourism infrastructure, also second homes	2	1	1	1	-
Tourism	4	Unauthorized access and camping, campfires, motorized access, entry of horses and pets, trash disposal, etc.	4	-	-	-	-
Fishing	4	Netting, angling, harpooning, hand fishing (shellfish), rarely dynamite; in bogs, lagoons and proximity to shores.	4	-	-	-	-
Fires	2	Mostly accidentally escaped fires of hunters, tourists, or farmers (campfires, smoking, slash-and-burn)	2	-	-	-	-

Table 6: Identified priority threats and reported key reasons for low enforcement deterrents

	Detection Challenge		Sanctioning Challenge	
Priority Threat	**# Parks**	**Frequently mentioned reasons**	**# Parks**	**Frequently mentioned reasons**
Livestock grazing (n=12)	4	No municipal retention facility for livestock, livestock not branded, absence of patrols in affected area	7	Unresolved land tenure (private and collective), overlap with indigenous lands, conflict potential, slow legal processes
Logging and wood extraction (n=9)	6	Furtive activity, avoidance behaviour	2	Overlap with indigenous lands, high conflict potential (incl. armed groups)
Agriculture (established, n=6)	0	-	6	Unresolved land tenure (private), overlap with indigenous lands
Agricultural frontier expansion (n=6)	2	Furtive activity	4	Unresolved land tenure (private), overlap with indigenous lands, conflict potential (incl. armed groups)
Hunting, extraction of fauna/flora (n=6)	6	Furtive activity, avoidance behaviour	-	-
Constructions (n=5)	0	-	3*	Unresolved land tenure (private), overlap with indigenous lands, leniency towards repairs, demolitions are rare
Tourism (n=4)	2	Furtive activity	2	Carrying capacity has not been defined
Fishing (n=4)	0†	-	4‡	Difference between (legal) subsistence and (illegal) commercial/sport fishing has not been operationalized
Fires (n=2)	2	Furtive activity	-	-

Respondents estimated that most of their enforcement created only very low deterrents. For 47 (87%) of individual priority threats, park staff stated that violators faced “very low” enforcement deterrents (i.e. less than 1% of the estimated expected benefits) when committing infractions within the past three years (Table 5). For three further priority threats, enforcement deterrents were still estimated to be “low” (1-10% of estimated benefits). Only two parks reported “high” deterrents (50-100% of estimated benefits) for a given priority threat: The staff of Isla de Salamanca had recently cracked down on illegal charcoal production with the support of the local police and army, a process that involved confiscation of boats, three days of minimum detention in prison, and criminal trials. The tiny sanctuary Los Colorados reported daily routine patrols to a roadside settlement within park boundaries to halt new constructions.

Key challenges to enforcement were roughly equally distributed between priority threats: Among 50 individual priority threats with low enforcement deterrents, 22 (44%) faced detection challenges and 28 (56%) faced sanctioning challenges. Distribution of enforcement challenges exhibited clear threat-specific patterns (Table 6): Furtive infractions, such as logging, hunting, extraction of flora and fauna, and fires, were more likely to face detection challenges. In contrast, infractions associated to a permanent field presence of the violator or his possessions, such as agriculture, construction, and livestock grazing (especially within fenced areas) were more likely to face sanctioning challenges. Illegal fishing in bogs, lagoons and coastal areas was reported to be more easily detectable than other furtive threats (with the exception of rare dynamite fishing), but also faced sanctioning challenges.

Workshop participants provided rather consistent responses about the reasons behind challenges to enforcement (Table 6). Low rates of sanctioning were generally attributed to the absence of an unambiguous legal basis recognized by all agencies responsible for enforcement, especially regarding land tenure. The large majority of assessed parks (13 of 15) contained areas claimed by private (and non-indigenous) actors as their property. Such claims included properties predating park establishment, as well as more recent settlements of domestic conflict refugees, protégées of armed groups, or wealthy second home owners. Claims would often be reinforced through agricultural, pastoral or residential use. In addition, five parks overlapped with indigenous reservations, the inhabitants of which were considered immune to park regulations both inside and outside of reservation boundaries. Sanctioning of illegal fishing (four parks) suffered from the absence of a clear, official, and operational distinction between legal subsistence fishing and other types of fishing; park staff tended to focus enforcement on small subsets of illegal fishing were the distinction was clearest (e.g. spearfishing and angling in Tayrona) or where local extinction was imminent (e.g. clam fishing in Isla de Salamanca). Tourism access restrictions were reportedly only enforced in parks whose carrying capacity had been officially estimated and implemented through a monitoring system for visitor frequency.

Detection challenges were mostly associated with furtive activities, i.e. those requiring only a short period of detectable illegal activity in the field (less than a day) to generate benefits to violators (Table 6). For logging and hunting, even parks with high patrolling density (e.g. Tayrona, Los Colorados) estimated to detect violators in less than 1% of the cases. Ten parks (67%) contained areas that had not been visited by guards for years, mostly due to limited accessibility; in six, such areas covered at least 50% of the territory. Furthermore, patrolling patterns in most parks allowed violators to adapt furtive activities to avoid detection. Guards would not regularly patrol outside of daylight hours; the only notable exception was nocturnal monitoring of turtle nests on the beaches of Tayrona and Sierra Nevada de Santa Marta. Only four parks scheduled patrols on weekends, and many reported evidence of higher infraction frequency during these days. Livestock grazing faced detection challenges in several parks where grazing animals were unmarked and could not be confiscated due to the absence of adequate municipal detention facilities.

Respondents frequently referred to the conflict potential of enforcement as an explanatory factor for low detection and sanctioning rates. In twelve parks (80%), park staff reported specific incidences of risks to their well-being resulting from enforcement in the recent past. These ranged from verbal aggression, ostracism, and damage to park equipment (punctured tires of park vehicles, stolen signposts, destruction of cabins) to threats of physical aggression (including machetes, fishing dynamite, and firearms). Assassinations of park employees, while infrequent, had occurred until the early 2000s, and several parks reported a presence of armed groups within and around their territory. To reduce the risk of retaliatory actions, park guards reported that they occasionally avoided high-conflict tasks, such as the formal identification of violators, confiscations, and the opening of sanctioning processes. Many park guards also indicated that they refrained from obligatory sanctioning if violators were poor, displaced, or had been living in or next to the park for long periods of time. Spatial strategies to conflict avoidance were rarely reported, and most guards reported to patrol high-conflict areas with higher frequency.

Discussion

Protected area regulations are believed to contribute to conserving biodiversity wherever mechanisms are present that incite regulatory compliance. If deterrence through enforcement is such a mechanism, as often suggested, then this mechanism seems currently unlikely to make a substantial contribution to the reduction of priority threats in the studied Colombian parks. According to park employees, most of the estimated enforcement deterrents were very low and unlikely to make a significant difference in the decision making of violators. Certainly, the observation of low enforcement deterrents does not preclude the possible existence of other mechanisms through which rules reduced threats. Indeed, there is ample evidence that individuals can be influenced to conform to formal rules or moral standards without the need for formal enforcement and economic sanctions, e.g. through moral suasion (Cialdini et al. 2006; Stern 2008; Ferraro & Price 2013). Neither did this study look at threats that potentially existed prior to the assessed time period but disappeared due to successful enforcement. However, the presence and persistence of detectable illegal activities supports the evidence for limitations to enforcement, and many park staff openly acknowledged and expressed frustration over the weaknesses in their park's prevailing enforcement regime.

An increase in the enforcement capacity of the studied parks, if desired, will likely take more than mere investments into guards, equipment, and infrastructure. Results provide a number of hypothetical pathways: One, the resolution of land tenure conflicts may be a precondition for improved regulatory compliance – a finding that would echo recent insights from 66 protected areas in the neighbouring Brazilian Amazon (Nolte et al. 2013a). Two, effective detection and sanctioning in the threat categories of fishing and tourism seems to rest on legally consistent and operational distinctions between permitted and sanctioned activities. Three, detection probabilities for furtive activities such as hunting and logging are unlikely to increase as long as predictability of the timing and location of patrols allow violators to engage in simple and inexpensive avoidance strategies. Four, efforts to increase enforcement capacity in the studied parks may fall short of expectations if they do not incorporate strategies to protect parks guards from perceived risks of retaliation.

Incorporating multiple threats, parks and enforcement steps in a single empirical analysis was helpful to identify these major patterns in enforcement. However, several limitations arise from the use of cross-sectional enforcement data. First, challenges in early enforcement steps (e.g. detection rates that are extremely low or zero) can impede the observation of potential challenges in later steps (e.g. sanctioning). Second, potential interactions between enforcement challenges may distort overall findings. For instance, the anticipation of potential conflict or slow legal processes could undermine the enthusiasm of park guards to detect or report infractions, but the magnitude of this effect remains largely indeterminable.

The identification of core challenges to enforcement is a first important step towards improving regulatory compliance in tropical parks. Improving the cost-effectiveness of enforcement, however, requires not only insights on the existence and respective importance of such challenges, but also on the estimated or actual cost-effectiveness of interventions to resolve them. Research into the respective costs and impacts of resolving land tenure, improving patrolling strategies, and protecting the well-being of park guards, among other things, would be valuable for the design of better enforcement strategies in Colombia and in other tropical countries.

CHAPTER 6: CONCLUSIONS

Following a decade of rapid growth and considerable global interest, the evaluation of the management effectiveness of protected areas has matured from a niche activity into a sizeable and established body of expertise, data, and knowledge, whose zenith is not yet in sight. PAME evaluations are carried out around the globe by many different organizations, at different scales, and with varying aims. PAME evaluation methods tend to evaluate a wide range of management aspects, following international recommendations that the "elements of [the management cycle][11] should, ideally, *all* be assessed if effectiveness of management is to be fully understood" (Hockings et al. 2006:11). However, the measurement of a key outcome of protected areas – conservation success – is often missing in PAME evaluations. Given that the conservation of biodiversity is a principal objective of protected areas, this situation seems disconcerting: Could it be it possible that widespread PAME evaluations measure something that has little to do with conservation success? And, if yes, what aspects of management should be measured to find out whether (and which) tropical protected areas are achieving this objective?

The four chapters of this dissertation seek to provide new answers to these questions by looking at the empirical relations between management aspects and outcomes of tropical protected areas. Complementing quantitative analyses of governance, management indicators, and avoided deforestation in hundreds of protected areas the Brazilian Amazon with field research on enforcement patterns in fifteen Colombia's national parks, they yield two overarching insights:

First, my research results challenges the notion that widespread PAME evaluation methods are able to measure the capacity of tropical forest protected areas to achieve their conservation objectives. Indicator scores of both METT and RAPPAM showed only weak associations with the empirical success of Brazil's Amazon protected areas to avoid forest fires and deforestation – arguably two of the most destructive threats to rainforest biodiversity. Those findings resonate with those of similar recent studies from the Brazilian Cerrado using RAPPAM (Carranza et al.

[11] Context, planning, inputs/resources, processes, outputs, outcomes

2014) and from Tanzania using METT (Lauren Coad, 2014, personal communication), which also observe that associations between PAME scores and anthropogenic land use change are weak. This pattern is a cause for concern, as METT and RAPPAM continue to be used for project prioritization and accountability by several of the world's largest conservation donors, including the World Bank, GEF, WWF, and many national governments. If such indicator systems are not closely linked to key biodiversity outcomes, relying on them for project prioritization and management could influence the design and implementation of protected area interventions in ways that reduce their conservation performance. It seems therefore important to understand how PAME indicator systems can be improved as to better reflect the capacity of a protected area to achieve their core objective.

Second, several chapters suggest that land use rights – their design, clarification, and enforcement – are key to understanding and predicting conservation outcomes in tropical protected areas. Governance types allowing sustainable resource use avoided significantly less deforestation than strict protected areas in the Brazilian Amazon (Chapter 3). The presence of unsettled land use rights was the only of 90 RAPPAM indicators to be significantly and consistently associated with failure to avoid deforestation (Chapter 4). In Colombia, interviewees consistently linked their failure of enforcing park regulations against deforestation threats to the unsettled land tenure situation within their park (Chapter 5). On one hand, these results may not seem surprising: After all, a protected area, as its very core, is a spatially explicit definition of rights, and as such exists only insofar as those rights are enforceable and enforced. On the other hand, however, land rights seem to have been assigned a back seat in the development of many widespread PAME methods: RAPPAM's indicator on land tenure and use right conflicts almost disappears within the set of 90 questions; and METT does not contain a land tenure indicator.

How is it possible that the two most widespread and accepted PAME evaluation methods have such difficulties in predicting a protected area's capacity to deliver conservation outcomes? In the absence of better evidence, answers remain speculative. However, several characteristics of METT and RAPPAM provide clues for a hypothesis: method developers may have sacrificed precision and predictive power in order to make their methodology attractive to their principal audiences – donors, governments, and NGOs. Three examples:

- **Form over function:** METT has 30 indicators; RAPPAM has 90, packaged in modules of five or ten. All are measured on the same scale (0 to 3 and “yes” to “no”, respectively). While this design appears attractive, there is no *a priori* reason why management effectiveness is best approximated by “round” numbers of identically scaled indicators. Instead, it seems possible that developers have occasionally compromised potentially optimal indicator numbers, choice and measurement to arrive at the desired format.

- **Indicator choice: “everything is important”:** METT and RAPPAM cover a wide range of indicators, including management aspects such as site design, planning processes, communication strategies, research programs, staff training, education programs, visitor facilities, and economic benefits, among many others. Such thematic breadth can be read as a testimony to the enormous diversity of expectations that are placed on protected areas by different interest groups. However, as results of this dissertation suggest, some management aspects may be more important than others in achieving core protected area objectives, and they may interact in complex ways. In spite of this, both RAPPAM and METT refrain from providing – potentially controversial – guidance on indicator analysis, and thus leave such decisions to the evaluation team. As a result, the most common interpretation method applied to RAPPAM and METT indicators consists of a simple averaging of all indicators, resulting in aggregate “effectiveness scores” whose meaning remains essentially obscure.

- **Subjective and imprecise language:** Many METT and RAPPAM indicators contain subjective qualifiers, such as “adequate”, “appropriate”, and “relatively”; this is particularly true for management aspects one would assume to be crucial for protected area performance, e.g. a protected area’s budget, staff, and equipment. Choosing subjective and imprecise wording may have been a response to the difficulties and controversies involved in defining standards or benchmarks for those indicators. However, a respondent’s judgment of “adequacy” is a function of his or her personal expectations and experience, which can vary across time and space. Protected area managers have also been shown to easily misinterpret the scope, scale, and timeframe of evaluations (Cook et al. 2014). Subjective and imprecise language can thus jeopardize the validity of comparative analyses that METT and RAPPAM were designed for.

A Way Forward

Improving the current practice of management effectiveness evaluation in protected areas is possible, but will require better theory and empirical knowledge. To inform decision making, we need to better understand the relations between management factors, conservation performance, and their cost-effective measurement. In my view, a comprehensive research program that aims to more fully inform and improve the evaluation and management of tropical protected areas could prioritize the following thematic directions:

- **Examining associations between diverse PAME methods and diverse conservation outcomes in different world regions and ecosystems.** Management indicators offer a low-cost alternative to biodiversity monitoring if they can predict conservation performance. However, results of this dissertation suggest that they do not – at least not in the studied cases. It therefore seems important to examine whether these findings hold true in other contexts: What are relationships between existing PAME indicators and anthropogenic land use change in other world regions and major ecosystems? Do different PAME indicator systems, whether widespread or not, exhibit better predictive capacity for conservation outcomes? What PAME indicators are closely associated with reductions in other categories of anthropogenic threats to biodiversity, such as logging and hunting? Efforts to extend the analyses of this dissertation to these new dimensions will be limited by the type, extent, and accessibility of existing PAME and outcome datasets. However, over time, such research efforts can contribute to a more robust evidence base on the predictive powers of current PAME evaluation systems.

- **Developing better indicators for management aspects believed to be critical for conservation success.** If protected area budget, staff, and equipment are indeed essential to conservation, as many experts seem to concur, those management aspects may have to be better measured to make them useful for effectiveness analyses and performance evaluations. Increasing the specificity, precision, and verifiability of indicators may be an important step forward. Subjective judgments of adequacy can supplement more objective measurement, but do not need to replace it. I applied those suggestions in the development of a new method to measure enforcement capacity (Chapter 5). Whether or not it allows a better prediction of conservation performance will require further empirical examination.

- **Developing conceptual models of the relative importance and interactions of management aspects.** Probably the largest void in the literature on PAME evaluation is the absence of conceptual models with explicit and testable assumptions about the relative importance of different management aspects and the functional interactions through which they influence conservation outcomes. The widespread practice of averaging diverse indicators implicitly treats all management aspects as equally important, additional and, as such, substitutable. However, as the example of the enforcement chain illustrates, different elements of management can interact in complex ways, and those interactions need to be understood and taken into account to predict protected area performance. The development of early conceptual models could benefit from an empirically grounded approach, e.g. by eliciting and aggregating the beliefs of protected area managers and decision makers on the importance and interactions of specific management aspects through causal cognitive mapping (Montibeller & Belton 2006; Klenk & Hickey 2011).

- **Testing for the existence and strength of causal linkages:** Associational analyses, as those presented above, can help to predict protected area performance from management indicators. However, in order to *improve* performance, decision makers need to make assumptions on *causal* directions and impacts: whether, how, and under which conditions will improvements in a given management aspect lead to improved conservation performance? Isolating, quantifying, and comparing causal impacts poses significantly higher challenges than associational analyses (Morgan & Winship 2007). Indeed, the success of such a causal research endeavor largely depends on the researcher's capacity to find or create conditions under which causal impacts can be estimated, which tends to be difficult in the field of conservation even for single intervention types (Margoluis et al. 2009). Given the diversity of interventions of potential interest to the conservation community and the costs involved in rigorous impact evaluations, causal research would probably be most cost-effective if it focused on only a small number of competing interventions that are assumed to exhibit close association to conservation success.

Results from earlier research steps, such as associative analyses and causal cognitive maps, could help to identify these priority interventions. The results of this dissertation also suggest that the clarification of land use rights in protected areas merits a closer look. Once the set of potential interventions has been narrowed down to a small set of interventions of interest, appropriate experimental or quasi-experimental studies can be designed and implemented.

- **Incorporating costs:** Given budget constraints, decisions on the prioritization of conservation interventions require an understanding of the relationship between the costs and expected impact of alternative scenarios. Comprehending and comparing the cost-effectiveness of competing conservation interventions for tropical protected areas thus requires knowledge not only of the causal direction and size of impacts, but also of the functional relationship between cost and such impacts. Empirical estimates of such relationships, however, are rarely found in the literature (Ferraro et al. 2011), probably owing to the many considerable challenges involved in their estimation. However, even simple empirical analyses of the implementation costs of conservation interventions is surprisingly rare in the literature. Therefore, even basic research on the actual allocation of conservation funds to protected areas and interventions types, coupled with an estimation of intervention impact, would already constitute a significant step forward towards a better understanding of cost-effectiveness.

Protected areas are likely to remain the most widespread instrument for the conservation of tropical biodiversity for decades to come. Evaluating their management effectiveness can be a valuable instrument for prioritizing and tracking interventions, and, ultimately, improving conservation performance. However, as this dissertation suggests, currently widespread evaluation methods may merit a process of revision and improvement if they are to become an effective vehicles for decision making. My research provides first insights on where and how such improvement is possible, and how it can be supported through empirical research. Even if we may never see a complete and empirically validated causal model that explains conservation performance of protected areas, improvements over the current *status quo* are certainly possible.

It is a long path towards more cost-effective conservation. Let’s start walking.

APPENDICES

APPENDIX A: SUPPLEMENTARY INFORMATION TO CHAPTER 2

Repeated sampling

All our estimates of the effect of protected areas on fire occurrence are based on 30 runs of repeated random sampling from the entire population of forest parcels in the Brazilian Amazon (2% and 5% for the first and second analysis, respectively), subsequent nearest neighbor matching (NNM), and averaging of the effect estimates and variances. We are not aware of precedents for this estimation strategy. But earlier analyses have shown that bootstrapping, a related technique allowing for the estimation of sampling distributions and confidence intervals through resampling, is invalid for NNM (Abadie & Imbens 2006).

For the purpose of our analysis, in which the argument is based on the direction and relative strength of estimated effects, we argue that our use of repeated sampling provides estimates that are as least as good, if not better, than those generated without repeated sampling. First, the estimation strategy that underlies each single run is well-established in the literature, all of which are based on NNM of simple random samples of forest parcels from a larger population (Andam et al. 2008, 2010; Gaveau et al. 2009; Joppa & Pfaff 2011; Nelson & Chomitz 2011). Second, when running our analyses with different random samples, we found that effect estimates varied between runs (Figure A1). Averaging across these effect estimates is likely to provide a more robust estimate of a protected area's effect on fire occurrence than reporting the results of any single run alone.

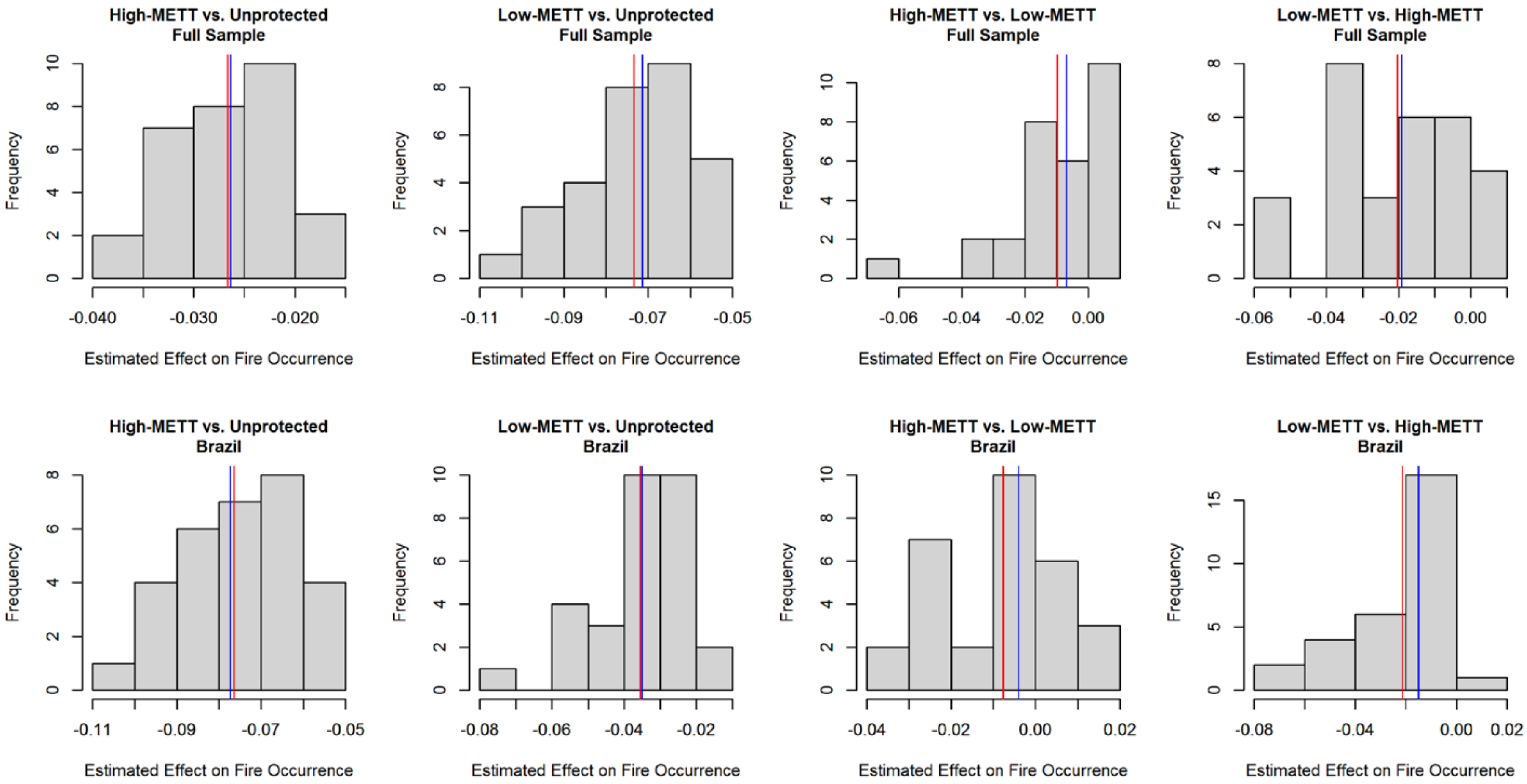

Figure A1: Histograms of estimates of protected area effects on fire occurrence across 30 runs with mean (blue line) and median (red line)

As a measure for the variance of our estimates, we averaged the Abadie-Imbens variance estimator (Abadie & Imbens 2006), used throughout earlier matching studies (Andam et al. 2008, 2010; Gaveau et al. 2009; Joppa & Pfaff 2011; Nelson & Chomitz 2011), across 30 runs (see also Methods). In addition, we provide histograms of all individual effect estimates (Figure A1). In conjunction with the results of our sensitivity checks below, our data highlight that although our general conclusions seem to be relatively robust to parameter changes, matching-based estimates of treatment effects remain vulnerable to both the choice of matching parameters and the characteristics of the chosen random sample, a source of variance that should be reported in future matching studies.

Matching Quality: Covariate Balance

The goal of matching is to create control groups that are similar to the treatment group in terms of their covariate distributions. Tables A1 and A2 show how covariate balance between treatment and control groups has improved through matching. Each reported value (means, differences, etc.) represents the mean value of 30 runs. With only a few exceptions, the standardized difference between the covariate means of treatment and control group has dropped considerably. A similar observation can be made for the mean difference of the empirical quantile-quantile plots (QQ plots). Thus, although small differences remain, the created control groups can be argued to be far more similar to the treatment group in terms of both the probability of treatment and fire occurrence than if control units had been sampled randomly from the pool of potential controls.

Table A1: Covariate balance before and after group matching (full sample)

	Full Sample Covariate		Mean Value Treatment Parcels	Mean Value Control Parcels	Standard-ized Difference	Mean difference QQ plot
High-METT vs. Unprotected	Average Elevation (m)	Unmatched	318.07	247.25	14.59	0.032
		Matched	315.47	312.78	0.56	0.014
	Average Slope (degree)	Unmatched	6.13	4.36	19.32	0.084
		Matched	6.08	6.05	0.37	0.023
	Travel Time to Major Cities (min)	Unmatched	1744.98	1545.57	20.98	0.066
		Matched	1743.58	1704.78	4.08	0.011
	Annual Precipitation (mm)	Unmatched	2382.74	2202.39	36.13	0.056
		Matched	2382.27	2370.75	2.31	0.009
	Distance to Forest Edge (km)	Unmatched	29.12	24.36	21.31	0.032
		Matched	29.16	28.89	1.19	0.005
Low-METT vs. Unprotected	Average Elevation (m)	Unmatched	200.04	247.25	-34.88	0.037
		Matched	200.17	181.21	13.80	0.048
	Average Slope (degree)	Unmatched	4.44	4.36	2.01	0.108
		Matched	4.42	4.73	-7.87	0.028
	Travel Time to Major Cities (min)	Unmatched	2365.73	1545.57	54.74	0.138
		Matched	2356.90	2273.24	5.59	0.018
	Annual Precipitation (mm)	Unmatched	2492.33	2202.39	52.21	0.090
		Matched	2493.13	2519.53	-4.70	0.022
	Distance to Forest Edge (km)	Unmatched	46.44	24.36	59.38	0.159
		Matched	44.08	42.67	4.05	0.015
High-METT vs. Low-METT	Average Elevation (m)	Unmatched	315.60	198.67	24.32	0.085
		Matched	215.06	219.10	-1.64	0.050
	Average Slope (degree)	Unmatched	6.09	4.40	18.56	0.125
		Matched	4.79	4.22	8.73	0.088
	Travel Time to Major Cities (min)	Unmatched	1747.34	2377.36	-65.99	0.116
		Matched	1832.29	1740.00	9.51	0.035
	Annual Precipitation (mm)	Unmatched	2382.22	2491.14	-22.02	0.042
		Matched	2490.72	2429.97	12.40	0.060
	Distance to Forest Edge (km)	Unmatched	29.13	46.52	-78.02	0.150
		Matched	32.36	30.24	9.06	0.041
Low-METT vs. High-METT	Average Elevation (m)	Unmatched	199.48	317.27	-86.89	0.086
		Matched	207.23	191.83	8.22	0.044
	Average Slope (degree)	Unmatched	4.42	6.11	-44.31	0.126
		Matched	3.31	3.57	-5.46	0.077
	Travel Time to Major Cities (min)	Unmatched	2375.67	1747.83	41.84	0.115
		Matched	1961.64	1878.05	7.06	0.028
	Annual Precipitation (mm)	Unmatched	2491.54	2382.33	19.61	0.042
		Matched	2396.85	2368.52	4.03	0.028
	Distance to Forest Edge (km)	Unmatched	46.55	29.18	46.77	0.150
		Matched	24.90	23.70	6.08	0.021

Table A2: Covariate balance before and after group matching (Brazilian sample)

	Brazilian Sample Covariate		Mean Value Treatment Parcels	Mean Value Control Parcels	Standard-ized Difference	Mean difference QQ plot
High-METT vs. Unprotected	Average Elevation (m)	Unmatched	105.31	146.57	-58.36	0.072
		Matched	104.51	108.27	-5.41	0.016
	Average Slope (degree)	Unmatched	3.21	3.01	6.84	0.076
		Matched	3.18	3.20	-0.53	0.018
	Travel Time to Major Cities (min)	Unmatched	1817.47	1513.69	28.16	0.074
		Matched	1821.87	1784.56	3.46	0.012
	Annual Precipitation (mm)	Unmatched	2391.95	2238.84	38.13	0.068
		Matched	2394.77	2386.10	2.17	0.025
	Distance to Forest Edge (km)	Unmatched	28.64	22.67	26.81	0.049
		Matched	28.61	28.27	1.51	0.007
Low-METT vs. Unprotected	Average Elevation (m)	Unmatched	188.27	146.57	48.53	0.091
		Matched	180.85	171.47	10.89	0.033
	Average Slope (degree)	Unmatched	4.15	3.01	38.94	0.125
		Matched	3.99	4.15	-5.38	0.019
	Travel Time to Major Cities (min)	Unmatched	2612.36	1513.69	74.33	0.190
		Matched	2597.53	2529.94	4.57	0.027
	Annual Precipitation (mm)	Unmatched	2475.29	2238.84	46.68	0.098
		Matched	2489.65	2517.50	-5.20	0.025
	Distance to Forest Edge (km)	Unmatched	51.91	22.67	77.59	0.218
		Matched	47.66	46.23	4.13	0.023
High-METT vs. Low-METT	Average Elevation (m)	Unmatched	105.39	188.64	-117.96	0.238
		Matched	100.28	103.27	-4.40	0.032
	Average Slope (degree)	Unmatched	3.22	4.18	-31.81	0.135
		Matched	2.43	2.22	10.32	0.091
	Travel Time to Major Cities (min)	Unmatched	1816.32	2618.85	-74.54	0.152
		Matched	1864.33	1802.79	5.61	0.031
	Annual Precipitation (mm)	Unmatched	2391.22	2474.94	-20.88	0.058
		Matched	2395.69	2351.61	10.43	0.051
	Distance to Forest Edge (km)	Unmatched	28.54	51.94	-104.99	0.195
		Matched	27.97	24.53	15.88	0.060
Low-METT vs. High-METT	Average Elevation (m)	Unmatched	188.47	106.03	95.79	0.235
		Matched	174.63	163.91	10.91	0.079
	Average Slope (degree)	Unmatched	4.16	3.25	31.09	0.131
		Matched	2.94	3.50	-21.67	0.097
	Travel Time to Major Cities (min)	Unmatched	2612.76	1814.81	54.04	0.151
		Matched	1987.24	1838.03	13.15	0.041
	Annual Precipitation (mm)	Unmatched	2474.68	2388.91	16.94	0.057
		Matched	2243.57	2243.61	-0.01	0.042
	Distance to Forest Edge (km)	Unmatched	51.67	28.50	61.50	0.194
		Matched	25.09	23.44	8.28	0.042

Sensitivity Tests

In the following, we justify the choice of our parameter settings for sampling, group splits, and matching. We then test the sensitivity of our effect estimates to the choice of these parameter settings by repeating the group matching analyses presented in the main paper with different parameter settings, and comparing the resulting effect estimates.

Designation Year

Our initial goal was to only include protected areas that existed during the period over which the outcome – deforestation fires – were monitored, i.e. 2000-2010. Protected area designation is usually preceded by long processes of public consultation and deliberation, during which deforestation patterns can already change, making the WDPA-listed designation year an imperfect indicator for the time after which we would expect observable effects. Enforcing 2000 as the upper limit on a site's designation year also reduced the original samples from 41 (three countries) and 33 (Brazil) observations to 26 and 19, respectively, resulting in a non-negligible loss of statistical power and decreased the subgroup balance of pixel numbers (39K:122K in the full sample). Choosing 2002 as the cutoff year increased sample size considerably (37 for all countries and 29 in Brazil) and provided better balance in terms of number of forest pixels (95K:137K), while still allowing the assumption that avoided deforestation in the observed period could mainly be attributed to the existence of the protected area. We therefore ran our analyses with 2002 as the cutoff year, but tested for the sensitivity of results to a cutoff year in 2000.

Group Splits: Equal site number vs. equal area

Our group comparison presented in the main paper splits the protected area sample into subgroups which have roughly the same number of sites (±1), in order for both groups to contain the same number of METT score observations. As protected areas vary in size, however, this resulted in an imbalance in terms of the number of forest pixels each group contained. We therefore ran additional analyses that split the protected area sample into groups that were more equal in terms of forest pixels, to test for the sensitivity of our results to this parameter choice.

Group Splits: 50%-50% vs. Upper AND LOWER QUARTILE

As illustrated in the main paper, effect estimates were strongly influenced by a slight change in the threshold on the METT composite score (from 1.33 to 1.22) used to split our sample into low and high-METT protected areas. We therefore tested whether differences between estimated effects of both groups on fire occurrence would be equally small when dropping protected areas close to the METT threshold. We therefore repeated our analysis with protected areas from the upper and lower quartile of METT composite scores as the high-METT and low-METT group, respectively. We are grateful to an anonymous reviewer for this suggestion.

Calipers: 0.5 vs. 1.0 vs. none

Calipers are imposed similarity thresholds for the maximum tolerated covariate distance between matched observation pairs. Treated observations for which no counterfactual untreated observation can be identified within caliper boundaries are considered incomparable and therefore excluded from the analysis. Calipers are usually defined in terms of the standard deviation of each covariate. Setting calipers involves trading off low bias (good covariate balance) for representativeness: Matching with small calipers (e.g. 0.25 standard deviations) will retain only very similar observations, but may drop many observations, reducing the analyst's ability to draw conclusions for the full sample. Matching with large or no calipers will retain many or all observations, but introduce substantial bias and covariate imbalance. We chose calipers of one standard deviation and tested for the sensitivity of results and statistical support to the application of stricter calipers (0.5 standard deviations).

Sensitivity tests: Results, covariate balance, and support

Figure A2 contrasts the matching quality of our alternative parameter sets (30 runs each) for high-METT vs. low-METT subgroups in terms of two indicators of interest in matching studies, namely 1) the level of statistical support, i.e. the ratio of retained observations after applying calipers and exact matching by country/state (values close to 1 are desirable), and 2) the degree to which matching achieved covariate balance, calculated as the means of the mean difference of empirical quantile-quantile plots across 30 runs and five covariates used in the Mahalanobis distance matching (values close to 0 are desirable). As Fig. A2 indicates, our preferred parameter set (on which our manuscript is based) leads to matching results of rather high quality both in terms of statistical support and covariate balance. Splitting the protected area samples into subgroups with similar areas provides estimates of comparable matching quality, especially in the case of Brazil. Imposing stricter calipers drastically reduces statistical support without necessarily improving covariate balance. Choosing 2000 as the cutoff year for site designation or matching upper vs. lower quartiles decreases both statistical support and covariate balance.

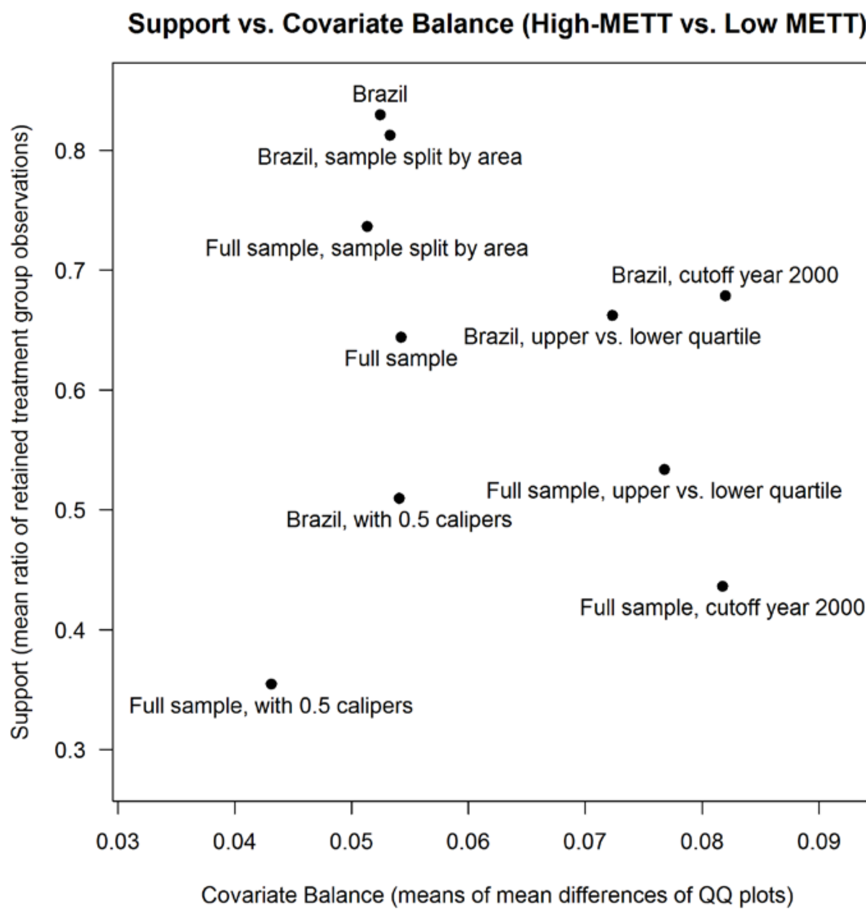

Figure A2: Statistical support and covariate balance of matching analyses with alternative parameter sets

Figure A3 provides effect estimates and standard deviations for each of our alternative parameter sets. Results indicate that the observations and conclusions drawn from the subgroup analysis in our paper would be similar for most parameter sets. Three exceptions apply. First, choosing the year 2000 as the cutoff for site designation not only worsens support and covariate balance, but also increases the volatility of our results. Brazilian estimates suggest that forest parcels in high-

METT sites have higher fire rates than matched controls from low-METT sites, whereas the relationship is inversed for the full sample. Second, splitting the entire sample by area suggests a strong effect for high-METT protected areas on fire occurrence. These results appear to be outliers and are based on a sample that appears somewhat imbalanced in terms of numbers of METT assessments (n_{low}:n_{high} = 24:13, with five of the high-METT observations coming from Peru and Bolivia). Third, considering only the upper vs. lower METT score quartiles of protected areas results in roughly similar effect estimates for the high-METT and low-METT groups (as compared to unprotected forest parcels) both for the full and the Brazilian sample, lending additional support to our conclusion that METT scores are weakly associated with success. Our observations highlight that the type of matching analysis we propose in this paper can remain vulnerable to parameter choice, and points to the importance of sensitivity tests in this and future analyses.

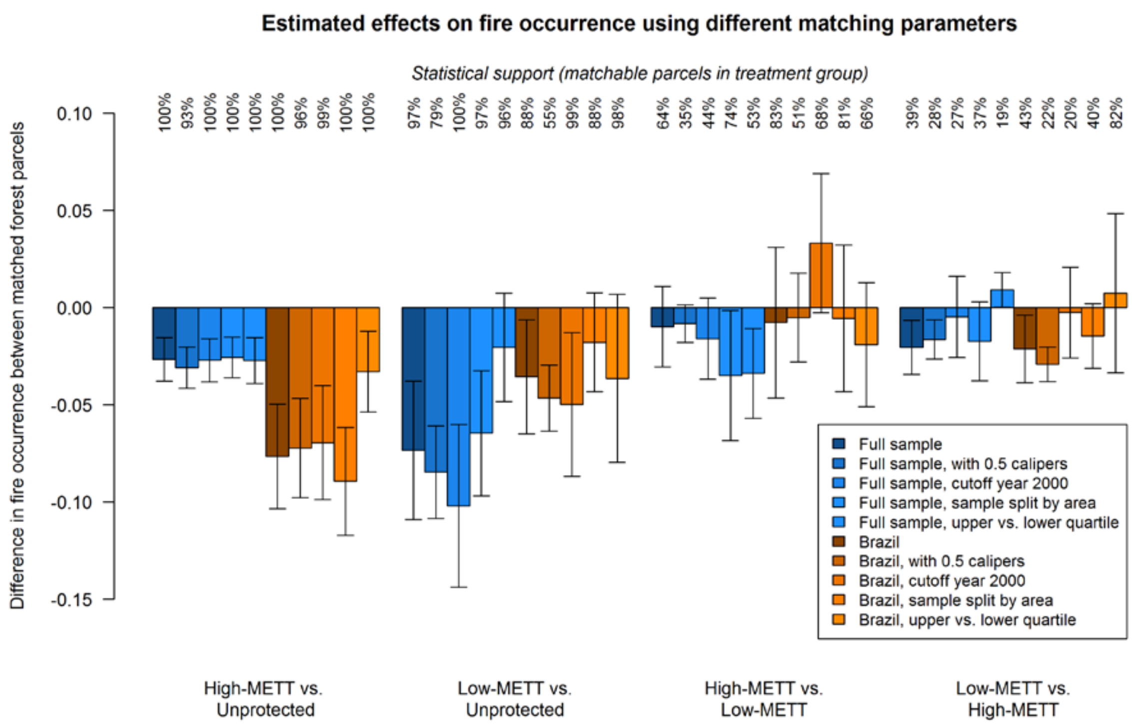

Figure A3: Estimates of effects on fire occurrence and standard deviations of sensitivity tests. Error bars proxy confidence intervals (average standard error * 1.96)

Differences between high-to-low-METT vs. low-to-high-METT matching

In all of our above analyses, we compared fire occurrence on similar forest parcels in high-METT and low-METT protected areas in two steps, first matching forest parcels from high- to low-METT areas, and then matching forest parcels from high- to low-METT protected areas. Regardless of the other matching parameters, the two comparisons generally lead to different sample sizes, different covariate distributions (see Table A1 and A2), and different effect estimates (see Fig. 1 and Fig. A3). At first glance, these results may seem counter-intuitive: In other statistical analyses, such as multiple regressions, a mere switch in the binary definition of "treatment" should lead to estimates that are the exact inverse of the estimates before the switch.

The explanation of the observed discrepancies between high-to-low vs. low-to-high METT matching lies in the way how the matching estimator affects the properties of the sample, i.e. the matched pairs of forest parcels. We conducted NNM matching with replacement in a context where covariate distributions differed between treatment forest parcels and potential controls: In spite of using rather generous calipers (1 SD), matching found comparable control parcels for an average of only 64% of forest parcels in high-METT protected areas in the full sample, and only 39% when matching in the other direction (Brazilian sample: 83% and 43%, respectively). Furthermore, observations of forest parcels tended to be clustered within the multidimensional covariate space (each protected area contributing forest parcels with a distinct multidimensional distribution of covariates). As we conducted matching with replacement, forest parcels were allowed to be recycled as controls for treated observations, and the number of times a forest parcel acted as a control may have been associated with either its probability of fire occurrence in the absence of protection, or the effect of protected areas in reducing it. For instance, it is not unlikely that forest parcels at the edges of distributions of individual protected areas were more likely to be recycled as control parcels. While an in-depth investigation of this phenomenon is beyond the scope of this paper, its existence does not seem to weaken our general conclusions about the weakness of associations between METT scores of protected areas and their effect on fire occurrence. Instead, we would recommend the switching of treatment definition as a robustness check for matching analyses whenever covariate distributions seem to differ considerably between treated units and potential controls.

METT Summary Statistics for Full Sample and Effectiveness Groups

Table A3 contains summary statistics of composite and individual METT scores for the full sample of protected areas for which individual effect estimates had been computed, as well as for each of the eight effectiveness groups defined by our definitions of relative effectiveness in reducing fire occurrence (Table 2). Our analysis of associations of individual METT indicators with “more” and “less effective” protected areas (Table 3) is based on the values in Table A3 (direction of difference and p-value of two-tailed t-tests).

Table A3: METT scorecard summary statistics* and score means of effectiveness groups

			Fire occurrence				Fire reduction				% Fire reduction				LOESS			
	Sample Mean	Standard Deviation	Mean of more effective areas	Mean of less effective areas	Difference	p-value, two-tailed t-test	Mean of more effective areas	Mean of less effective areas	Difference	p-value, two-tailed t-test	Mean of more effective areas	Mean of less effective areas	Difference	p-value, two-tailed t-test	Mean of more effective areas	Mean of less effective areas	Difference	p-value, two-tailed t-test
Composite METT Score	1.34	0.41	**1.35**	**1.32**	**0.03**	**0.83**	**1.38**	**1.31**	**0.07**	**0.64**	**1.40**	**1.29**	**0.12**	**0.46**	**1.39**	**1.21**	**0.19**	**0.26**
12: Staff numbers	0.93	0.53	0.83	1.09	-0.26	0.15	1.00	0.89	0.11	0.57	1.00	0.86	0.14	0.49	0.95	0.89	0.06	0.73
15: Current budget	1.07	0.65	1.06	1.09	-0.04	0.88	1.09	1.06	0.04	0.89	1.21	0.93	0.29	0.27	1.10	1.00	0.10	0.58
16: Security of budget	0.93	0.72	1.06	0.73	0.33	0.19	0.64	1.12	-0.48	0.06	1.00	0.86	0.14	0.62	1.05	0.67	0.39	0.13
18: Equipment	1.28	0.96	1.28	1.27	0.01	0.99	1.45	1.17	0.29	0.42	1.64	1.00	0.64	0.07	1.40	1.00	0.40	0.29
2: Regulations	1.41	0.82	1.22	1.73	-0.51	0.14	1.64	1.28	0.36	0.29	1.43	1.43	0.00	1.00	1.40	1.44	-0.04	0.89
3: Law enforcement	1.20	0.58	1.14	1.27	-0.13	0.57	1.40	1.07	0.33	0.15	1.18	1.23	-0.05	0.84	1.18	1.25	-0.07	0.75
6: Boundary demarcation	1.86	0.69	1.83	1.91	-0.08	0.76	1.91	1.83	0.08	0.80	1.71	1.93	-0.21	0.41	1.80	2.00	-0.20	0.41
28: Access assessment	1.04	0.69	1.06	1.00	0.06	0.82	1.27	0.88	0.39	0.14	1.29	0.77	0.52	0.05	1.16	0.78	0.38	0.18
30: Monitoring & evaluation	0.96	0.74	0.94	1.00	-0.06	0.85	1.09	0.88	0.21	0.50	1.00	0.92	0.08	0.80	0.89	1.11	-0.22	0.54
10: Research	0.96	0.79	1.06	0.80	0.26	0.38	0.90	1.00	-0.10	0.72	1.21	0.77	0.45	0.14	1.15	0.50	0.65	0.02
9: Resource inventory	1.28	0.84	1.33	1.18	0.15	0.63	1.27	1.28	-0.01	0.99	1.29	1.36	-0.07	0.82	1.40	1.00	0.40	0.21
7: Management plan	1.38	1.21	1.39	1.36	0.03	0.96	1.18	1.50	-0.32	0.51	1.50	1.29	0.21	0.65	1.45	1.22	0.23	0.63
8: Regular work plan	1.52	0.99	1.61	1.36	0.25	0.53	1.55	1.50	0.05	0.91	1.57	1.50	0.07	0.85	1.60	1.33	0.27	0.52
17: Budget management	1.35	0.80	1.33	1.36	-0.03	0.93	1.80	1.06	0.74	0.01	1.67	1.08	0.59	0.07	1.50	1.00	0.50	0.15
21: State & comm. neighbors	1.25	0.85	1.53	0.78	0.76	0.02	1.25	1.25	0.00	1.00	1.55	1.08	0.46	0.18	1.47	0.71	0.76	0.05
22: Indigenous people	2.15	1.07	2.00	2.50	-0.50	0.34	2.50	2.00	0.50	0.34	1.83	2.43	-0.60	0.36	2.00	2.50	-0.50	0.34
23: Local communities	1.16	1.03	1.07	1.27	-0.20	0.63	0.90	1.33	-0.43	0.29	0.69	1.67	-0.97	0.02	0.94	1.62	-0.68	0.11
27: Condition assessment	2.10	0.90	2.33	1.73	0.61	0.14	2.36	1.94	0.42	0.19	2.36	1.79	0.57	0.10	2.35	1.56	0.79	0.10

* Includes all protected areas with METT scores for which individual effects on fire occurrence had been calculated (n=29)

APPENDIX B: SUPPLEMENTARY INFORMATION TO CHAPTER 3

Data

Protected areas

We considered all protected areas included in the World Database of Protected Areas (WDPA) (UNEP-WCMC 2011) situated in the Brazilian Legal Amazon. We used spatial data from the 2010 version of the WDPA as it included the original boundaries of protected areas that had recently been subject to downsizing as a result of their failure in stemming deforestation (Araújo & Barreto 2010). For instance, the National Forest Bom Futuro had been significantly downsized in 2010 to exclude deforestation that had occurred between 2000 and 2010. We used 2012 data from the National Cadaster of Protected Areas (CNUC) of the Brazilian Ministry of the Environment to ensure that our pool of potential controls (unprotected forest parcels) did not any contain parcels situated in recently established protected areas or protected areas with expanded boundaries. We excluded from the pool of potential controls all unprotected forest parcels situated within 10km buffers around any protected area (both WDPA and CNUC) to reduce the vulnerability of our results to potential local spillover effects (Andam et al. 2008).

Deforestation

We used two different deforestation datasets to draw on their respective strenghts in detecting tropical deforestation. The fine-grained PRODES dataset published by the Brazilian Institute for Space Research (INPE) is based on ~30m resolution LandSat imagery and thus capable of detecting deforestation on relatively small patches of forests (Câmara et al. 2006). However, the low temporal resolution of LandSat imagery (biweekly images) hampers the detection of deforestation due to frequent cloud cover. PRODES' particularly high rate of error in early years (up to 2000) prompted us to use only 2001-2005 data for our first period of analysis. Our second deforestation measure, the Gross Forest Cover Loss (GFCL) published by the South Dakota State University (Hansen et al. 2010), is based on data from the Moderate Resolution Imaging

Spectroradiometer (MODIS). With daily return rates, MODIS satellites are more likely to encounter cloud-free conditions. However, the lower resolution of their sensors (~250m) reduces their ability to detect small-scale deforestation patches (Hansen et al. 2008). We ran separate analyses with both datasets and contrasted their respective results throughout.

Covariates

Probabilities of deforestation pressure and protection are influenced by a number of location-specific characteristics, most notably the suitability of a given plot for agriculture, ease of access, and distance to markets (Andam et al. 2008; Joppa & Pfaff 2010; Nelson & Chomitz 2011). We use the following covariates to control for differences in deforestation pressure

- *Agricultural Suitability:* Elevation and slope influence a forest parcel's suitability for agriculture (Nelson & Chomitz 2011). Similarly, the occurrence of seasonal flooding has been shown to influence agricultural suitability and the probability of forest conversion(Newton et al. 2011). We extracted average slope and average elevation from data provided by the International Institute for Applied Systems Analysis (Fischer et al. 2007), and identified seasonally flooded areas using the GlobCover 2005 dataset based on the European Space Agency's Envisat platform (Arino et al. 2009).

- *Forest Cover:* At ~1km resolution, low average tree cover on a forest parcel can indicate existing forest fragmentation and deforestation. Furthermore, the probabilities of forest conversion detected by GFCL are a function of baseline tree cover (Carroll et al. 2011). We used tree cover estimates provided by the MODIS-based Vegetation Continuous Fields dataset (Collection 3) to control for this covariate (Hansen et al. 2003).

- *Distance to Forest Edge:* Strongly influencing physical accessibility, distance to forest edge has been shown to be strongly associated with deforestation (Andam et al. 2008). We computed distance to forest edge as the shortest Euclidian distance of a given forest parcel to 1) parcels with less than 25% forest cover (VCF) 2) rivers (ESRI hydropolygons) and 3) major roads (Soares-Filho et al. 2006).

- *Travel Time to Major Cities:* Accessibility to markets is an important predictor of deforestation patterns (Nelson & Chomitz 2011). We used the algorithm, datasets and assumptions of an existing travel time dataset from the European Union's Joint Research

Center (Nelson 2008) to compute our own travel time estimates using 1) improved and more detailed Brazilian road data (Soares-Filho et al. 2006) as well as 2) a land cover map that reflected baseline land cover conditions in the year 2000 (MODIS Land) (Oak Ridge National Laboratory Distributed Active Archive Center 2011).

- *State*: Brazil's federal states can exercise considerable autonomy in devising state-level policies which can influence deforestation pressure and its spatial distribution. We use state boundaries provided by the Global Administrative Areas database (www.gadm.org) to control for this covariate.

We did not include distance to roads as a covariate in our analysis. Roads facilitate physical access to forest parcels and the transport of timber and agricultural products to markets. However, in the Brazilian Amazon, roads are only one element of transport infrastructure, with river travel being the main means of travel and transport in remote areas of the basin. We argue that 1) our estimates of travel time to major cities capture such interactions between road and river travel better than an estimate of distance to roads, and that 2) our estimates of distance to forest edge, with forest edge including major roads and rivers, capture the remainder of local-level variation in physical accessibility.

Methods

Estimating Deforestation Pressure

Matching is a quasi-experimental method that seeks to mimic random assignment of treatment by identifying artificial control groups of untreated units which differ from treated units in all relevant aspects but the treatment itself. Matching estimators rely on the assumption that treatment selection is on observables, i.e. that the observable covariates used in the matching procedure account for all differences between treatment and control units which are associated with both the probability of treatment (protection type) and outcome (deforestation). Given the absence of randomly controlled trials of the assignment of protection to forest parcels, an explicit test of the validity of this assumption is not possible. Assessments of the validity of matching estimators therefore have to rely on 1) a sound theoretical and empirical argument for the choice of covariates, as well as an 2) assessment of the extent to which matching was able to balance covariates between control and treatment groups.

Choice of Covariates. Above, we listed the covariates included in our matching estimator, together with an empirical and theoretical rationale for the inclusion of each. Controlling for baseline forest cover, political boundaries, agricultural suitability, accessibility, and distance to markets has been considered both necessary and sufficient by a large number of matching studies that assess the impact of protection on deforestation and/or forest fires (Andam et al. 2008; Joppa & Pfaff 2011; Nelson & Chomitz 2011; Pfaff et al. 2013; World Bank 2013). One study from Costa Rica tests the sensitivity of matching estimates to using an extended set of covariates, including poverty, population density, and immigration, and finds results to be similar (Andam et al. 2008). While we cannot explicitly test the extent to which matching successfully mimics random assignment, we consider the existing theoretical and empirical support for our choice of covariates sufficient to trust in the extent to which our estimator successfully controls for the most relevant joint bias in treatment assignment and deforestation outcomes.

Covariate Balance. Matching relies on the existence of a pool of control units whose covariates are sufficiently similar to the pool of treatment units to qualify as matches (statistical support). Whether matching has been successful can be assessed by comparing covariate distributions between treated units and control units both before and after matching. A commonly used indicator to assess such similarity is the mean difference of empirical quantile-quantile (eQQ) plots of covariates in the treatment and control group (Andam et al. 2008). To obtain an aggregate balance indicator for each of the 292 protected areas, we averaged the standardized mean difference of eQQ plots across 30 repetitions and our six continuous covariates (matching was exact for categorical covariates). We then examined the distributions of the 292 estimates using Kernel density estimators, weighing each balance indicator by the number of matched forest parcel. We also examined distributions for each protection type separately.

Our results indicate that matching dramatically improved covariate balance for all protected areas in our sample (Figure B7). Matching reduced the mean of our 292 balance estimates from 6.13 to 0.07. Furthermore, matching achieved similar improvements of covariates balance for all protection types (Figure B8), suggesting that remaining differences in covariates were not biased towards either protection type. We therefore consider our matching estimator to have successfully controlled for differences in observable covariates between forest parcels in control and treatment groups.

Dropped Forest Parcels. Causal inference through matching relies on the existence of control units that are sufficiently comparable to the pool of treated units to qualify as observations of counterfactual outcomes (statistical support). We follow earlier matching studies in removing protected forest parcels if no control parcels could be found within one standard deviation of each covariate (calipers). Calipers retained 91.5% of forest parcels from the treated sample, distributed roughly equally amongst protection types (Strict Protection 91.7%, Sustainable Use: 92.6%, Indigenous Lands: 90.9%). Visual inspection of the results suggests that protected areas with a high rate of dropped forest parcels are situated in both high and low pressure areas for all three protection types. The counterfactual outcome (deforestation pressure) cannot be observed for these dropped parcels. However, the large percentage of retained pixels and their distribution among protection types suggests that our results are likely to hold for the full sample of forest parcels.

Leakage. Leakage occurs when treatment influences the outcomes on untreated units. If protection of a given set of parcels leads to an increase (or decrease) deforestation on unprotected parcels, a comparison of protected and unprotected units will overestimate (or underestimate) the effects of protection. A recent study did not find evidence for leakage occurring as the result of the creation of protected areas in the Brazilian Amazon (Soares-Filho et al. 2010). Nevertheless, we limited the risk of an influence of differences in local leakage on our findings by excluding from our pool of potential control parcels a 10km buffer around all protected areas and military areas that had been created until 2010. Although protection types may differ in the extent to which they engender leakage, the fact that our pool of control parcels covers a vast region reduces the probability that controls of different protection types may be differently affected by the leakage problem. While we cannot rule out the possibility that leakage is occurring, we do not consider its possible existence to alter our findings about the differential impacts of protection types.

Density Estimation

We used Kernel density estimators to assess the skewness of the protection-type specific distributions of estimated deforestation pressure, and to examine the shift in these distributions that occurred between 2000 and 2005 as a result of newly designated areas in all categories. We used R's *density* function with a Gaussian kernel and default bandwidth computation, and weighted observations by the number of matched forest parcels. We estimated density for each protection type separately (Figure B2).

Transformations

We found that distributions of original deforestation pressure estimates were strongly skewed towards low levels of deforestation pressure (Figure B2, left). As a result, a small number of high-pressure protected areas were able to drive the differences in the aggregate estimates of pressure and impact (see main text). We also observed a strongly skewed distribution of observed deforestation rates, whose variance increased with higher estimated deforestation pressure (Figure B3). In order to reduce such heteroskedasticity and to allow for an estimation of pressure-specific effectiveness of protection types that would take advantage of the full sample, we transformed both observed deforestation rates and estimates of deforestation pressure. We did not use a logarithmic transformation due to the existence of real zeros in both variables. We found that a double square root transformation resulted in less skewed distributions and was therefore more amenable to subsequent regressions (Figure B2, right).

Regressions

Non-parametric. We used locally weighted scatterplot smoothers (LOESS, using R's *loess* function, span=1) to non-parametrically estimate observed deforestation rates as a function of deforestation pressure. 95% confidence intervals were computed based on the standard errors of the LOESS prediction. We applied separate LOESS estimators for each protection type, time period (2000-05 vs. 2006-10), protected area sample (established ≤ 2000 vs. ≤ 2005), and deforestation dataset (PRODES vs. GFCL), and compared the resulting functions (Figure 5, Figures B3-B6).

Linear regressions: We used linear regressions to test the strength of the differences in pressure-specific observed deforestation between protection types. We regressed observed deforestation rates on estimated deforestation pressure (both transformed) and included dummy variables for sustainable use areas and indigenous lands. We ran models with three distinct specifications for each dataset and time period: 1) without interactions between pressure and protection types 2) with interactions between pressure and protection types 3) with interaction between pressure and indigenous lands only (Table B1). The latter corresponds to our non-parametric observation that deforestation rates in indigenous lands responded differently to deforestation pressure than deforestation rates in strictly protected and sustainable use areas (see main text).

Table B1: Results of weighted regressions of observed deforestation rates on estimated deforestation pressure and protection types (transformed data)

Independent Variables	Without Interactions[12]	With Interactions	Interactions with Indigenous Lands only
PRODES Deforestation 2001-05, protected areas established ≤ 2000			
Intercept	0.086 ***	0.014	0.010
Deforestation Pressure (^ 0.25)	0.233 ***	0.498 ***	0.515 ***
Sustainable Use Area[13]	0.056 **	0.035	0.044 *
Indigenous Land	0.011	0.124 ***	0.128 ***
Sustainable Use Area x Pressure		0.029	
Indigenous Land x Pressure		-0.382 ***	-0.399 ***
[Adjusted R^2]	[0.259]	[0.390]	[0.394]
PRODES Deforestation 2006-10, protected areas established ≤ 2005			
Intercept	0.036 **	0.021	-0.004
Deforestation Pressure (^ 0.25)	0.351 ***	0.403 ***	0.489 ***
Sustainable Use Area	0.043 ***	0.009	0.050 ***
Indigenous Land	0.011	0.046 +	0.071 ***
Sustainable Use Area x Pressure		0.158 +	
Indigenous Land x Pressure		-0.135 +	-0.221 ***
[Adjusted R^2]	[0.351]	[0.386]	[0.381]
Gross Forest Cover Loss 2000-05, protected areas established ≤ 2000			
Intercept	0.018	-0.014	-0.023
Deforestation Pressure (^ 0.25)	0.327 ***	0.468 ***	0.510 ***
Sustainable Use Area	0.051 **	0.023	0.042 *
Indigenous Land	-0.025	0.029	0.039 +
Sustainable Use Area x Pressure		0.075	
Indigenous Land x Pressure		-0.211 *	-0.252 ***
[Adjusted R^2]	[0.478]	[0.525]	[0.526]

[12] Significance codes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, + $p < 0.1$

[13] Protection types are dummy variables. The omitted protection type is Strict Protection.

Independent Variables	**Without Interactions**[12]	**With Interactions**	**Interactions with Indigenous Lands only**
Gross Forest Cover Loss 2005-2010, protected areas established ≤ 2005			
Intercept	0.011	0.009	-0.010
Deforestation Pressure (^ 0.25)	0.395 ***	0.402 ***	0.474 ***
Sustainable Use Area	0.037 **	0.009	0.042 **
Indigenous Land	0.015	0.028	0.047 **
Sustainable Use Area x Pressure		0.143	
Indigenous Land x Pressure		-0.051	-0.124 *
[Adjusted R^2]	[0.423]	[0.434]	[0.431]

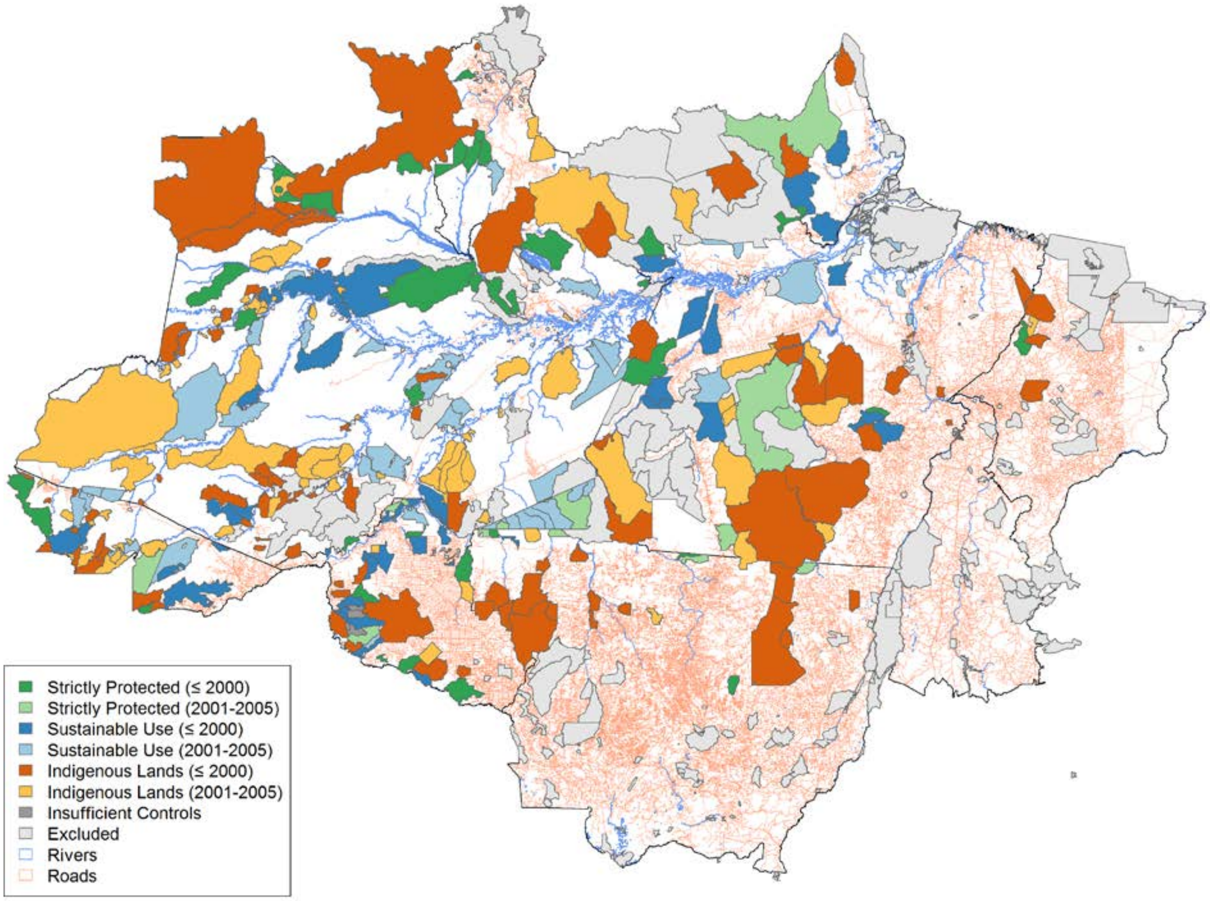

Figure B1: Schematic map of the Brazilian Amazon protected areas included in Chapter Three. Excluded areas include protected areas established post-2005, Environmental Protection Areas (APAs), and protected areas outside the humid forest tropical biome, with less than 50% tree cover, or with less than 200 forest parcels in 2000. Sources: (Soares-Filho et al. 2006; UNEP-WCMC 2011)

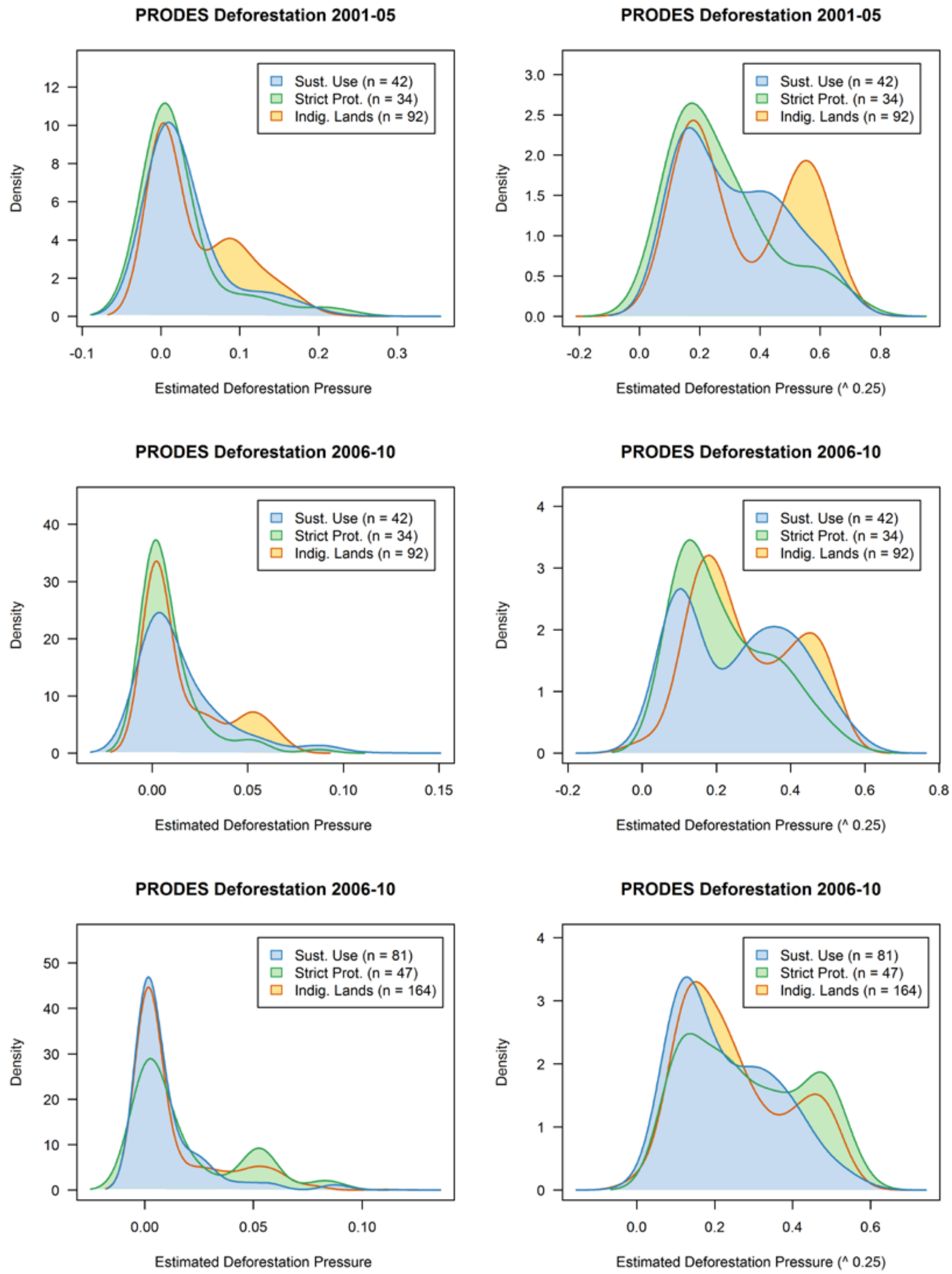

Figure B2: Density distributions of original (left) and transformed (right) deforestation pressure estimates for protected areas established ≤ 2000 (top and middle, with 2001-05 and 2006-10 estimates, respectively) and 2005 (bottom, 2006-10 estimates). Observations were weighted by the number of matched forest parcels.

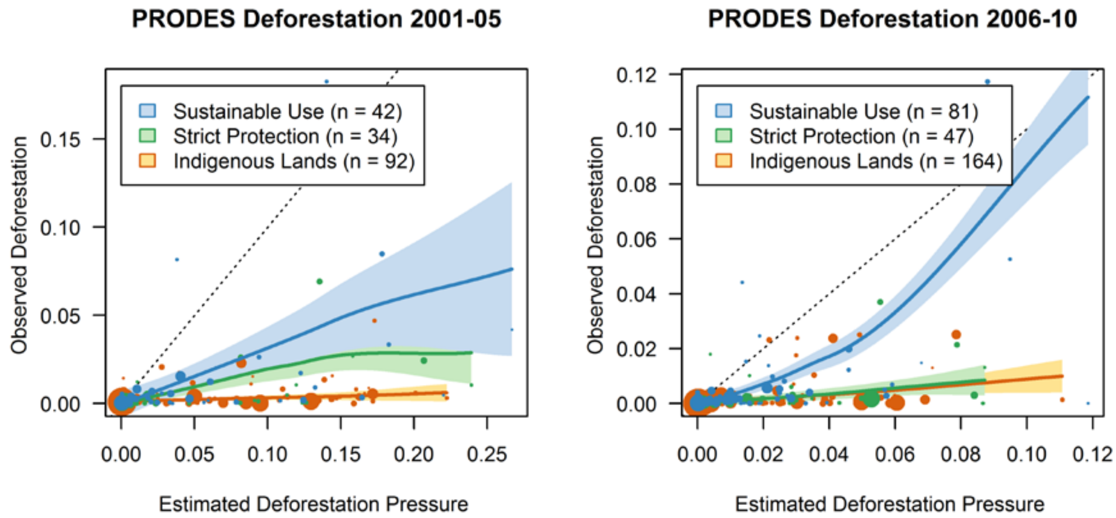

Figure B3: As Figure 5, but based on original data without transformation

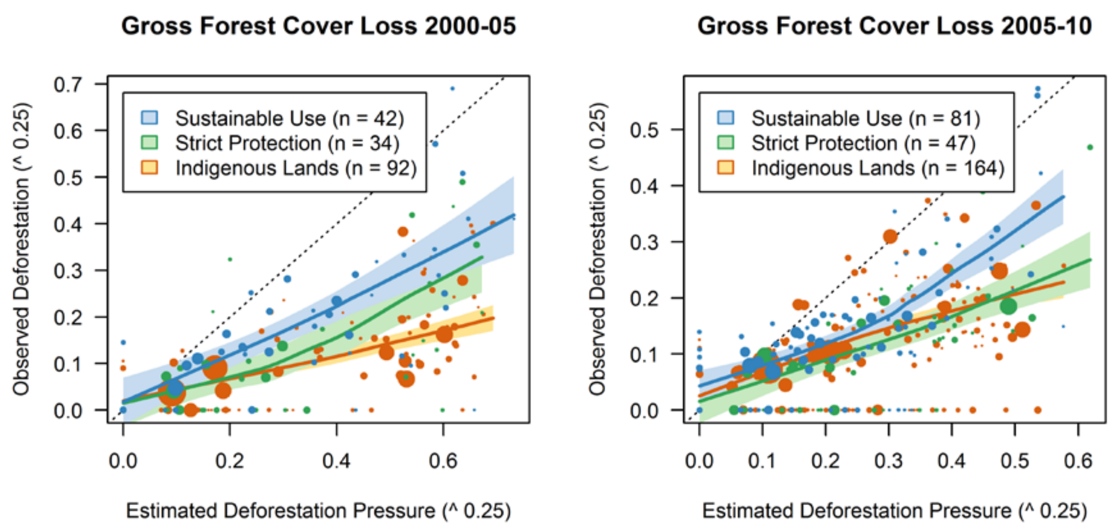

Figure B4: As Figure 5, but using Global Forest Cover Loss instead of PRODES

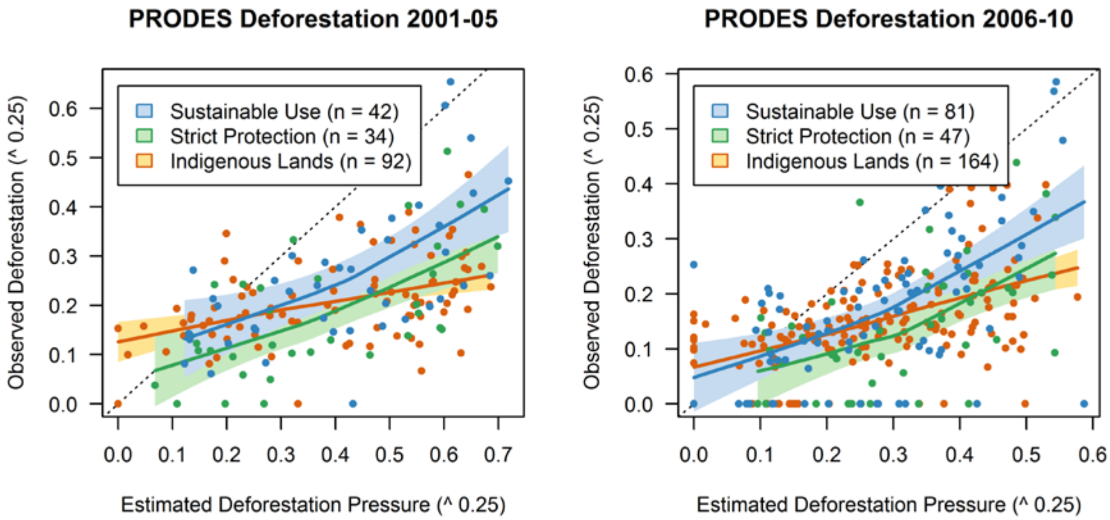

Figure B5: As Figure 5, but without weighting protected areas by number of matched forest parcels

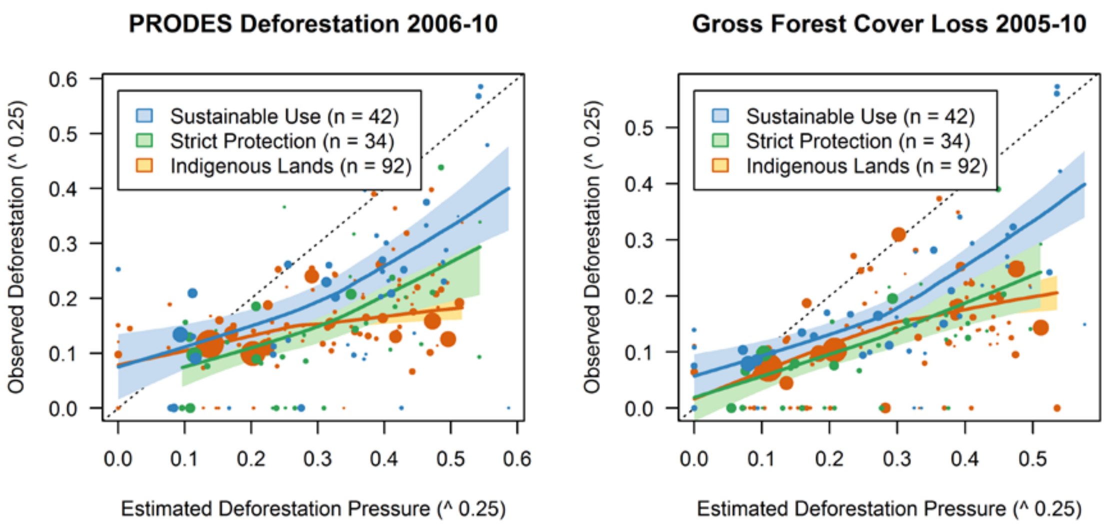

Figure B6: As Figure 5 (right) and Figure B3 (right), but excluding protected areas declared between 2000 and 2005 from the sample

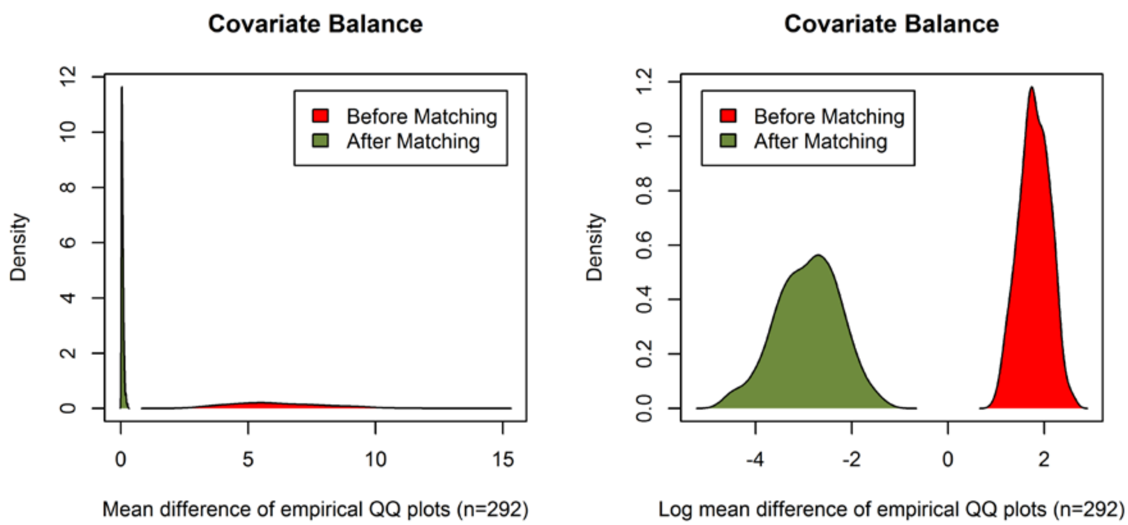

Figure B7: Density distributions of mean standardized differences of empirical quantile-quantile plots (raw and log), averaged across 30 repetitions and six continuous covariates for each of the 292 protected areas considered in our analysis.

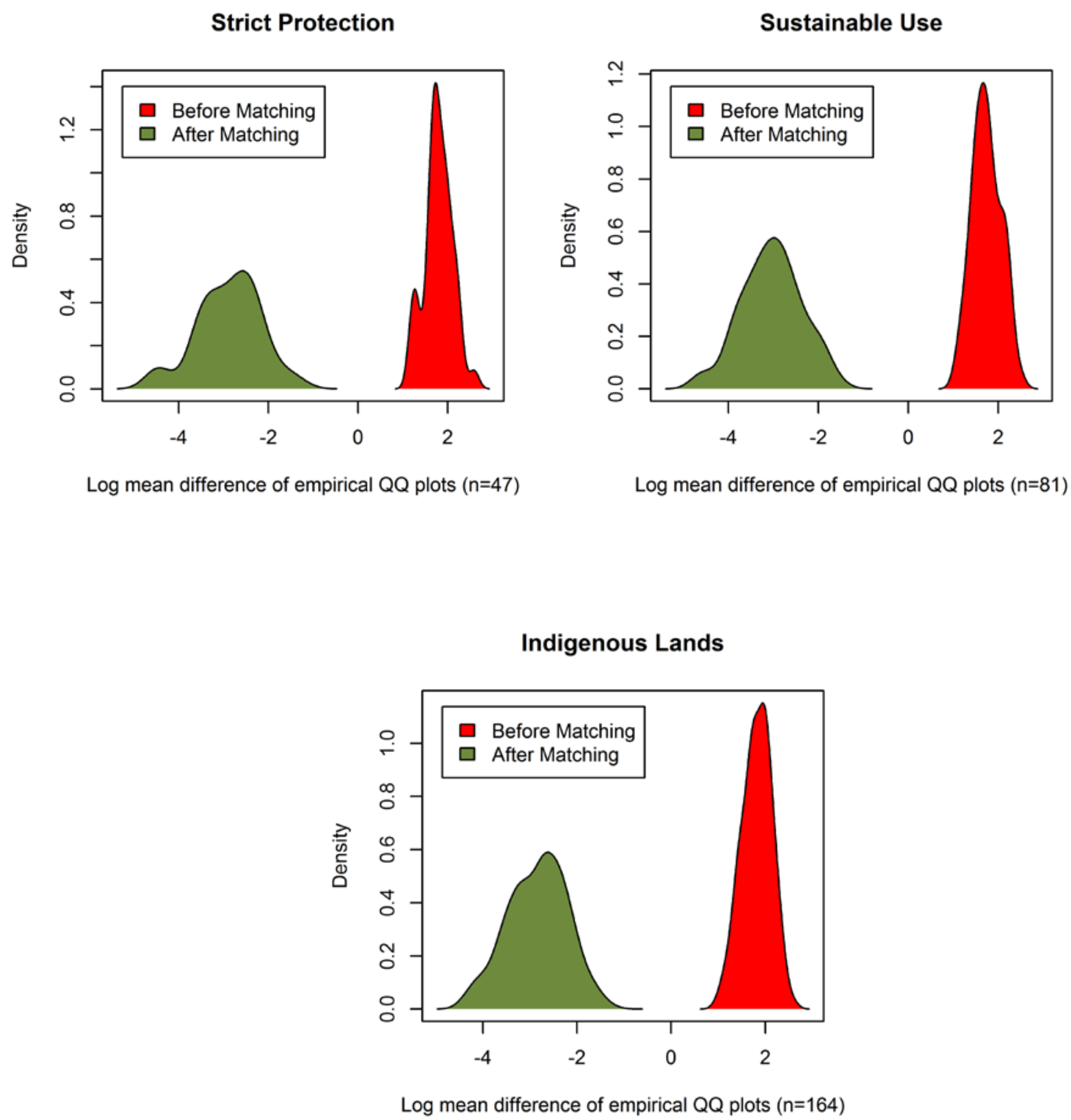

Figure B8: Density distributions of mean standardized differences of empirical quantile-quantile plots (log), averaged across 30 repetitions and six continuous covariates, by protection type.

REFERENCES

Abadie, A. and G.W. Imbens. 2006. Large Sample Properties of Matching Estimators for Average Treatment Effects. Econometrica **74**:235–267.

Adeney, J.M., N.L. Christensen, and S.L. Pimm. 2009. Reserves protect against deforestation fires in the Amazon. PloS one **4**:e5014.

Akella, A.S. and J.B. Cannon. 2004. Strengthening the Weakest Links: Strategies for Improving the Enforcement of Environmental Laws Globally. Conservation International, Washington DC.

Albers, H.J. and E.J.Z. Robinson. 2013. A review of the spatial economics of non-timber forest product extraction: Implications for policy. Ecological Economics **in press**:

Andam, K.S., P.J. Ferraro, A. Pfaff, G.A. Sanchez-Azofeifa, and J.A. Robalino. 2008. Measuring the effectiveness of protected area networks in reducing deforestation. Proceedings of the National Academy of Sciences **105**:16089–94.

Andam, K.S., P.J. Ferraro, K.R.E. Sims, A. Healy, and M.B. Holland. 2010. Protected areas reduced poverty in Costa Rica and Thailand. Proceedings of the National Academy of Sciences **107**:9996–10001.

Angelsen, A. 2008. Moving Ahead with REDD.Bogor, Indonesia

Angelsen, A. 2010. Policies for reduced deforestation and their impact on agricultural production. Proceedings of the National Academy of Sciences **107**:19639–44.

Angelsen, A., S. Brown, C. Loisel, L. Peskett, C. Streck, and D. Zarin. 2009. Reducing Emissions from Deforestation and Forest Degradation (REDD): An Options Assessment Report.Washington, USA

Aragão, L.E.O.C., Y. Malhi, N. Barbier, A.A. Lima, Y. Shimabukuro, L. Anderson, and S. Saatchi. 2008. Interactions between rainfall, deforestation and fires during recent years in the Brazilian Amazonia. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences **363**:1779–85.

Araújo, E. and P. Barreto. 2010. Formal threats to protected areas in the Amazon. The State of the Amazon **16**:1–10.

Arino, O., J.R. Perez, V. Kalogirou, P. Defourny, and F. Achard. 2009. Globcover 2009. European Space Agency, Paris, France.

ARPA. 2010. ARPA: Financial and operational management of ARPA. WWF Brasil, Brasília, Brazil.

Assunção, J., C.C. Gandour, and R. Rocha. 2012. Deforestation Slowdown in the Legal Amazon: Prices or Policies ? Climate Policy Initiative, Rio de Janeiro, Brazil.

Barber, C.P., M. a. Cochrane, C. Souza, and A. Veríssimo. 2012. Dynamic performance assessment of protected areas. Biological Conservation **149**:6–14.

Bartholome, E. and A.S. Belward. 2005. GLC2000: A New Approach to Global Land Cover Mapping from Earth Observation Data. International Journal of Remote Sensing **26**:1959–77.

Bertzky, B., C. Corrigan, J. Kemsey, S. Kenney, C. Ravilious, C. Besançon, and N. Burgess. 2012. Protected Planet Report 2012: Tracking progress towards global targets for protected areas. IUCN, Gland, Switzerland / UNEP-WCMC, Cambridge, UK.

Börner, J., S. Wunder, S. Wertz-Kanounnikoff, G. Hyman, and N. Nascimento. 2011. REDD sticks and carrots in the Brazilian Amazon: Assessing costs and livelihood implications. CCAFS Working Paper No. 8. CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS), Copenhagen, Denmark.

Brooks, T.M., R.A. Mittermeier, G.A.B. da Fonseca, J. Gerlach, M. Hoffmann, J.F. Lamoreux, C.G. Mittermeier, J.D. Pilgrim, and A.S.L. Rodrigues. 2006. Global biodiversity conservation priorities. Science **313**:58–61.

Bruner, A., R.E. Gullison, R.E. Rice, and G.A. da Fonseca. 2001. Effectiveness of parks in protecting tropical biodiversity. Science **291**:125–8.

Bruner, A.G., R.E. Gullison, and A. Balmford. 2004. Financial Costs and Shortfalls of Managing and Expanding Protected-Area Systems in Developing Countries. BioScience **54**:1119–1126.

Butchart, S.H.M. et al. 2010. Global biodiversity: indicators of recent declines. Science (New York, N.Y.) **328**:1164–8.

Câmara, G., D. de Morisson Valeriano, and J. Vianei Soares. 2006. Metodologia para o Cálculo da Taxa Anual de Desmatamento na Amazônia Legal (PRODES). Instituto Nacional de Pesquisas Espaciais, São José dos Campos, Brazil.

Carranza, T., A. Manica, V. Kapos, and A. Balmford. 2014. Mismatches between conservation outcomes and management evaluation in protected areas: A case study in the Brazilian Cerrado. Biological Conservation **173**:10–16.

Carroll, M., J. Townshend, M. Hansen, C. Dimiceli, R. Sohlberg, and K. Wurster. 2011. MODIS Vegetative Cover Conversion and Vegetation Continuous Fields. In B. Ramachandran, C. O. Justice, & M. J. Abrams, eds. Land Remote Sensing and Global Environmental Change. Remote Sensing and Digital Image Processing. Springer New York, New York, NY.

CBD. 2010. COP 10 Decision X/31. Convention on Biological Diversity, Montreal, Canada.

CBD. 2004. Programme of Work on Protected Areas (PoWPA). Convention on Biological Diversity, Montreal, Canada.

CBD and UNEP. 2010. Strategic Plan for Biodiversity 2011-2020. Secretariat of the Convention on Biological Diversity, Montreal, Quebec, Canada.

Chape, S., J. Harrison, M. Spalding, and I. Lysenko. 2005. Measuring the extent and effectiveness of protected areas as an indicator for meeting global biodiversity targets. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences **360**:443–55.

Chhatre, A. and A. Agrawal. 2008. Forest commons and local enforcement. Proceedings of the National Academy of Sciences **105**:13286–91.

Cialdini, R.B., L.J. Demaine, B.J. Sagarin, D.W. Barrett, K. Rhoads, and P.L. Winter. 2006. Managing social norms for persuasive impact. Social Influence **1**:3–15.

Coad, L. et al. 2013. Progress towards the CBD protected area management effectiveness targets. Parks1–125.

Coad, L., F. Leverington, C. Nolte, and M. Hockings. 2011. Management Effectiveness Tracking Tool: The Global Database. University of Oxford, Oxford, UK.

Cochrane, M.A. 2011. The past, present, and future importance of fire in tropical rainforests. In M. Bush, ed. Tropical Rainforest Responses to Climate Change. Springer, Berlin.

Collaborative Partnership on Forests. 2012. 2012 Study on Forest Financing. Advisory Group on Finance, Collaborative Partnership on Forests, United Nations Forum on Forests, New York, NY.

Cook, C.N., R.W.B. Carter, and M. Hockings. 2014. Measuring the accuracy of management effectiveness evaluations of protected areas. Journal of Environmental Management **139**:164–71.

Cook, C.N. and M. Hockings. 2011. Opportunities for improving the rigor of management effectiveness evaluations in protected areas. Conservation Letters **4**:372–382.

Corbera, E. and H. Schroeder. 2011. Governing and implementing REDD+. Environmental Science & Policy **14**:89–99.

Davalos, L.M., A.C. Bejarano, and H.L. Correa. 2009. Disabusing cocaine: pervasive myths and enduring realities of a globalised commodity. International Journal on Drug Policy **20**:381–6.

Dudley, N. 2008. Guidelines for applying protected area management categories. International Union for the Conservation of Nature, Gland, Switzerland.

Dudley, N., A. Belokurov, O. Borodin, L. Higgins-Zogib, M. Hockings, L. Lacerda, and S. Stolton. 2004. How effective are protected areas? A preliminary analysis of forest protected areas by WWF - the largest ever global assessment of protected area management effectiveness. WWF, Gland, Switzerland.

Dudley, N., A. Belokurov, L. Higgins-Zogib, M. Hockings, S. Stolton, and N. Burgess. 2007. Tracking progress in managing protected areas around the world. WWF, Gland, Switzerland.

Ervin, J. 2003a. Rapid Assessment of Protected Area Management Effectiveness in Four Countries. BioScience **53**:833–841.

Ervin, J. 2003b. WWF Rapid Assessment and Prioritization of Protected Area Management (RAPPAM) Methodology. WWF, Gland, Switzerland.

Ferraro, P. and S. Pattanayak. 2006. Money for nothing? A call for empirical evaluation of biodiversity conservation investments. PLoS biology **4**:e105.

Ferraro, P.J. 2008. Asymmetric information and contract design for payments for environmental services. Ecological Economics **65**:810–821.

Ferraro, P.J. 2009. Counterfactual Thinking and Impact Evaluation in Environmental Policy. In M. Birnbaum & P. Mickwitz, eds. Environmental program and policy evaluation: Adressing methodological challenges. New Directions for Evaluation. American Evaluation Association, Fairhaven, MA.

Ferraro, P.J., K. Lawlor, K.L. Mullan, and S.K. Pattanayak. 2011. Forest Figures: Ecosystem Services Valuation and Policy Evaluation in Developing Countries. Review of Environmental Economics and Policy **6**:20–44.

Ferraro, P.J. and M.K. Price. 2013. Using nonpecuniary strategies to influence behavior: evidence from a large-scale field experiment. The Review of Economics and Statistics **95**:64–73.

Fischer, G., F. Nachtergaele, S. Prieler, H.T. van Velthuizen, L. Verelst, and D. Wiberg. 2007. Global Agro-Ecological Zones Assessment for Agriculture. IIASA, Laxenburg, Austria.

Fuller, R.A., E. McDonald-Madden, K.A. Wilson, J. Carwardine, H.S. Grantham, J.E.M. Watson, C.J. Klein, D.C. Green, and H.P. Possingham. 2010. Replacing underperforming protected areas achieves better conservation outcomes. Nature **466**:365–7.

Gardner, T. a et al. 2008. The cost-effectiveness of biodiversity surveys in tropical forests. Ecology letters **11**:139–50.

Gaveau, D.L., J. Epting, O. Lyne, M. Linkie, I. Kumara, M. Kanninen, and N. Leader-Williams. 2009. Evaluating whether protected areas reduce tropical deforestation in Sumatra. Journal of Biogeography **36**:2165–2175.

GEF. 2008. GEF Benefits Index (GBI) for Biodiversity: Initial and Revised. Global Environment Facility, Washington DC.

GEF. 2009. GEF Focal Area: Biodiversity. Global Environment Facility, Washington, DC.

Geldmann, J., M. Barnes, L. Coad, I.D. Craigie, M. Hockings, and N.D. Burgess. 2013. Effectiveness of terrestrial protected areas in reducing habitat loss and population declines. Biological Conservation **161**:230–238.

Gibson, C., J. Williams, and E. Ostrom. 2005. Local Enforcement and Better Forests. World Development **33**:273–284.

Gordon and Betty Moore Foundation. 2006. Andes Amazon Initiative - Overview. Gordon and Betty Moore Foundation, Palo Alto, CA.

Groombridge, B. and M. Jenkins. 2002. World Atlas of Biodiversity. University of California Press, Berkeley, CA.

Hansen, M.C. et al. 2013. High-resolution global maps of 21st-century forest cover change. Science (New York, N.Y.) **342**:850–3.

Hansen, M.C., R.S. Defries, J.R. Townshend, M. Carroll, C. Dimiceli, and R.A. Sohlberg. 2003. Global Percent Tree Cover at a Spatial Resolution of 500 Meters: First Results of the Modis Vegetation Continuous Fields Algorithm. Earth Interactions **7**:1–15.

Hansen, M.C., Y. Shimabukuro, P. Potapov, and K. Pittman. 2008. Comparing annual MODIS and PRODES forest cover change data for advancing monitoring of Brazilian forest cover. Remote Sensing of Environment **112**:3784–3793.

Hansen, M.C., S. V. Stehman, and P. V. Potapov. 2010. Quantification of global gross forest cover loss. Proceedings of the National Academy of Sciences **107**:8650–8655.

Hayes, T. 2006. Parks, People, and Forest Protection: An Institutional Assessment of the Effectiveness of Protected Areas. World Development **34**:2064–2075.

Hijmans, R.J., S.E. Cameron, J.L. Parra, P.G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climated surfaces for global land areas. International Journal of Climatology **25**:1965–1978.

Hilborn, R., P. Arcese, M. Borner, J. Hando, G. Hopcraft, M. Loibooki, S. Mduma, and A.R.E. Sinclair. 2006. Effective enforcement in a conservation area. Science **314**:1266.

Hirsch, P.D., W.M. Adams, J.P. Brosius, A. Zia, N. Bariola, and J.L. Dammert. 2010. Acknowledging Conservation Trade-Offs and Embracing Complexity. Conservation Biology **25**:1–6.

Hockings, M., S. Stolton, N. Dudley, and R. James. 2009. Data Credibility: What Are the "Right" Data for Evaluating Management Effectiveness of Protected Areas? In M. Birnbaum & P. Mickwitz, eds. Environmental program and policy evaluation: Adressing methodological challenges. New Directions for Evaluation. Jossey-Bass, Hoboken, NJ.

Hockings, M., S. Stolton, F. Leverington, N. Dudley, J. Courrau, P. Valentine, and S. Editor. 2006. Evaluating Effectiveness: A framework for assessing management effectiveness of protected areas. 2nd edition. IUCN, Gland, Switzerland.

Holmes, G. 2007. Protection, Politics and Protest: Understanding Resistance to Conservation. Conservation and Society **5**:184–201.

Infield, M. and A. Namara. 2001. Community attitudes and behaviour towards conservation: an assessment of a community conservation programme around Lake Mburo National Park, Uganda. Oryx **35**:48–60.

Joppa, L. and A. Pfaff. 2010. Reassessing the forest impacts of protection: the challenge of nonrandom location and a corrective method. Annals of the New York Academy of Sciences **1185**:135–49.

Joppa, L.N. and A. Pfaff. 2011. Global protected area impacts. Proceedings of the Royal Society B **278**:1633–8.

Joppa, L.N. and A. Pfaff. 2009. High and far: biases in the location of protected areas. PloS one **4**:e8273.

Justice, C. 2002. The MODIS fire products. Remote Sensing of Environment **83**:244–262.

Kasparek, M., R. Mack, C. Mayer, and K. Riha. 2010. Biodiversity in German Development Cooperation 2010. Deutsche Gesellschaft für Technische Zusammenarbeit (GTZ) GmbH, Eschborn, Germany.

Keene, M. and A.S. Pullin. 2011. Realizing an effectiveness revolution in environmental management. Journal of Environmental Management **92**:2130–5.

Kinouchi, M. 2012. Efetividade de gestão e análise comparativa das aplicações do método RAPPAM nas unidades de conservação federais. In VII Congreso Brasileiro de Unidades de Conservação, Sep 24-27, 2012, Natal, Brazil.

Kis-Katos, K. and J.H. Gonçalves da Silva. 2010. Economic Causes of Deforestation in the Brazilian Amazon: A Panel Data Analysis for 2000s. In Beiträge zur Jahrestagung des Vereins für Socialpolitik 2010. Leibniz-Informationszentrum Wirtschaft, Kiel, Germany.

Klenk, N.L. and G.M. Hickey. 2011. A virtual and anonymous, deliberative and analytic participation process for planning and evaluation: The Concept Mapping Policy Delphi. International Journal of Forecasting **27**:152–165.

Kull, C. 2002. Madagascar aflame: landscape burning as peasant protest, resistance, or a resource management tool? Political Geography **21**:927–953.

Laurance, W.F. et al. 2012. Averting biodiversity collapse in tropical forest protected areas. Nature2–6.

Leverington, F., K.L. Costa, J. Courrau, H. Pavese, C. Nolte, M. Marr, L. Coad, N. Burgess, B. Bomhard, and M. Hockings. 2010. Management effectiveness evaluation in protected areas – a global study. Second edition 2010. University of Queensland, Brisbane, Australia.

Leverington, F., H. Pavese, and K.L. Costa. 2007. Management effectiveness evaluation in Latin America and the Caribbean. Report to the InterAmerican Biodiversity Information Network, Brisbane, Australia.

Liu, J., M. Linderman, Z. Ouyang, L. An, J. Yang, and H. Zhang. 2001. Ecological degradation in protected areas: the case of Wolong Nature Reserve for giant pandas. Science (New York, N.Y.) **292**:98–101.

Malamud, B.D., J.D.A. Millington, and G.L.W. Perry. 2005. Characterizing wildfire regimes in the United States. Proceedings of the National Academy of Sciences **102**:4694–9.

Margoluis, R., C. Stem, N. Salafsky, and M. Brown. 2009. Design Alternatives for Evaluating the Impact of Conservation Projects. In M. Birnbaum & P. Mickwitz, eds. Environmental program and policy evaluation: Addressing methodological challenges. New Directions for Evaluation.

Martin, A.S. and J.F. Rieger. 2003. The Parks in Peril Site Consolidation Scorecard. The Nature Conservancy, Arlington, VA, USA.

Mascia, M.B. and S. Pailler. 2011. Protected area downgrading, downsizing, and degazettement (PADDD) and its conservation implications. Conservation Letters **4**:9–20.

McCarthy, D.P. et al. 2012. Financial Costs of Meeting Global Biodiversity Conservation Targets: Current Spending and Unmet Needs. Science **338**:946–949.

Miller, D.C., A. Agrawal, and J.T. Roberts. 2013. Biodiversity, Governance, and the Allocation of International Aid for Conservation. Conservation Letters **6**:12–20.

Montibeller, G. and V. Belton. 2006. Causal maps and the evaluation of decision options—a review. Journal of the Operational Research Society **57**:779–791.

Mora, C. and P. Sale. 2011. Ongoing global biodiversity loss and the need to move beyond protected areas: a review of the technical and practical shortcomings of protected areas on land and sea. Marine Ecology Progress Series **434**:251–266.

Morgan, S.L. and C. Winship. 2007. Counterfactuals and causal inference: methods and principles for social research. Cambridge University Press, New York, NY.

Morton, D.C., R.S. Defries, J.T. Randerson, L. Giglio, W. Schroeder, and G.R. Van Der Werf. 2008. Agricultural intensification increases deforestation fire activity in Amazonia. Global Change Biology **14**:2262–2275.

Nauclér, T. and P.A. Enkvist. 2009. Pathways to a Low-Carbon Economy: Version 2 of the Global Greenhouse Gas Abatement Cost Curve. McKinsey & Company, New York, NY.

Nelson, A. 2008. Travel Time to Major Cities: A Global Map of Accessibility. Global Environment Monitoring Unit, Joint Research Centre of the European Commission, Ispra, Italy.

Nelson, A. and K.M. Chomitz. 2011. Effectiveness of Strict vs. Multiple Use Protected Areas in Reducing Tropical Forest Fires: A Global Analysis Using Matching Methods. PLoS ONE **6**:e22722.

Nelson, A. and K.M. Chomitz. 2009. Protected Area Effectiveness in Reducing Tropical Deforestation. The World Bank, Washington DC, USA.

Nepstad, D. et al. 2006. Inhibition of Amazon Deforestation and Fire by Parks and Indigenous Lands. Conservation Biology **20**:65–73.

Newton, P., W. Endo, and C. a. Peres. 2011. Determinants of livelihood strategy variation in two extractive reserves in Amazonian flooded and unflooded forests. Environmental Conservation **39**:97–110.

NIMA. 2000. Vector Map Level 0 (Digital Chart of The World), Edition 5. National Imagery and Mapping Agency, Reston, VA.

Nolte, C. and A. Agrawal. 2013. Linking Management Effectiveness Indicators to Observed Effects of Protected Areas on Fire Occurrence in the Amazon Rainforest. Conservation Biology **27**:155–65.

Nolte, C., A. Agrawal, and P. Barreto. 2013a. Setting priorities to reduce deforestation in Amazon protected areas: Are we choosing the right indicators? Environmental Research Letters **8**:1–7.

Nolte, C., A. Agrawal, K. Silvius, and B. Soares-Filho. 2013b. Governance regime and location influence avoided deforestation success of protected areas in the Brazilian Amazon. Proceedings of the National Academy of Sciences **110**:4956–4961.

Nolte, C., F. Leverington, A. Kettner, M. Marr, G. Nielsen, B. Bomhard, S. Stolton, S. Stoll-Kleemann, and M. Hockings. 2010. Protected Area Management Effectiveness Assessments in Europe. A review of application, methods and results. Bundesamt für Naturschutz (BfN), Bonn, Germany.

Oak Ridge National Laboratory Distributed Active Archive Center. 2011. MODIS subsetted land products. Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, TN.

Oates. 1999. Myth and reality in the rain forest: how conservation strategies are failing in West Africa. University of California Press, Berkeley, CA.

Olson, D.M. et al. 2001. Terrestrial Ecosystems of the World: A New Map of Life on Earth. BioScience **51**:933–8.

Parker, C., J. Brown, J. Pickering, E. Roynestad, N. Mardas, and A.W. Mitchell. 2009. The Little Climate Finance Book. Global Canopy Programme, Oxford, UK.

Parker, C., M. Cranford, N. Oakes, and M. Leggett. 2012. The Little Biodiversity Finance Book. 3rd edition. Global Canopy Programme, Oxford, UK.

Pedlowski, M., E.A.T. Matricardi, D. Skole, S.R. Cameron, W. Chomentowski, C. Fernandes, and A. Lisboa. 2005. Conservation Units: a new deforestation frontier in the Amazonian state of Rondônia, Brazil. Environmental Conservation **32**:149.

Pereira, H.M., L.M. Navarro, and I.S. Martins. 2012. Global Biodiversity Change: The Bad, the Good, and the Unknown. Annual Review of Environment and Resources **37**:25–50.

Petursson, J.G., P. Vedeld, and M. Sassen. 2012. An institutional analysis of deforestation processes in protected areas: The case of the transboundary Mt. Elgon, Uganda and Kenya. Forest Policy and Economics.

Pfaff, A., J. Robalino, E. Lima, C. Sandoval, and L.D. Herrera. 2013. Governance, location & avoided deforestation from protected areas: greater restrictions can have lower impact, due to differences in location. World Development **in press**:

Pivello, V.R. 2011. The Use of Fire in the Cerrado and Amazonian Rainforests of Brazil: Past and Present. Fire Ecology **7**:24–39.

Porter-Bolland, L., E. Ellis, M.R. Guariguata, I. Ruiz-Mallén, S. Negrete-Yankelevich, and V. Reyes-García. 2012. Community managed forests and forest protected areas: An assessment of their conservation effectiveness across the tropics. Forest Ecology and Management **268**:6–17.

Pueyo, S., P.M.L. de Alencastro Graça, R.I. Barbosa, R. Cots, E. Cardona, and P.M. Fearnside. 2010. Testing for criticality in ecosystem dynamics: the case of Amazonian rainforest and savanna fire. Ecology Letters **13**:793–802.

Robinson, E., A. Kumar, and H. Albers. 2010. Protecting Developing Countries' Forests: Enforcement in Theory and Practice. Journal of Natural Resources Policy Research **2**:25–38.

Saatchi, S.S., R.A. Houghton, R.C. Dos Santos Alvalá, J. V. Soares, and Y. Yu. 2007. Distribution of aboveground live biomass in the Amazon basin. Global Change Biology **13**:816–837.

Salafsky, N., R. Margoluis, K.H. Redford, and J.G. Robinson. 2002. Improving the Practice of Conservation: a Conceptual Framework and Research Agenda for Conservation Science. Conservation Biology **16**:1469–1479.

Scharlemann, J.P.W., V. Kapos, A. Campbell, I. Lysenko, N.D. Burgess, M.C. Hansen, H.K. Gibbs, B. Dickson, and L. Miles. 2010. Securing tropical forest carbon: the contribution of protected areas to REDD. Oryx **44**:352–357.

Schmitt, C.B. et al. 2009. Global analysis of the protection status of the world's forests. Biological Conservation **142**:2122–2130.

Schwartzman, S. and B. Zimmerman. 2005. Conservation alliances with indigenous peoples of the Amazon. Conservation Biology **19**:721–727.

Sekhon, J.. 2007. Multivariate and Propensity Score Matching Software with Automated Balance Optimization: The Matching Package for R. Journal of Statistical Software **42**:1–51.

Sheil, D. 2001. Conservation and Biodiversity Monitoring in the Tropics : Realities , Priorities , and Distractions. 1999–2002.

Sheil, D. 2002. Why doesn't biodiversity monitoring support conservation priorities in the tropics? Unasylva **53**:50–54.

Soares-Filho, B. et al. 2010. Role of Brazilian Amazon protected areas in climate change mitigation. Proceedings of the National Academy of Sciences **107**:10821–10826.

Soares-Filho, B.S., D.C. Nepstad, L.M. Curran, G.C. Cerqueira, R.A. Garcia, C.A. Ramos, E. Voll, A. McDonald, P. Lefebvre, and P. Schlesinger. 2006. Modelling conservation in the Amazon basin. Nature **440**:520–3.

Stern, M.J. 2008. Coercion, voluntary compliance and protest: the role of trust and legitimacy in combating local opposition to protected areas. Environmental Conservation **35**:200.

Stolton, S., M. Hockings, N. Dudley, K. MacKinnon, T. Whitten, and F. Leverington. 2007. Management Effectiveness Tracking Tool. Reporting Progress at Protected Area Sites. 2nd Edition. WWF, Gland, Switzerland.

Sunderland, T.C.H., C. Ehringhaus, and B.M. Campbell. 2008. Conservation and development in tropical forest landscapes: a time to face the trade-offs? Environmental Conservation **34**:276–279.

Sutinen, J.G. 1987. Enforcement of the MFCMA : An Economist's Perspective. Marine Fisheries Review **49**:36–43.

Terborgh, J.W. 2004. Requiem for Nature. Island Press, Washington DC.

Tollefson, J. 2012. Brazil set to cut forest protection. Nature **485**:19.

Trumper, K., M. Bertzky, B. Dickson, G. van der Heijden, M. Jenkins, and P. Manning. 2009. The Natural Fix? The Role of Ecosystems in Climate Mitigation. World Conservation Monitoring Center, United Nations Environment Programme, Cambridge, UK.

UNEP-WCMC. 2011. World Database of Protected Areas. United Nations Environmental Programme - World Conservation Monitoring Centre, Cambridge, UK.

Venter, O., W.F. Laurance, T. Iwamura, K.A. Wilson, R.A. Fuller, and H.P. Possingham. 2009. Harnessing carbon payments to protect biodiversity. Science **326**:1368.

Verissimo, A., A. Rolla, M. Vedoveto, and S. de M. Futada. 2011. Protected Areas in the Brazilian Amazon: Challenges and Opportunities. IMAZON, Belém, Brazil.

Vitel, C.S.M.N., P.M. Fearnside, and P.M.L. de Alencastro Graça. 2009. Análise da inibição do desmatamento pelas áreas protegidas na parte sudoeste do Arco de desmatamento. In J. C. N. Epiphanio & L. S. Galvão, eds. Anais XIV Simpósio Brasileiro de Sensoramento Remoto, Natal, Brazil 2009. Instituto Nacional de Pesquisas Espaciais, São José dos Campos, Brazil.

Vuohelainen, A.J., L. Coad, T.R. Marthews, Y. Malhi, and T.J. Killeen. 2012. The Effectiveness of Contrasting Protected Areas in Preventing Deforestation in Madre de Dios, Peru. Environmental Management **50**:645–63.

Waylen, K., A. Fischer, P.J.K. McGowan, S.J. Thirgood, and E.J. Milner-Gulland. 2010. Effect of Local Cultural Context on the Success of Community-Based Conservation Interventions. Conservation Biology **24**:1119–1129.

World Bank. 2013. Crossing Paths: Role of Protected Areas and Road Investments in Brazilian Amazon Deforestation. Economic Sector Report. World Bank, Washington DC, in press.

Zimsky, M., P. Ferraro, F. Mupemo, J. Robinson, and N. Sekhran. 2010. Results of the GEF Biodiversity Portfolio Monitoring and Learning Review Mission, Zambia. Enhancing Outcomes and Impact through Improved Understanding of Protected Area Management Effectiveness. Global Environment Facility, Washington DC.