Beyond Win-Wins: Understanding diverse impacts of complex protected area governance arrangements on human wellbeing and conservation in tropical forests

Robin Loveridge

PhD

University of York

Environment and Geography

June 2021

Thesis abstract

The role of protected areas has expanded from focusing on nature conservation, to also include human wellbeing and equitable governance objectives. Some contend that these objectives are mutually supporting, with 'win-win' outcomes possible for people and nature. However others consider these objectives to be competing, necessitating trade-offs. In response to the expanding remit of protected areas, complex governance arrangements have evolved. Yet understanding of how novel governance arrangements impact multiple objectives remains limited. In this PhD I aimed to evaluate how innovations in protected area governance have led to complex impacts on human wellbeing and conservation. The main contributions were first, to develop a generalised methodology for measuring human wellbeing in the context of development and conservation projects. The Wellbeing Indicator Selection Protocol provides a step-by-step guide to developing context-specific wellbeing indicators based on local perspectives and aligned with established wellbeing theory through the integration of quantitative and qualitative approaches. Second, I assessed the impact of community forests combined with forest certification in Tanzania as an example of recent innovations in protected area governance seeking to achieve win-wins. I found that certified community forests positively impacted wellbeing, conservation effectiveness and reduced gender inequality, though did not reduce elite capture. Finally, I undertook an in-depth analysis of the pathways by which certified community forests impact human wellbeing, governance effectiveness and forest restoration. I identified positive interacting pathways between certified community forests, equitable governance and wellbeing, though also trade-offs between conservation and agriculture. By integrating the contrasting perspectives of different actors and research methods, this thesis has contributed to advancing understanding of protected area governance and a shift away from the panacea of win-win solutions, towards a more in-depth understanding of what works where, for whom and why.

Acknowledgements

I owe a great debt of thanks to so many people who have mentored and guided me over the years. This thesis is the culmination of thirty plus years spent in awe of nature in all its intricacies and complexity and a burning desire to understand how people and planet can live in greater harmony. Thanks especially to my two main supervisors Andy Marshall and Susie Sallu, and supporting supervisor Marion Pfeifer for opening my mind with their insights and perspectives and indulged my questions and musings. Huge thanks to Johan Oldekop, Phil Platts, Steven Rushton, Julia Latham, Henry Travers and Emily Woodhouse whose generosity in sharing their technical specialisms and insights was invaluable in steering the development of the different manuscripts. Thanks to my research funder ESRC, University of York, Tanzania Forest Service and the Tanzania Commission of Science and Technology for the opportunity to undertake this research. Thanks to my fantastic research assistants Ignatus Joel Pesha, Mercy Mgaya, Petro Nnyti and Lilian Fredy, Abi Willis and all the staff at Mpingo, Conservation and Development Initiative and all interviewees who gave their time to participate in my studies and who made my time in Tanzania so unforgettable. Thanks to my academic peers and friends, in particular Will Mitchell and Jez Cusack with whom I've shared in equal measure the joys of intellectual lightbulb moments, adventures in the field and who've helped me navigate through the lows of academic frustrations with a sense of humour on this PhD journey. Thanks of course to my parents and sister and brother; especially my mother whose compassion has taught me to see things from different points of view and father, whose integrity and gritty determination I admired and will always remember with pride. An everlasting love and thanks to my wonderful partner in crime Kate, whose energy, encouragement and belief in me sustains and inspires me to be bold and think big. You are my home. Finally, thanks to my baby son Rhys, whose delight in life is a reminder of how wonderful the world is, and should be.

Table of Contents

Author's Declaration

I declare that this thesis is a presentation of original work and I am the sole author, except where contributors are noted below. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

Colleagues who contributed to the data chapters are included as co-authors on the respective chapters. However, I played the dominant role in all study design, data collection, data analysis, interpretation and writing. In accordance with the University of York submission guidelines, separate signed declarations have been provided by all co-authors and myself for the three data chapters submitted to academic journals as detailed below.

Statement of PhD candidate's contribution to manuscript: 'Measuring human wellbeing: a protocol for selecting local indicators'

We the authors of the paper titled '**Measuring human wellbeing: a protocol for selecting local indicators**' affirm that Robin Loveridge, The PhD candidate, provided the major contribution to this paper. Specifically, Robin led the conceptualisation, fieldwork, data processing and analysis and manuscript development. Ignatus supported fieldwork coordination, Susannah and Andrew co-supervised the research.

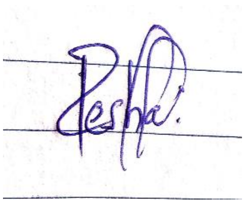
The manuscript has been published in Environmental Science & Policy:

Loveridge, R., Sallu, S.M., Pesha, I.J. and Marshall, A.R., (2020). Measuring human wellbeing: A protocol for selecting local indicators. *Environmental Science & Policy*, 114, 461-469.

Yours sincerely,

Robin Loveridge 

Susannah Sallu 

Ignatus Joel Pesha 

Andrew R Marshall 

Statement of PhD candidate's contribution to manuscript: 'Certified community forests positively impact human wellbeing and conservation effectiveness and improve the performance of nearby national protected areas'

We the authors of the paper titled '**Certified community forests positively impact human wellbeing and conservation effectiveness and improve the performance of nearby national protected areas**' affirm that Robin Loveridge, The PhD candidate, provided the major contribution to this paper. Specifically, Robin led the conceptualisation, fieldwork, data processing and analysis and manuscript development. Susannah and Andrew co-supervised the research, Marion and Daniel provided guidance on remote sensing analyses, Johan gave advice on statistical matching analyses, Phil gave advice on statistical analyses, Julia provided graphical design assistance and Mercy helped coordinate research and shared practical insights in the field.

The manuscript has been resubmitted to Conservation Letters after being accepted with minor revisions on the first submission.

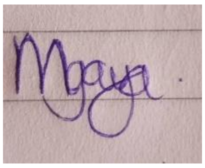
Yours sincerely,

Robin Loveridge

Susannah M Sallu

Marion Pfeifer

Johan A Oldekop

Mercy Mgaya 

Daniel A da Silva 

Julia Latham

Philip J Platts 

Andrew R Marshall

Statement of PhD candidate's contribution to manuscript: 'Pathways to win-wins or trade-offs? How certified community forests impact forest restoration and human wellbeing'

We the authors of the paper titled '**Pathways to win-wins or trade-offs? How certified community forests impact forest restoration and human wellbeing**' affirm that Robin Loveridge, The PhD candidate, provided the major contribution to this paper. Specifically, Robin led the conceptualisation, fieldwork, data processing and analysis and manuscript development. Andrew and Susannah co-supervised the research, Marion and Steven gave advice on structural equation modelling techniques and supported fieldwork and conceptual development of the research design.

The manuscript has been submitted to Philosphical Transactions B.

Yours sincerely,

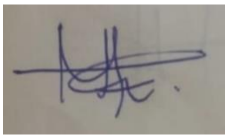
Robin Loveridge

Andrew R Marshall

Marion Pfeifer

Steven Rushton

Petro Nnyinti

Lilian Fredy 

Susannah M Sallu 

Chapter one: General Introduction

‘Sometimes we call those national protected areas by the name “Grandmother’s Fields” – *Shamba la bibi*. It can have two meanings:

First, by Grandmother we are talking about Queen Elizabeth because those National Forest Reserves and protected forests were introduced in the colonial time. So, Grandmother’s Fields can mean that those forests belong to Queen Elizabeth, so you must stay away. You cannot go there to do anything or take any benefit.

The second meaning is that the grandmother is your own grandmother and she wants to take care of you. You can go to her fields and do what you want, don’t worry, grandmother won’t mind.’

Village chairperson, Kilwa district – Alternative local narratives of protected areas as either a colonial fortress, or open access resource.

1.1 Thesis overview

Tropical forests are crucial for global biodiversity and provide a number benefits to society (Gibson et al., 2011). Benefits include both the material, or direct benefits of food and marketable products (MEA, 2005), as well as less tangible, indirect benefits such as supporting human health and sustaining cultural identities (Karjalainen, Sarjala and Raitio, 2010; Finer et al., 2008). Recognition of these linkages between people and nature has led to the emergence of a 'win-win' narrative, suggesting that positive forest conservation outcomes will also improve human wellbeing (Wells and Brandon, 1992).

Protected Areas (PAs) have increasingly become a focus of the win-win debate, with the role of PAs expanding from a primary focus on conserving nature, to also include improving human wellbeing (Watson et al., 2014; Naidoo et al., 2019). However, the performance of PAs to meet these twin goals is contested since to date PAs have demonstrated highly variable outcomes (Oldekop, et al., 2016; Gavin et al., 2018). Furthermore, PAs can have diverse impacts on different local actors. For example, PAs may limit resource access, impacting the ability of PA adjacent communities to cope with times of scarcity (Atela et al., 2015; Dressler et al., 2016). These unintended consequences often go unaccounted for in traditional impact evaluation designs, which have tended to prioritise assessment of a narrow suite of economic indicators (Woodhouse, et al., 2015). Partial accounting of PA impacts and a lack of understanding of how PAs affect different actors creates uncertainty regarding who the winners and losers of PA governance are. Evaluations of PA impacts suggest that win-wins have been rare, contentious and instead, trade-offs between people and nature and between different actors have been the norm (McShane et al., 2011).

Over the past 40 years, three dominant trends in forest PA governance have been (1) formalisation of forest tenure through establishment of strictly protected areas; (2) decentralisation of PA governance through the establishment of Community Forests (CFs) and other shared tenure systems and power sharing agreements, (3) the introduction of market-based approaches such as forest certification and carbon credits (Agrawal, Chhatre and Hardin, 2008). However, within these broad trends there is much differentiation. For example, forest certification projects may be superimposed on CFs to create novel governance forms (e.g. certified CFs), resulting in increasingly sophisticated and complex governance arrangements.

Innovations in PA governance share a common ambition to integrate the needs of people and planet to create sustainable landscapes (Sayer et al., 2013; Campbell, Sayer and Walker, 2018). Yet the success of such approaches has remained limited (Reed et al., 2020). As PA governance continues to grow in complexity, advances in PA impact evaluations are needed to keep pace with this real-world complexity in order to provide policy-relevant insights and help to improve the outcomes of PA governance for people and planet. However, research on PAs has lagged behind on-the-ground experimentation, with few rigorous impact evaluations of novel forms of PA governance (Burivalova et al., 2017; Hajjar, Engbring and Kornhauser, 2021). Therefore, to contribute to bridging this research-practitioner gap, I aim to evaluate how innovations in PA governance have led to complex impacts on human wellbeing and conservation effectiveness.

I have written the thesis with the three central data chapters (chapters two to five) in the style of scientific papers, each contributing to the overall thesis aim. While this format has resulted in some repetition of rationale and method between chapters, each chapter can be understood as a stand-alone contribution with the style of each chapter following the formatting guidelines of the journal.

In **chapter one** I provide an introduction to the research area and context for this thesis through a critical review of relevant literature on the impacts of PAs on human wellbeing and conservation effectiveness. In this review I identify common failings of PA governance approaches and the methods used to evaluate them in order to identify research gaps to inform the thesis aim, objectives, methodology and focus of subsequent chapters.

Chapter two is based on an identified knowledge gap concerning lack of clarity on a clear methodology for measuring human wellbeing. To contribute to this knowledge gap, I provide a novel approach for measuring a human wellbeing in the context of conservation and development projects. The paper is titled 'Measuring human wellbeing: a protocol for selecting local indicators' and has been published in Environmental Science & Policy (Loveridge et al., 2020).

In **chapter three** I provide an impact evaluation of certified CFs in Tanzania as an example of recent innovations in PA governance seeking to achieve win-wins. In this chapter I aim to contribute to advancing the ability of impact evaluations to adequately represent complex real-world situations by recognising that PAs do not exist in a vacuum independent of their surroundings, integrating complex landscape characteristics into the impact evaluation design. This chapter applies the human wellbeing assessment method from chapter two and is titled 'Certified community forests positively impact human wellbeing and conservation effectiveness and improve the performance of nearby national protected areas'. At the time of submitting this thesis, the manuscript of chapter three has been resubmitted to Conservation Letters following a favourable first review in which the manuscript was accepted with minor revisions. The outcome of the resubmission is now awaited.

Chapter four takes a deeper look at the underlying logic and assumptions underpinning recent trends in PA governance by assessing the pathways by which certified CFs impact wellbeing and conservation, specifically testing whether improvements in equitable governance or financial benefits are more important pathways. At the time of submitting this thesis, this chapter titled 'Pathways to win-wins or trade-offs? How certified community forests impact forest restoration and human wellbeing' has been submitted by invitation to a themed issue of Philosophical Transactions B that focuses on forest restoration in the UN decade of restoration. Reviewer feedback in awaited.

In the final **chapter five** I synthesise key findings, provide a summary discussion of the implications and limitations of the research for PA practice and recommend directions for future research.

1.2 Win-wins: A brief, barren marriage of people and nature?

The concept of win-wins for people and nature is premised on the core assumption that human wellbeing is strongly linked to the natural environment and therefore good environmental management will benefit people (MA, 2003; Tallis et al., 2008). Win-wins are intuitively appealing (Christensen, 2003), offering benefits for all and no drawbacks. This unilaterally positive rationale continues to attract support for integrated conservation and development projects (ICDPs), which explicitly aim to benefit local people and conservation simultaneously (Wells and Brandon, 1992). However, win-wins are a highly contentious concept and the uniting of social and environmental objectives has been referred to as a 'brief, barren marriage' (Redford and Saunders, 1992). An analysis of World Bank projects seeking to deliver positive social and environmental outcomes found that only 16 % achieved win-wins (Tallis, Kareiva & Marvier, 2008).

In the following section I critically review key barriers to our understanding and realisation of win-wins in tropical forest protected area (PA) governance. First, I explore conceptual ambiguities that hamper a clear understanding of if a win-win has been achieved. Second, I evaluate methodological approaches used to causally attribute the impact of PAs identifying strengths and weaknesses of different approaches to correctly attributing whether win-wins have been achieved. Third, in the light of conceptual and methodological limitations of existing studies, I evaluate the literature on PA impacts to see whether win-wins have been achieved in practice, paying particular attention to understanding how complex social outcomes have been characterised. Finally, I summarise common failings and weaknesses of existing governance approaches and evaluations in order to identify research gaps and new frontiers for research on the social and environmental impacts of PA governance.

1.3 Conceptual ambiguity and the gap between policy and practice

Win-wins are defined as joint positive outcomes for both people and the environment (Bennett et al., 2015). Trade-offs occur where either a positive social outcome is associated with a negative environmental outcome, or vice versa (Howe et al., 2014; Carr et al., 2021). Coarse scale global analyses indicate that environmental declines negatively affect quality of life, suggesting that conserving nature could be a win-win (Brauman et al., 2020). However finer scale analyses identify negative associations for people living adjacent to biodiverse areas including elevated levels of crop damage from wild animals and poorer access to health and education infrastructure (Naughton-Treves, 1997; Galvin, Beeton and Luizza, 2018). Local communities therefore often find themselves at odds with conservation. Agriculture is the main livelihood of rural communities in developing countries, however agricultural expansion is the primary driver of forest loss in the tropics (Gibbs et al., 2010).

Beyond a simple dichotomous definition of a win-win, there is also a need to consider who the winners and losers are. For example, poorer members of communities are often the most dependent on forest resources and so conservation efforts to limit access may disproportionately impact these already vulnerable people and exacerbate pre-existing inequalities (Kumar, 2002). Therefore, there is a need to look more deeply into proposed win-wins by disaggregating communities to understand how benefits and costs are distributed between actors in order to account for unequal impacts (Daw et al., 2011).

Protected areas (PAs) have long been the centre-piece of efforts to conserve nature. However increasingly attention is also being paid to how PAs impact human wellbeing (Agarwala et al., 2014; Naidoo et al., 2019). The Durban Accord (World Parks Congress, 2003) asserts that PAs should as a minimum 'do no harm' to local communities. More specific guidelines have since been endorsed in International Agreements including the Convention on Biological Diversity (2010), which states that 'PAs should be established and managed in close collaboration with and through equitable processes that recognize and respect the rights of indigenous and local communities and vulnerable populations and such costs and benefits of the areas are fairly shared'. Further underscored by the New York Declaration on Forests, Goal 10 to 'strengthen forest governance and transparency'. Equitable governance concerns notions of fairness (McDermott, 2013; Friedman et al. 2018), in relation to three dimensions of governance; (1) Distribution - the fair distribution of benefits and costs; (2) Recognition – respect for the rights and values of different actors (Schreckenberg, Franks and Martin, 2016), (3) Procedure - the fair participation of actors in decision making, relating to transparency of information, accountability of managers and equitable dispute resolution (Borrini-feyerabend, Dudley and Jaeger, 2013).

To understand if a win-win has been achieved, we must first understand what constitutes a win. Positive forest conservation outcomes tend to be evidenced through indicators of change in either extent or quality of forest habitat (FAO, 2020; Parkes, Newell and Cheal, 2003; Pfaff et al., 2014). Change in extent is often measured through remote sensing analyses of forest loss data layers (e.g. Hansen et al., 2013). Positive forest quality change (restoration), and negative change (degradation) in large-scale analyses are commonly assessed by change in remote sensing data layers of vegetation indices (e.g. Pfeifer et al., 2016, 2012). In small-scale studies, field measurements of forest structure may be undertaken (e.g. Kalonga, Midtgaard and Eid, 2015). Alternative ecological indicators have been proposed that build on the concept of ecosystem health, composed of three interrelated components: (1) Vigour – the activity, or primary productivity of an ecosystem; (2) Organisation - the diversity and number of interactions between system components; (3) Resilience - a system's capacity to maintain structure and function in the presence of stress or rebound from disturbance (Rapport, Costanza and McMichael, 1998). However the stochasticity and potentially long time periods between disturbance events under natural experimental conditions makes the concept an impractical indicator of conservation outcomes for short to medium term interventions.

In terms of social indicators, improving human wellbeing has become a focal concept for measuring social progress and has been enthusiastically adopted as a goal of international environmental and sustainable development policy (UNDP, 2015; CBD, 2016). The concept has also become a focus of global conservation NGOs, exemplified by the mission statement of Conservation International aim of 'sustainable care for nature and human wellbeing'. Despite its popularity at the policy level, there remains ongoing debate about how wellbeing should be conceptualised and measured (Dasgupta, 2001, p. 24; OECD; 2013).

The last 40 years have seen a shift in development discourse from a narrow focus on economic development, to more broadly include health, education, gender equity and what is needed to participate in society (Haq, 1996; Alkire, 2002; UN, 2015). It is now widely accepted that measuring the process of development through economic proxy indicators alone is inadequate, since an increase in income does not capture the multiple interactions and interdependence of the different aspects of people's lives (Gough et al., 2007; Daw et al., 2011). For example, a basic level of education is required to support literacy, which affects employment opportunities and the ability to read a newspaper and engage in political activities. Equally, education can be a stronger predictor of health than income. Therefore purely economic indicators fail to adequately represent non-economic aspects of peoples' lives (Haq, 1996). In the 1980s and 1990s the ethical and empirical strength of these arguments achieved a shift in development discourse from a focus on economic development to 'human centred development' (Max-Neef, 1989; Haq, 1996). This human orientated conception of development was defined by Sen (1999) as 'a process of expanding the real freedoms people enjoy'. The concept of human wellbeing encompasses this broader notion of multidimensional development, building on an understanding of what people need to participate and flourish in society (Max-Neef, 1989; Alkire 2002). Both human wellbeing and human development share the same theoretical root in the capabilities approach of Sen, with human wellbeing adopted as a holistic focal concept for assessing progress in human development (1999).

Various definitions of wellbeing exist, though none are unanimously accepted (Brown and Westaway, 2011). A widely adopted definition developed by the Wellbeing in Developing Countries research group defines wellbeing as 'a state of being with others, which arises where human needs are met, where one can act meaningfully to pursue one's goals and where one can enjoy a satisfactory quality of life' (Gough et al., 2007, p. 46). A detailed historical review of the human wellbeing concept is provided in chapter 2, noting that while there is a growing consensus concerning universally relevant, broad domains of wellbeing, there remains ambiguity concerning how the concept should be measured (chapter 2).

The ambiguity concerning how to measure human wellbeing has meant that evaluations of the impact of PAs have evidenced wellbeing in various ways, yet a full accounting of social impacts is often lacking (de Lange et al., 2016). Assessments have tended to prioritise a narrow conception of human wellbeing, focussing on economic impacts through externally defined indicators that are readily observable (de Lange et al., 2016). This risks discounting local priorities and non-economic impacts (Woodhouse, et al., 2016). For example how limiting forest access may impact the ability of communities to cope in times of scarcity (Atela et al., 2015; Dressler et al., 2016). Evaluations that assess only a simplistic conception of wellbeing may characterise PAs as a social win if they result in financial benefits, when significant detrimental and unintended non-economic impacts may be missed. To more comprehensively evaluate the impact of PAs, there is a need to consider broader, locally relevant measures of human wellbeing (Woodhouse et al., 2016; Sterling et al., 2017). Lack of clarity on how to measure wellbeing represents a significant barrier to the realisation of genuine win-wins in PA governance because inconsistencies in measurement and narrow framings of social impacts hampers identification and adaptive management to overcome diverse and unintended negative PA impacts.

1.4 Methodological challenges of attributing impact

To estimate the impact of an intervention, a counterfactual approach is often used; asking what would have happened in the absence of the intervention (Schleicher et al., 2020). By this approach the difference in the response variable between a treatment and control group is used to assign impact to the intervention. Randomised control trials (RCTs) are often considered the 'gold standard' of impact evaluation (Gertler et al., 2016), where it is possible to replicate experimental conditions by giving equal probability for sample units to be assigned to either the treatment or control groups. RCTs are common in the medical studies, where a random sample of the population might be given a new treatment and a randomly assigned control group given a placebo. However PAs are established non-randomly in space for strategic reasons, such as biodiversity values and remote locations with limited alternative land-use or economic potential (Joppa and Pfaff, 2010). Consequently PAs are systematically different from the wider landscape in terms of factors, other than governance practices that are likely to influence wellbeing and conservation effectiveness (Ferraro, Hanauer and Sims, 2011). For example, Joppa et al., (2008) illustrate that forest protected areas in the Amazon and Congo show low rates of forest cover loss. However, this is due primarily to their remoteness rather than differences in how the forests are governed, so that these PAs are *de facto* protected by their relative inaccessibility and lower anthropogenic pressure rather than their protected status (Joppa, et al., 2008).

To isolate the impact of PAs from other confounding influences, quasi-experimental designs are employed that seek to mimic randomisation under non-experimental conditions (de Lange, Woodhouse and Milner-Gulland, 2016). Increasingly popular in conservation science is the use of statistical matching to identify treatment and control units with balanced distributions of confounding variables (Schleicher et al., 2020). For example, Andam et al., (2008) compared deforestation rates in Costa Rica's government managed PAs with control sites matched on criteria likely to influence land clearance; agricultural productivity, distance to forest edge, distance to roads and distance to the nearest city. A key assumption of matching studies is that by controlling for confounding variables, treatment assignment in sampled units is, as if random (Oldekop et al., 2019). Confounding variables are commonly referred to in the impact evaluation literature as exogenous variables – affecting both the treatment and response (Engle et al., 1983). By controlling for exogenous variation, any difference between treatment and control units is assumed to be caused by the treatment.

Impact evaluation designs employing matching use a subtly, but profoundly different set of assumptions to attributing causation compared with other modelling approaches in the natural sciences. Standard ecological modelling approaches employ reverse causality; seeking to explain the maximum amount of variation in a given response variable y, from a number of explanatory variables. Impact evaluation frameworks ask forward causality questions, seeking to isolate the effect of a single explanatory variable x (e.g. the PA intervention) on the response variable y (e.g. human wellbeing). By reducing any bias introduced by differences in confounding explanatory variables between treatment and control groups enhances the causal inference of an experiment to attribute differences to the intervention (Stuart, 2010). Research designs employing forward causality have been suggested to be better suited to hypothesis testing of particular causal drivers (e.g. forest protection) as they generate more precise estimates of hypothesised effects (Oldekop et al., 2020). The ability of such studies to discount alternative explanations and attribute social and ecological outcomes to the intervention has led to calls for their widespread adoption to improve the rigour of impact evaluations (Baylis et al., 2016). However their use remains limited (Burivalova et al., 2017).

Impact estimation using statistical matching depends on the availability of appropriate control units (Green, et al., 2013), with model extrapolation less accurate outside the region of shared overlap of confounding variables between treatment and control groups (Stuart, 2010). Assessing the appropriateness of potential controls depends on the availability of appropriate datasets and knowledge to account for all relevant confounding factors (Schleicher et al., 2020). In complex, real-world scenarios, PAs do not exist in a vacuum (Baylis et al., 2016). In reality, PAs are influenced by multiple other development and conservation interventions, for which there may be only patchy information available

(Schleicher et al., 2020). PAs also cause impacts beyond their own boundaries. For example the displacement of deforestation to areas adjacent to PAs, commonly referred to as "leakage" or "spillover effects" (Pfaff and Robalino, 2017). To date evidence of spillover effects is mixed. Spillover effects may have negative effects on overall conservation effectiveness by increasing forest loss in adjacent areas (Ewers and Rodrigues, 2008; Pfeifer et al., 2012). Alternatively, other studies have identified a 'halo effect' in which PA adjacent areas are better protected than expected (Gaveau et al., 2012). Conversely, Andam et al. (2008) found no evidence of leakage from PAs in Costa Rica. Furthermore as new PAs are established there is increasing connectivity between PAs under various forms of governance, leading to a range of PA network configurations and a need to understand how PAs interact with each other (Barton, Blumentrath and Rusch, 2013; Thornton, Branch and Murray, 2020). To more comprehensively account for the net impact of PAs, studies should therefore consider the interactions between PAs and their surroundings (DeFries, Karanth and Pareeth, 2010; Baylis et al., 2016).

The reliance of quantitative impact evaluations on *a priori* hypotheses concerning key drivers and secondary data sources with which to test them is a potential research design weakness. Quantitative research designs tend to be deductive in the sense that they test a pre-determined hypothesis that there is a difference between two sample populations in the chosen response caused by a limited number of pre-determined explanatory variables. A criticism of such deductive research designs is that they suggest a simplistic, mechanistic relationship between cause and effect (Hall, 2003) e.g. assuming the same stimulus will consistently lead to the same response, irrespective other contextual factors. Such designs have their origins in the natural sciences. However, this deductive approach has some problems of application in the social sciences. One reason is because intentional theories of human behaviour suggest that human responses are conditional, adjusted based on the social conditions (Poteete et al., 2010). This suggests that a complex range of responses are possible depending on the specific context. When responses are conditional on a wide range of drivers, deductive research designs identify the normative, average response to a treatment and so are useful in identifying relationships between two variables. But they have been suggested to provide little inference concerning causality of why the relationship exists as they lack contextual information (George and Bennett, 2005). Therefore, in complex real-world situations, purely deductive designs are limited in their ability to identify and attribute causal relationships amongst multiple potential drivers (Gerring, 2007).

Top-down framings have been suggested to create a 'bottleneck' in dialogue (Evans, Murphy and Jong, 2014), which may miss locally relevant and unanticipated explanations for PA impacts (Rode, Gómez-Baggethun and Krause, 2014; Turnhout, 2014). The focus of conventional impact analyses on estimating the size of impacts provides little causal explanation of exactly how PAs affect social and environmental outcomes (Ferraro and Hanauer, 2015); what is the means by which the observed impacts are achieved? These more in-depth insights are needed to help PA managers refine existing approaches to improve performance (Persha, Agrawal and Chhatre, 2011).

An alternative to quantitative impact evaluations is to employ qualitative approaches. This includes participatory approaches and case studies. Participatory approaches focus on identifying the perceptions of different actors in PA governance and often take a more exploratory approach; asking open-ended questions, rather than using pre-determined hypotheses (Franks, Booker and Roe, 2018; Schreckenberg et al., 2010a). By its exploratory nature, qualitative methods can be used to identify new causal relationships and develop theory (Mahoney and Goertz, 2006). However qualitative approaches are unable to generate quantitative estimates that are comparable between sites, which are needed to test new theories (Oldekop et al., 2020). In contrast, quantitative approaches are well suited to the empirical testing of theory (Poteete et al., 2010).

Case studies provide a more in-depth analysis of a study system than quantitative approaches by focussing on identifying heterogeneous outcomes in order to evaluate the range of contextual factors that determine impacts (Woodhouse et al., 2016; Collier et al., 2004). By focussing on variation and identifying anomalies and outlying cases, qualitative case study approaches support critical analysis of the limits of general patterns and simplistic explanations (George and Bennett 2005). This focus on causal heterogeneity can highlight contrasting explanations to quantitative assessments, which identify the normative response by estimate the average effect size (e.g. PA impact) of an intervention (Gerring, 2007). For example, exploring the underlying reasons why a PA has positively impacted one village, but not another can provide more nuanced insights into how PAs affect social outcomes. In this way, qualitative methods are well suited to exploring complex phenomenon by using a range of contextual factors to unpack the relationships between multiple potential drivers and conditional responses (Poteete et al., 2010).

A final methodological approach for attributing PA impacts is to use a conceptual model, also termed a theory-of-change, in which the hypothesised pathways to social and ecological impacts are represented in a graphical model, specifying the assumptions (Romero and Putz, 2018). This approach can help to clarify the means by which PAs impact social and ecological outcomes by explicitly articulating causal pathways and identifying what evidence is needed to assess short and long-term changes along these pathways (de Lange, et al., 2016; Ferraro and Hanauer, 2015). For example, a payment for ecosystem services project might create a theory of change on the assumption that received financial payments, lead to an attitude change towards conservation and improved conservation outcomes, with associated data collected for each step. This approach depends on logical assumptions rather than comparison with a control group. The reduced survey effort required by focussing only on intervention sites without control sites has made this approach popular with applied conservation organisations, where delivering rather than evidencing impact may be prioritised (Kapos et al., 2009; Fauna & Flora International, 2018). However the absence of a comparison group makes this approach weak in terms of estimating impact effect sizes (Woodhouse, De Lange and Milner-Gulland, 2016). The theory of change approach has also been criticised for being overly linear in the way that causal pathways are characterised, failing to adequately represent complex systems (Wilkinson, Hills and Barbrook-johnson, 2021). Furthermore, the relationships between outcomes are rarely tested through quantitative means (Ferraro and Hanauer, 2015; Adedokun, Childress and Burgess, 2011). Instead theories of change are generally used to provide a logical framework to identify complementary sources of information to support a hypothesised pathway of change (Gertler et al., 2016).

The methodological approaches reviewed here have their origins in distinct scientific disciplines spanning the quantiative social sciences of medicine and economics, to more qualitative, participatory social sciences and the natural sciences. Each approach is aligned with a particular perspective and prioritises different forms of scientiifc validity. The more quantiative approaches tend to prioritise external validty – the extent to which findings can be generalised to different contexts (Campbell and Stanley, 1963). By being able to extrapolate findings to other contexts and make generalisable recommendations, quantitative approaches lend themselves to policy engagement. However top-down research designs risk marginalising the importance of local values, context-specific factors and may lack local relevance (Woodhouse et al., 2015).

Qualitative research methods tend to provide more contextual information (Collier, Bradley and Seawright, 2004). By their exploratory nature, qualitative methods can be used to identify new causal relationships and develop theory (Mahoney and Goertz, 2006).

Qualitative research methods can therefore be seen to enhance internal validity; the ability of studies to attribute findings to particular drivers of change (Yue, 2012).

Qualitative studies tend to be designed from a bottom-up perspective, building on the opinions and views of study participants. In this way they support ecological validity, defined by Yue (2012) in relation to case study research as the extent to which the researchers' findings reflect the lived experience of those whom the researchers are studying. Ecological validity ensures that local relevance is retained, promoting rather than marginalising the priorities of study participants (Howard et al., 2016). However a weakness of qualitative approaches is that they may be limited in their ability to generalise findings to other contexts because: (1) designs tend to select particular study sites for specific reasons which may not be representative of the wider population; (2) the focus on collecting in-depth case-specific information makes it harder to generalise findings to other contexts, thereby limiting external validity.

The integration of top-down and bottom-up research perspectives provides a potential solution to the weaknesses of each approach to attribute PA impacts. For example, controlling for complex confounding influences in statistical matching may require the integration of qualitative and place-based knowledge to inform the selection of appropriate controls where only incomplete secondary data is available (e.g. Mitchell et al., 2018). Furthermore, a theory of change approach can be complementary to statistical matching, since a logical understanding of the study system is required to identify the direction of the causal pathway. For example the establishment of a PA might lead to an increase in tourism revenue and improved wellbeing of local communities (Braber, et al., 2018). However causation could flow in the opposite direction as an increase in tourism revenue might result from improved wellbeing of local communities allowing them to invest in tourism infrastructure (Schleicher et al., 2020). Therefore, a clearly defined theory of change is needed to attribute the direction of causation. To continue this example, an increase in tourism revenue could result from confounding factors such as improved accessibility to the PA resulting from the construction of new road networks. To exclude this alternative explanation, matching approaches that control for confounding variables are needed. Top-down and bottom-up perspectives have opposing strengths and weaknesses in terms of attributing causation. The interweaving of contrasting methodological approaches through application of mixed-methods approaches may add additional methodological complexity, but presents opportunities for developing novel insights and more robust assessments.

1.5 Trends in PA governance and their impacts on human wellbeing and conservation effectiveness

Since the beginning of the colonial era the dominant pattern has been for tropical forests to be owned and governed by central governments (Barrow et al., 2012). Over the past 40 years there have been three distinct trends in forest governance with complexity of governance arrangements increasing over time (Figure 1). First, forest tenure arrangements were increasingly formalised through the establishment of strictly protected areas (IUCN category I protected areas). These are managed exclusively by central governments for biodiversity conservation (Agrawal, Chhatre and Hardin, 2008). However in many developing countries central agencies lack the capacity to effectively enforce restrictions on use, resulting in nominal 'paper parks', *de facto* open access conditions and overexploitation (Nasi et al., 2008). Since the 1980s, a second trend has been towards decentralisation of governance responsibility (Arnold, 2001; World Bank, 2008). A common PA governance approach that exemplified this trend is Community Forests (CFs). Following the IUCN protected area categorisation we refer to CFs as a particular form of category VI protected areas, where sustainable use of natural resources is permitted (Dudley, 2008). As with other category VI protected areas, sustainable use is promoted as a *means* to achieve nature conservation (Dudley, 2008). What distinguished CFs from other forms of category VI protected areas is that communities are officially recognised by the national government as having collective governance responsibility and decision-making power over the area of land. CFs are distinct from production forests, as CFs have nature conservation as a primary objective. They are also distinct from private protected areas as the land is held under a collective title, rather than by a private entity.

CFs are created with the intention of simultaneously improving forest conservation and the wellbeing of local communities by permitting forest access and vesting rights and responsibilities to local actors as a means to incentivise sustainable use (Persha, Agrawal and Chhatre, 2011). Governmental and international donor support has led to a rapid rise in designation of CFs and communities now have formal land tenure rights covering 15% of the world's forests (Rights and Resources Initiative, 2019). A third trend has been the rise in market-based approaches such as forest certification, exemplified by Forest Stewardship Council (FSC) certification and Payment for Ecosystem Services (PES), exemplified by reducing emissions from deforestation and forest degradation (REDD+) projects (Agrawal, Chhatre and Hardin, 2008). This approach seeks to capitalise on increasing recognition of the monetary value of ecosystem services provided by forests to generate additional revenue flows to sustainably manage forests and compensate communities for the opportunity costs of avoided use (Miles and Kapos 2008). The introduction of market-based approaches adds further governance complexity as these governance arrangements may be superimposed on existing PAs (Hajjar, Engbring and Kornhauser, 2021). For example, forest certification might be introduced in CFs requiring the integration of international standards with national legislation in certified CFs (Figure 1.1). The following section reviews the impact of these major trends in forest governance on the wellbeing of adjacent communities and conservation effectiveness.

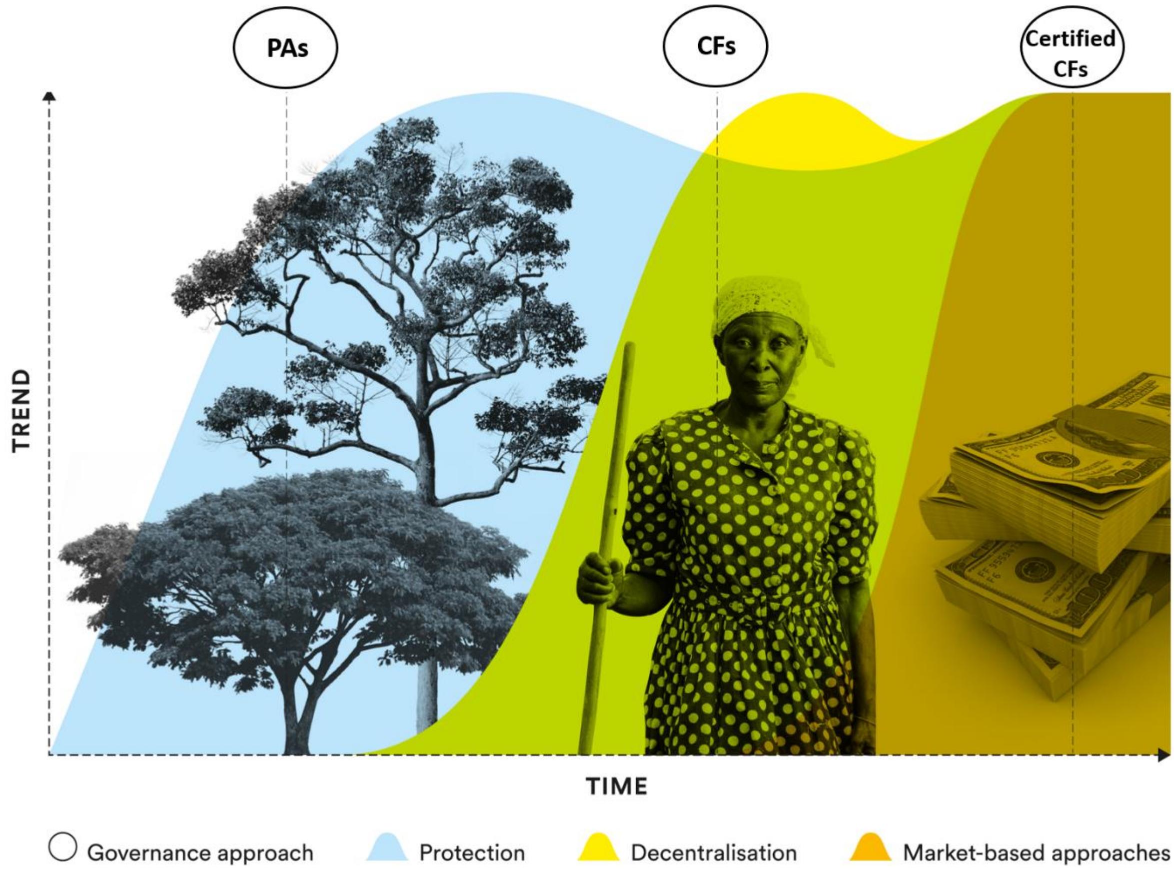


Figure 1.1 Major trends in forest protected area governance through time showing the emergence of community forests (CFs) and certified CFs alongside strictly protected areas

(PAs). Trend lines and colour layering show how new trends have added to the governance arrangements of earlier trends creating additional governance complexity. Figure produced for publication of chapter three.

Global scale analyses show that PAs governed by either centrally governments or locally communities generally do slow the rate of natural land cover loss relative to unprotected sites under similar development pressure (Joppa and Pfaff, 2011). However finer scale analyses of PA networks within particular countries or assessments of particular PAs illustrate a more mixed story of conservation effectiveness. Andam et al. (2008) show that Costa-Rica's protected area network overseen by the central government reduced the rate of tropical forest loss relative to control sites. An assessment of the effectiveness of CFs in Madagascar at halting deforestation found that CFs did not reduce deforestation rates relative to control sites with no formal forest tenure system in place (Rasolofoson et al., 2015). In contrast to these findings, Bray et al., (2008) compared deforestation rates between CFs and strictly protected areas in Mexico and found that CFs showed lower deforestation rates. A pan tropical study by Porter-Bolland et al., (2012) found the same pattern. Taken together, these studies suggest that both centrally governed PAs and CF strategies can be effective at achieving positive conservation outcomes, although effectiveness varies between countries. Coarse distinctions between national versus locally governed PAs is not a reliable predictor of positive or negative impacts (Persha, Agrawal and Chhatre, 2011; Pullin et al., 2013).

The picture of how PA governance influences wellbeing outcomes is complicated by ongoing uncertainty as to how wellbeing should be measured and the different metrics used by contrasting studies (Pullin et al., 2013). However global analyses show that PAs that supported local empowerment were more likely to have positive socioeconomic outcomes (Oldekop et al., 2016). Centrally governed, strictly protected areas have been criticised for their negative impacts on local communities due to factors highlighted by more qualitative studies such as forced evictions and conflicts with communities (West, Igoe and Brockington, 2006), disruption to agricultural and other local livelihoods (Naughton-Treves, 1997) and eroding cultural heritage of indigenous groups (Mahanty, 2018). However, a study of strictly protected areas in Cambodia showed that where targeted community development interventions are also in place, positive socio-economic outcomes can also occur (Beauchamp et al., 2018). Similarly a comparable study of a national park in Tanzania showed that communities living in closer proximity to the national park had access to better education and water provision facilities than communities further from the park with benefits accruing due to tourism revenue (Baird, 2014).

Negative social outcomes have also been observed in CFs, as conflicts can be created over resource access within and between villages (Gross-camp et al., 2019) and increasing inequality between local elites and non-elites, referred to as 'elite capture' (Coad et al., 2008; Roe et al., 2009). A national scale study in Tanzania using census data representative of multiple dimensions of human wellbeing showed a spatial link between improvements in household food security in areas surrounding CFs compared with non-CF areas (Pailler et al. 2015). But household wealth and health outcomes were not significantly different. Wealthy households benefited disproportionately from CFs than poor households and CF benefits increased the longer that CFs had been in place (Pailler et al., 2015). Gross-camp (2017) also found greater food insecurity and better water access in CF sites than non-CF sites and evidence of increasing inequality as more wealthy community members were more likely to increase in wealth than poorer community members. However support for CFs within these communities remained in spite of the absence of tangible benefits due to the control over forests that CFs afford villages (Gross-camp, 2017). As such studies of the social impacts of CFs tend to show modest and often non-significant social impacts (e.g. Rasolofoson et al., 2017), or trade-offs between different domains of wellbeing such as income and material wellbeing versus land use rights and the freedom domain of wellbeing (Hajjar et al., 2020).

Studies commonly examine either the social or conservation outcomes of different PA governance approaches, but few studies explore both in the same analysis and this is a recognised gap in the literature (Hajjar and Oldekop, 2018). One exception is that by Sims (2010) which combined socio-economic information from a poverty mapping exercise with satellite imagery estimates of change in forest cover to compare localities in Thailand that had a high proportion of the land under strict forest protection with localities that had a lower proportion of land protected. The study found that areas that had a higher proportion of land protected had significantly reduced forest clearance. Furthermore, the study found that localities with a higher proportion of land protected had a decrease in the ratio of households in poverty. An apparent win-win. Sim (2010) attributes this to wildlife tourism revenue offsetting the opportunity costs of protection. However measures of the Gini coefficient showed that income inequality increased more in localities that had a higher proportion of land protected.

Only one study by Oldekop et al., (2019) has reported the impact of both wellbeing and conservation findings for CFs programme in the same analysis through a national study of CFs in Nepal. The study used statistical matching to control for a suite of confounding drivers of deforestation and development, and through this identification strategy, the authors assert that assignment to treatment and control units is 'as if random' (Oldekop et al. 2019). Response variables used are deforestation and secondary census data to calculate a universal multidimensional poverty index (MPI) and found that the presence of CFs increases the likelihood of win-win outcomes. A global meta-analysis of the published literature on PA impacts showed that: (1) PAs which reported positive social outcomes were more likely to also report positive conservation outcomes; (2) positive development outcomes was a better predictor of conservation success than forest tenure (Oldekop et al., 2016). This suggests that trade-offs between conservation and wellbeing may not be inherent, but that positive wellbeing outcomes and conservation outcomes may be mutually supporting.

1.6 The introduction of market-based approaches

Reduced levels of forest exploitation impose an opportunity cost on users – the forgone benefits from alternative land uses (Adams, Pressey and Naidoo, 2010; Poudyal et al., 2017). Despite CFs providing tenure security, the short-term costs of reduced extraction rates may be too high to deter users from over-exploitation. This short-term opportunity cost is a clear barrier to sustainable use and win-wins. Recognition of these challenges following the low success of PA decentralisation trends of the 1980s, and building on the concept of ICDPs came a growing emphasis from the 1990s onwards towards the use of market-based instruments to incentivise sustainable use (Naughton-Treves, Holland and Brandon, 2005; Agrawal, Chhatre and Hardin, 2008). This has largely been in payment for ecosystem services (PES) projects that aim to overcome this barrier by providing financial compensation to users where sustainable use practices can be demonstrated (Wunder, 2007).

The following section reviews evidence of the effectiveness of two prominent market-based instruments that seek to deliver win-wins for tropical forest management; REDD+ and FSC certification, drawing on lessons learnt and focussing primarily on market-based approaches implemented in CFs. Due to their relative recentness, few assessments of the effectiveness of REDD+ projects are yet available. Many REDD+ pilot projects have ended before becoming fully operational with only approximately one third of REDD+ projects successfully selling verified carbon credits (Duchelle et al., 2018). Exploratory qualitative studies of the impact of these aborted projects have identified community disappointment, growing distrust of conservation organisations and a lack financial compensation (Massarella et al., 2018).

A study of REDD+ in Nepalese CFs suggests that the opportunity costs of harvesting restrictions cannot be adequately offset by REDD+ monetary payments (Bastakoti and Davidsen, 2017). While assessments of the equity of the REDD+ compensation mechanism have suggested that the benefits are more accessible to more educated community members with secure land tenure, thereby contributing to pre-existing inequalities within communities (Calvet-Mir et al., 2015; Scheba and Rakotonarivo, 2016). A meta-analysis of the published literature on REDD+ projects in CFs suggests that in general, these projects limit the ability of communities to dynamically manage their natural resources by locking communities into carbon contracts that prohibit exploitation (Hajjar, Engbring and Kornhauser, 2021). This has serious implications for non-material dimensions of wellbeing, such as the security of communities in times of scarcity. These shortcomings have led some practitioners to suggest that REDD+ projects are halted until these issues can be resolved (Leggett and Lovell, 2012). In terms of conservation, early results are more promising. A randomised control trial experiment of villages enrolled in a REDD+ project in Uganda showed that villages enrolled in the scheme had half the deforestation rate of control villages (Jayachandran et al., 2017). However more rigorous studies are required to corroborate these findings.

FSC certification was officially launched in 1994 and is a global initiative working across different forest management types from large scale private plantations to local community run forests and the most widespread certification in the tropics (Burivalova et al., 2017). The following section reviews FSC projects relating to the sustainable management of natural forests, rather than plantations. ‘FSC forest management certification confirms that the forest is being managed in a way that preserves the natural ecosystem and benefits the lives of local people and workers, all while ensuring it sustains economic viability’ (FSC, 2018). As such FSC certification aims to achieve win-wins by providing a price premium to forest managers on the sale of forest products if sustainable forest management can be demonstrated.

Over the 25 years of FSC operation various evaluations have taken place. However rigorous evaluations of tropical forest projects are rare and in many cases it is not possible to interpret robust trends due to small sample sizes and confounding variables that were not accounted for (Romero et al., 2017; Burivalova et al., 2017). This challenge is summarised by

Ven and Cashore (2018) as 'the difficulties of establishing causality amidst a seemingly infinite spectrum of confounding variables'.

A study in Malaysia by Imai et al., (2009) and a study in Gabon by Medjibe et al., (2013) used only a single FSC site and single control site which limits the ability of these studies to generalise their findings. While a study in Indonesia compared four FSC sites with six control sites and found no difference in logging induced carbon emissions (Griscom et al., 2014). This study selected sites of similar size, slope and elevations. However as with the CF evaluations described above, the forest management histories of the different sites were not discussed and so it is not possible to conclusively attribute the findings to the recent FSC initiatives. Kalonga et al., (2014) explored the socio-economic impacts of FSC in Tanzania in two villages with certified CFs compared to two non-CFM villages and showed that communities with certified CFs were able to obtain a higher price for timber sales than control villages, had higher average salaries and lower income inequality. In one of the most extensive FSC evaluation Cerutti et al., (2014) assessed the development outcomes of certified and non-certified sites in Republic of Congo, Gabon, and Cameroon. The authors report improved living conditions in villages engaged in FSC certification. But the authors note that this finding might be due to other confounding factors such as area, access to market and spill-over effects from other neighbouring development initiatives. In terms of conservation outcomes, Kalonga et al., (2015) suggest that two FSC certified forests in Tanzania had less degraded forest structure than four control forests, however the analysis does not account for pre-selection bias of FSC sites and so the mechanism causing this outcome is uncertain. Evaluation of conservation impacts of FSC in Mexico and Indonesia using remotely sensed data suggest that FSC certification had small, but observable impacts on rates of deforestation (Blackman et al., 2015; Miteva, Loucks and Pattanayak, 2015). Systematic reviews highlight weaknesses of research designs limiting the ability of studies to causally attribute impacts due to the poor quality control groups for comparison (Komives et al., 2018).

Overall, due to the short duration of REDD+ projects, few examples of rigorous impact evaluations exist. The rigorous studies available in the published literature suggest that both REDD+ and FSC certification can demonstrably contribute to positive conservation outcomes, but there is less reliable evidence of positive development outcomes. Studies fail to account for confounding variables and this hampers causal inference of the mechanism driving the development and conservation outcomes observed. As with assessments of forest management strategy, few REDD+ or FSC certification assessments analyse conservation and development outcomes simultaneously and this is a barrier to understanding whether a trade-off or win-win paradigm is dominant.

1.7 Repeated failings of PA governance to achieve win-wins

Overall, the above studies suggest that trade-offs are more common than win-wins in PA governance (McShane et al., 2011; Sunderland and Campbell, 2007). Positive impacts on human wellbeing appear to be more rare than positive impacts on conservation effectiveness. Yet despite their rarity, win-wins do exist. Indeed, positive social outcomes have been observed to be a more reliable predictor of positive conservation outcomes than coarse distinctions between PA governance approach (Oldekop et al., 2016; Pullin et al., 2013). These associations begin to provide insight into how win-wins might be achieved through incentivising pro-conservation behaviour. The wide variability in performance within PA governance categories indicates that site specific contextual factors may provide more causal explanation for PA impacts than the overarching governance approach. This suggests the need to look more deeply into site-specific governance dynamics to understand PA impacts.

In many developing countries central agencies lack the capacity to effectively enforce restrictions on use, resulting in nominal 'paper parks', *de facto* open access conditions and overexploitation (Brandon, Redford and Saunders, 1998; Nasi et al., 2008). A second issue is that in many developing countries, where state control and accountability is weak, informal lines of patronage, in the form of undeclared rents paid to officials, can act as major inefficiency, limiting the ability of communities to realise the benefits of forest tenure and usage rights. These informal institutions act as a strong disincentive for officials to relinquish control over forest resources (Nelson, 2011). As such Agrawal and Ribot (1999; Agrawal, 2007) advocate that to understand the effectiveness of a policy, it is not the content of the policy document that is of interest, but how the rights of stakeholders change in practice. The realities of the extent to which policy reforms are implemented are therefore a key determinant of whether win-wins are achieved. Furthermore understanding the site specific histories of engagement with NGO, private sector and other outside forces may influence the opportunities available to communities to benefit from CFs. Indeed studies have shown that length of engagement in a particular PA strategy is a good predictor of efficiency (Pailler et al., 2015). Under short donor funding cycles, NGO led initiatives such as CFs may end before the barriers to achieving win-wins are overcome.

Forest tenure reforms that aim to divest control from central government to local communities have faced criticism for delivering incomplete decentralisation; only partially divesting control of forest management from central government to local levels (Murombedzi, 2011). In several countries in Sub-Saharan Africa the rhetoric of decentralisation has not been supported by substantive reforms to institutions (Nelson, 2011). In such cases land tenure remains with the state, or control is shifted from central government to local government (Meshack et al. 2006), rather than directly to the local residents. Instead only access rights for valuable products such as timber and wildlife are transferred to communities who are still obliged to pay rents to local governments, which can reduce the profitability of forest-based livelihoods for community members (Vyamana 2009; Blomley et al., 2011). These limited reforms focus on increasing the usage rights of communities to resources that remain under state control, rather than comprehensive transfer of land tenure and land-use decision making (Murombedzi, 2011). Despite these systematic differences in how CFs have been implemented, meta-analyses to date have not assessed nuanced differences between CF approaches in different countries. Meta-analyses are limited to coarse distinctions between forest governance types. Greater Integration of these political ecology and rights-based perspectives into assessment of CFs may provide further insights into why different approaches succeed and fail. For example, a consequence of incomplete decentralisation is that decision-making control is vested in local elites affiliated with local government, who disproportionately benefit from land-use decisions, thereby reinforcing local inequalities (Shackleton et al. 2002; Roe et al. 2009). This may explain why cases of elite capture are documented even within decentralised systems of forest management.

In terms of the market-based approaches of REDD+ and FSC certification, these initiatives have tended not to deliver the promised economic dividend for sustainable management practices (Lund et al., 2017; Humphries et al., 2012). As such the opportunity costs of reduced forest exploitation by local communities have not been fully compensated for. Additionally the distribution of benefits has not been equitable (Gurney et al., 2015). This influences the legitimacy of interventions in the eyes of stakeholders (Luttrell et al., 2013). Therefore, negative development impacts of forest management on local communities may result in a reciprocal negative impact on conservation outcomes as community members are not incentivised to change their unsustainable patterns of forest use. Incomplete compensation to local communities for reducing levels of forest exploitation is a major barrier to achieving win-wins in tropical forest management. This benefit deficit might be overcome by iterative refinement to improve the performance of the existing, underperforming approaches. Incremental change of existing approaches through adaptive management has the benefit of avoiding the conservation fad, boom-bust paradigm and disengagement of target communities (Massarella et al., 2018; Scoones, 2016). However this approach is undramatic and may lack novelty and the “hook” often needed to attract donor support (Redford, Padoch and Sunderland, 2013, pg 437). Iterative refinement of existing initiatives through adaptive management therefore faces challenges where projects require long-term donor funding. This funding reality is particularly problematic given that the management approaches reviewed are predominantly NGO driven, not economically viable in their own right and dependent on donor funding support.

Forest related PES projects seek to provide additional compensation payments for reduced levels of forest exploitation, thereby incentivising conservation and/or sustainable use (Fletcher et al., 2016). REDD+ has generated nearly £8 billion (Norman and Nakhooda, 2014). However the performance-based nature of these payments mean that projects require an introductory phase of project implementation, which has often not led to the long-term financing hoped for (Lund et al., 2017). REDD+ projects are conceptually complex and tend to be driven by outside actors such as donors or NGOs, rather than communities, the target beneficiaries (Sandbrook, 2010; Latham, 2013). Analyses of REDD+ projects have documented many misunderstandings between implementing organisations and community participants, leading to disappointment and anger when the promised payments do not arrive (Massarella et al., 2018; Scheba and Rakotonarivo, 2016). High variability of the price of carbon credits and the novelty and unpredictability of a carbon market still in its infancy have led to few projects generating regular carbon payments (Fletcher et al., 2016). Some commentators have therefore labelled REDD+ a ‘conservation fad’ that has failed to deliver on its promise of fully compensating for reduced forest exploitation (Redford, Padoch and Sunderland, 2013; Lund et al., 2017). In part this is due to the failure of the Paris climate agreement to create binding national caps on carbon emissions, so demand for carbon credits has remained low and a viable long-term carbon market is unlikely to emerge (Angelsen et al., 2017).

Given these largely market-based failings, forest certification that seeks to augment existing timber markets, rather than establish a new market for carbon may provide a more reliable way to generate compensation payments for reduced forest exploitation. Timber sales based on an established market are more tangible and less abstract than carbon sales and so may present less opportunity for misunderstanding and distrust with intended beneficiaries.

In this respect FSC projects have been reportedly well received by target communities (Kalonga, et al., 2014). However studies of community based enterprises implementing FSC certified timber sales in Brazil suggest that some remote communities struggle to access markets effectively in order to realise the price premium potential of FSC certification (Humphries et al., 2012). Cost-benefit analyses have also shown that the small scale-production of some community based FSC projects means that net profits are marginal and so long-term financial viability is questionable (Humphries et al., 2012). Furthermore the large start-up and verification costs and technical capacity required to provide adequate reporting to satisfy certification audits act as major barriers to the viability of such approaches to deliver long-term development benefits, with much of the financial benefit being absorbed by accreditation bodies (Blackman and Guerrero, 2012, Burivalova et al., 2017). As with REDD+, the need for technical capacity and slow financial return requires continued donor and technical support to deliver payments. These findings suggest that FSC certification projects often engage with the market sub-optimally and this limits the ability of this market-based approach to provide significant benefits to communities. Finally, given that meta analyses suggest that positive development outcomes may be a prerequisite for positive conservation outcomes (Oldekop et al., 2016), the poor performance of CFs to deliver positive social change may explain poor conservation performance. The modest/negligible social benefits accrued to communities from CFs may not be sufficient to incentivise pro-conservation behaviour.

Linking communities with global markets through certification programmes requires understanding of global, as well as local scale processes and engagement with a wider range of actors (Wiersum, Humphries and van Bommel, 2013). The added complexity inherent in combining market-based approaches with local forest tenure requires greater management capacity. This reduces the autonomy of CFM and takes the control of forest management and delivery of benefits further from communities who lack the technical expertise to engage in such schemes without significant training. The importance of such diminished agency should not be overlooked, given that a sense of control has been identified as a major reason for the popularity of CFs among communities (Gross-camp, 2017). Therefore market-based approaches implemented in CFs should seek to empower and retain community decision-making ability, allowing communities to make informed choices about how they wish to engage with global markets. To engender local support, informed decision-making and equitable benefit sharing are necessary.

Overall, the preceding literature review suggests that important contributing factors for positive social and environmental outcomes in contemporary PA governance are:

(1) supportive political environment, particularly national regulations that recognise and respect the tenure rights of different actors;

(2) equitable governance processes, including institutional transparency and accountability of PA governing institution;

(3) access and integration into stable and supportive markets local, national or international markets;

(4) ability of PA governing institution to enforce regulations;

(5) in-depth knowledge of socio-ecological context;

(6) technical capacity and adaptive governance arrangements supporting ongoing refinement of governance in the face of changing requirements;

(7) long-term financial support to overcome initial implementation challenges, build trust and lasting benefits.

These enabling factors suggest the need for PA governance arrangements to align with wider socio-economic processes across scales from local, site-based governance to national political processes and international market integration or funding from donor agencies. This requires effective engagement with a diverse range of actors and a high degree of proficiency in skills that go beyond the conventional training of natural resource scientists and managers. Specifically, community engagement and governance aligned with local values and norms, business, political and marketing expertise. The need for integration of PAs with the wider socio-economic system places a high demand on PA governing institutions to develop diverse expertise and suggests the need for interdisciplinary approaches to evaluate PA effectiveness.

1.8 Knowledge gaps and challenges for evaluating win-wins in PA governance

This review has identified three areas in which data gaps exist that inhibit the evaluation of win-wins in PA governance; (1) conceptual gaps that create uncertainty about how a win-win should be framed and understood; (2) methodological gaps concerning how the impact of

PAs should be measured; (3) explanatory gaps concerning why PA governance approaches have such variable impacts.

A key conceptual knowledge gap concerns how to understand social impacts of PAs. While there is increasing agreement on a general approach to measuring wellbeing, the flexibility and deliberate adaptability of the concept causes some ambiguity as to how exactly the concept should be measured. In the absence of clear guidelines, assessments have characterised wellbeing impacts in various ways, often prioritising economic indicators and potentially missing unintended non-economic impacts (Woodhouse et al., 2016). The concept of perverse outcomes is well established in conservation science (Maron et al., 2016). For example market based-approaches, which prioritise economic benefits for participating communities, have also been associated with negative social impacts, such as weakening community cohesion and creating social conflicts. Furthermore, the increasing focus on wellbeing of individuals rather than the economy as a whole, suggests the need to disaggregate communities to understand how different groups area affected (Fry et al., 2015). In terms of appropriately framing research on win-wins, this suggests that a social win cannot be discerned by a single aggregated result. Instead, research designs should seek to answer the more granular question of who are the winners and the losers? Therefore, win-win research on PA impacts should move beyond simplistic framings to better understand impacts on the diversity of different actors in adjacent communities.

Methodological knowledge gaps identified in this review have centred on the weaknesses of existing designs and need for integration between quantitative and qualitative approaches. To date few assessments of PAs incorporate both social and environmental data. This lack of a broader, interdisciplinary framing of analyses is an initial barrier to understanding whether a win-win or trade-off paradigm is dominant (Ferraro and Hanauer, 2015). In the field of impact evaluation statistical matching is increasingly preferred, due to its ability to mimic experimental conditions and isolate PA impacts (Baylis et al., 2016; Schleicher et al., 2020). However its effectiveness depends on correctly identifying suitable controls (Ferraro and Hanauer, 2015), which may be enhanced through incorporating qualitative evaluation criteria. Top-down, quantitative research designs that test *a priori* hypotheses are common in PA impact evaluations. However prioritising particular perspectives and ways of knowing risks marginalising more subjective or locally important PA impacts (Massarella, Sallu and Ensor, 2020), such as how PA governance may impact perceived fairness, legitimacy and willingness to support conservation interventions (Oyanedel, 2020). The dominance of top-down framings have also been referred to by Sayer and Wells (2004) as the “tyranny of the log frame", causing the design of rigid analysis frameworks to test pre-defined outcomes, rather than allowing a more exploratory approach to enable identification of unplanned impacts. The field of impact evaluation might be advanced by greater use of mixed-methods that integrate the strengths of statistical matching methods with more qualitative approaches to ensure consideration of relevant context-specific factors.

Despite significant evaluation effort, the wide variation in observed PA impacts makes it challenging to develop generalisable theory about the success of different governance approaches. There remains ongoing debate about the relative merits of centrally versus locally governed PAs, market-based approaches and the validity of the win-win concept (Lele et al., 2010; Muradian et al., 2013; Hajjar et al., 2020). Christensen (2004) describe how "the myth of win-win solutions created a culture in which overly ambitious projects proliferated, based on weak assumptions and little evidence". There is a clear need for more rigorous evaluation of the assumptions and logic underpinning these key concepts and PA approaches in order to develop a robust evidence base to inform effective decision-making (Ferraro and Hanauer, 2014a). However explaining why some PAs succeed and others fail is hampered by the growing complexity and of PA governance arrangements as new approaches are superimposed on existing arrangements. This makes it challenging to disentangle the relative impacts of different approaches, such as local governance versus forest certification (Gavin et al., 2018; Figure 1). Furthermore the diversity of governance arrangements within each governance approach have increased as new approaches are iteratively refined and adapted to local contexts (Ingram, Ros-Tonen and Dietz, 2015). For example, simplified forms of forest certification have been developed that have less stringent requirements to reduce the technical capacity required to participate in forest certification schemes (Burivalova et al., 2017). The specifics of the governance arrangements and local context may be more reliable predictors of success than the broad governance category (Ostrom, 2007). Therefore, there is a need for PA evaluations to embrace the complexity of real-world situations, enabling a deeper examination of the specifics of why particular impacts are observed.

1.9 Aims and objectives

I aim to evaluate how innovations in PA governance have led to complex impacts on human wellbeing and conservation effectiveness. To meet this aim and in the light of identified knowledge gaps relating to PA governance evaluation and win-wins, this thesis has the following objectives:

To critically review current trends in tropical PA governance and their impacts on human wellbeing and conservation effectiveness (Chapter 1).

To develop appropriate methods for measuring human wellbeing in different contexts in order to resolve conceptual ambiguity (Chapter 2).

To evaluate the impact of certified CFs on human wellbeing and conservation effectiveness as an example of recent innovations in PA governance seeking to achieve win-wins (Chapter 3).

To test the assumptions and logic underpinning certified CF governance approaches (Chapter 4).

To advance robust interdisciplinary analysis techniques to provide a well-rounded representation of the situation on-the-ground, thereby providing useful insights for conservation practice that contribute to bridging research-practitioner divides (Chapters 2-5)

To synthesise key knowledge and methodological contributions, discuss the challenges and limitations of evaluating and implementing PA governance for win-win outcomes and identify frontiers for future research (Chapter 5).

1.10 Research approach

1.10.1 Positionality

Prior to undertaking this PhD my academic training was primarily from the natural science disciplines of biology and ecology. This theoretical understanding has been supplemented by several years of 'real-world' experience of working in applied conservation for international conservation NGOs, focusing on site-based conservation of tropical forests in Africa and South-East Asia, which demonstrated to me the complexity of conservation challenges. Natural sciences tend to take a top-down perspective, focusing on external validity (Campbell and Stanley, 1963). This perspective tends to adopt a positivist research philosophy, assuming that an objective trust is observable, prioritising quantitative data collection, large sample sizes and a deductive, deterministic research approach, e.g. assuming that x leads to y.

In contrast, the social sciences and particularly the qualitative social sciences tend to adopt a research philosophy orientated more towards constructivism-interpretivism (Creswell, 2012). This research philosophy takes the position that there are multiple ways of knowing stemming from the subjective beliefs and experience of different actors and therefore, multiple truths (Creswell, 2012). This philosophical lens prioritises 'understanding the complex world of lived experience from the point of view of those who live it' (Schwandt 1994: 221). Therefore, taking a bottom-up perspective and emphasising ecological validity (Yue, 2012). Natural and social science disciplines thereby reflect contrasting ontological (the nature of reality) and epistemological (the relationship between the researcher and research subject) positions (Creswell, 2012).

In this PhD I have been keen to expand my understanding to explore alternative disciplinary perspectives of conservation challenges. By studying taught units in preparation for commencing field research in new disciplines, including qualitative methods and development economic theory, I developed a reflexive understanding of how my beliefs and experience might colour the research process and indeed, my research findings. Through the research process, my research philosophy has evolved into one of pragmatism. This philosophy sits at the intersection of objectivity and subjectivity, recognising truth as what works in a particular context relative to the perceptions of different actors (McCaslin, 2012). This position has emerged, partly in response to my supervisory team, which has been composed of representatives from both natural science and qualitative social science traditions. Therefore, throughout the research process I have sought to articulate points of tension stemming from contrasting disciplinary approaches, navigating these divides to forge an interdisciplinary 'middle-way' between philosophical positions. For example, in chapter two I set out to develop an approach for measuring human wellbeing in different contexts. In the process, I articulate the tension between universal and locally specific measures and suggest an approach that samples from different local actors, recognising a pluralistic perspective, while using statistical simplification techniques to reduce the complexity of identified indicators.

My research aspires to be relevant for informing practical site-based conservation decision-making. I have found the philosophy of pragmatism an appropriate lens in this regard for exploring the complexities of conservation challenges because it recognises (1) that no panacea exists for success, but rather embraces contextual complexity; (2) the perceptions of local actors are important drivers of behaviour and ultimately, the success of conservation interventions (Bennett, 2016).

1.10.2 Methodology

I adopted an exploratory mixed methods design, undertaken at landscape scale. Mixed methods research focuses on collecting, analysing and mixing both quantitative and qualitative data in a single study. Its central premise is that the use of quantitative and qualitative approaches in combination provides a better understanding of research problems than either approach alone (Creswell and Clarke, 2017). Here I adopted an exploratory mixed-methods design that first ground-truthed and expanded existing theory through an initial research phase of exploratory qualitative enquiry, followed by a quantitative testing phase of identified key relationships. I drew on and suggested approaches for integrating (1) methodological best practices from the quantitative disciplines of conservation science, ecology, economics and impact evaluation with more exploratory qualitative approaches from the social sciences, (2) top-down and bottom-up research perspectives.

By integrating the strengths of qualitative research to develop locally grounded theory in complex real-world settings, and the strengths of quantitative approaches to empirically test hypotheses and make generalisable conclusions, I aimed to develop robust analytical approaches to the different forms of scientific validity discussed. This cross-validation of research findings, gathered through different methodologies, if in agreement, has the additional benefit of giving further assurance that the analysis provides a coherent, robust understanding of the research problem. The methodology I adopted seeks to shine analytical lenses on PA governance from both top-down and bottom-up perspectives in order to build a more comprehensive understanding of intended and unintended impacts. All three data chapters adopted this approach. In chapter two I explored integration of perspectives by developing methods for analysing local conceptions of wellbeing (bottom-up perspective) that are consistent with accepted universal theory (top-down perspective). In chapter three I integrated qualitative approaches into a statistical matching research design by incorporating qualitative criteria e.g. (political stability) and an expert panel review to validate the appropriateness of candidate control villages identified through statistical matching. For chapter four I led exploratory focus groups and undertook key informant interviews to inform the design of a causal model that disaggregated the different means by which certified CFs impact human wellbeing and conservation effectiveness. I then tested the causal model by quantitative and qualitative analyses.

I undertook this methodological approach at landscape scale since landscapes are generally complex, multi-use environments in which governance arrangements must seek to balance multiple, often conflicting social and conservation objectives (Sayer et al., 2017). This spatial framing thereby provided an appropriate scale for exploring complex social-conservation trade-offs in action. In this way the scale of the study was targeted to provide practical insights to inform conservation practitioner perspectives and policy-relevant findings.

An additional practical reason for undertaking research at landscape scale was that PAs are increasingly being shown to have impacts beyond their own borders and to be impacted by the surrounding environment (DeFries, Karanth and Pareeth, 2010). Positive buffering effects may be present as PAs enhance protection of neighbouring land-uses, or cause negative spillover effects by displacing deforestation to neighbouring sites (Pfaff and Robalino, 2017). Accurate assessment of PA impacts should seek to account for these interaction effects and consider how PAs are interrelated with surrounding land-uses (DeFries et al., 2007). A landscape-scale framing therefore embraced this real-world complexity of multiple-interacting landscape features.

1.11 Study area and research context

I apply and test this methodology in Tanzania, which provides an excellent test case of the challenges of reconciling PA conservation with improving human wellbeing because 38% of the country is protected (IUCN, 2020), while national development policy is explicitly targeting improvements in human wellbeing (MFP, 2016). Tanzania has a history of state-controlled PA governance, established during the German, then British colonial eras from the 1920s onwards. However the country has also received significant international funding for experimentation with new trends in PA governance. PA decentralisation initiatives have been supported from the 1990s, leading to the Forest Act of 2002 which created the legal premise for communities to independently govern and retain revenues from forests on village land (MNRT 1998, 2002). To date approximately 2.5 million ha (10%) of Tanzania's forests have been designated as CFs (Blomley et al., 2019). Market-based forest conservation approaches including Reduced Emissions from Deforestation and Degradation and forest certification projects have also received significant international investment since the early 2000s (Blomley et al., 2019).

At the time of commencing my research, decentralised forest governance and the institutions, including participating communities, NGO's, donors governmental actors were coming under increasing pressure to justify the positive impact of this governance paradigm (Nelson, 2011; Basnyat et al., 2018). The national political context, as with other countries in Asia and Africa, was tending towards a preference for recentralisation of natural resource governance (e.g. Magessa, Wynne-jones and Hockley, 2020). Set against a backdrop of this potential sea change in natural resource governance, my research was designed to test the impact of community forest governance to provide an evidence base to help inform future directions for forest governance in Tanzania.

I undertook research in two focal landscapes (Figure 1.2). First, field data collection for chapter two was undertaken through pilot field visits to the Udzungwa-Kilombero Landscape in Morogoro and Iringa regions of central Tanzania where existing logistics support could be provided by University of York partner organisation Reforest Africa. This landscape encompasses the Udzungwa Mountains and Greater Kilombero Valley and is part of a recognised biodiversity hotspot, though threatened by logging and increasing demand for land for farming (Willcock et al., 2016). The landscape was dominated by strictly protected areas and provided a good initial location for exploring human wellbeing and PA and forest relations of PA adjacent communities within a simple governance context.

Through an additional pilot field trip I then established a partnership with a Tanzanian NGO Mpingo Conservation Development Initiative (MCDI), who work with district government and local communities to support the establishment of CFs combined with FSC forest certification in Lindi and Pwani regions of Tanzania in the Kilwa-Rufiji Landscape. The landscape was again dominated by PAs, including traditional strictly PAs and certified CFs. In 2009 communities in this landscape became the first in Africa to obtain a FSC certification for CFs (FSC, 2018). This programme provided an innovative example of recent advances in PA governance with direct relevance to the aim and objectives of this thesis and so data collection for chapters three and four were undertaken in this landscape.

In this landscape, MCDI supports communities to integrate the requirements for forest certification with national CF legislation. Additional measures include requirements to evidence 1. forest inventory assessments to verify that harvesting is being undertaken sustainably (FSC principles 5, 6 and 9), 2. transparent governance procedures are in place to ensure fair benefit sharing among community members and access to dispute resolution mechanisms (FSC principle 4 and 7), 3. promotion of gender equality through equal provision of employment, training opportunities and positions on decision-making councils (FSC principle 2; FSC, 2015). The NGO works alongside district government to provide training support to participating communities to ensure that these requirements are met and to develop a coherent governance system that integrates national legislation and FSC certification requirements.

The dominant religions in the study regions were Christianity and Islam, with a greater proportion of people identifying as Christian in Morogoro and Iringa regions, while a greater proportion of people identified as Muslim in the coastal region of Lindi where the Arabic influence from coastal trade is stronger. Livelihoods centred on cultivated agriculture, with subsistence forest use for food, building materials, firewood collection and traditional cultural practices still an important part of daily lives (Latham et al., 2017).

Figure 1.2. Top left: map of Tanzania detailing the Udzungwa-Kilombero Landscape (blue), Rufiji-Kilwa Landscape (green), major city of Dar es Salaam (black dot), international borders (grey). Top-right: the mountain village of Udekwa in the Udzungwa-Kilombero landscape. Bottom-left: Typical mud brick and straw roofed house in Makangaga village, Rufiji-Kilwa landscape. Bottom-right: Focus group with women participants from Nainokwe village, Rufiji-Kilwa landscape.

The process of establishing certified CFs in the region requires that communities send a letter of interest to the NGO requesting support to establish the CF. The NGO will then begin investigating the potential to establish a CF following the criteria set out in chapter 3 supporting information. Through their long-term presence in the region dating back to 2003, the NGO has sort to build trust and be embedded within the social life of the villages in which it works. In this way the NGO has developed an organic, bottom-up approach to expansion, working with villages who first request their support, rather than imposing their presence in a top-down way.

Protected areas have potential to be located non-randomly in space due to particular social and environmental features (Joppa and Pfaff 2010). Certified CFs likewise have potential to be preferentially located in villages with abundant natural resources and high levels of pre-existing community cohesion, supporting effective collaborative governance. There is therefore potential for these pre-existing characteristics to confound estimates of certified CF impact. A mixed quantitative and qualitative matching approach was used to overcome this selection bias in CF establishment (chapter three). For example, certified CFs were matched with control villages who had also sent a letter requesting support. Thereby matching treatment and control groups based on demonstrated evidence of commitment to collective action.

I partnered with MCDI as they had a high level of technical understanding of supporting community forest governance combined with FSC certification. I worked from the NGO's head office in Kilwa in between field visits to the villages engaged in certified CF governance and so I had the opportunity to be, to an extent, embedded in the life and workings of the NGO. This provided the opportunity for me to develop a good understanding of the process by which the NGO supports community forest governance. Over several workshop sessions I was able to document how representatives from the NGO perceived that certified community forests impact conservation and human wellbeing. Through key informant interviews with both MCDI representatives, as well as district government and district and national representatives from Tanzania Forest Service, I was also able to develop an understanding of the practical factors driving forest loss in the region, and the challenges for implementing a successful model of sustainable forest governance. Incorporating this practitioner knowledge was an important component in the co-creation of the research design of data chapter four and crucial for identifying potentially confounding variables to control for in the matching study in chapter three, to support impact estimates that appropriately control for local contextual drivers. The partnership therefore provided an excellent means of integrating and embedding the local context and practitioner knowledge into my research.

The research was also designed to feedback the research findings to the partner NGO and local communities. A feedback presentation on the findings of the study was given to the NGO and government representatives. Feedback sessions were also provided in village council meetings on my final fieldwork trip. However, my research aimed primarily to provide an evidence base to inform national level decision-making and have relevance to global academic debates concerning shifting trends in forest PA governance and their efficacy. The research questions were also chosen *apriori* before beginning work with the government, NGO and village conservation practitioners. By prioritising this national scale and academic audience, the extent to which the findings were directly operable for translating into conservation actions was likely limited. Positioning the research as an external assessment (see ethical considerations) of the NGO supported intervention, rather than a more integrated research-practitioner co-design of the research questions was likely a constraint on the reciprocal embedding of the research findings and their practical utility for directly informing on-the-ground conservation actions.

1.12 Ethical considerations

Permission to undertake research was granted by The Tanzania Commission for Science and Technology (COSTECH) and the research design approved by the University of York Environment Department Ethical Review Committee. Free prior informed consent was given by all study participants prior to the commencement of research. Permission to commence research was also granted first by district government by visiting district government offices and making a presentation about my research objectives, then on subsequent trips sharing initial findings. Then seeking permission by each village chief upon arriving in study villages and where possible being directly introduced by the village chief to members of the village to reassure villagers that my visit was indorsed by the chief and that there should be no concerns about speaking to an 'outsider'.

My research aimed to assess the impact of the partner organisation MCDI, who have a clear mandate of being an advocate for community forest governance. Therefore to undertake an assessment of the impact of this governance approach, it was important that the research's quantitative impact assessments were not biased by this relationship. MCDI's staff were well known in the study villages and so to keep an unbiased separation from the NGO, I recruited an independent team of Tanzanian research assistants. All research in the villages began with an explanation of the research's positionality, stating our independence, being led by an independent research team from the UK.

International visitors to Tanzania are often perceived to live a very separate life from the lifestyle and slower pace of life common in rural Tanzania. To help breakdown these barriers between researcher and research participant I took care to travel by local transport to arrive in villages by the same means that villagers would. I also undertook a three-month Swahili language training course prior to initiating fieldwork so that I could converse as much as possible directly with villagers and build a connection with study participants, rather than rely exclusively on interpreters. I also camped centrally in communal village areas to encourage opportunities for conversations in the evenings and promote mutual exchange of ideas and to participate in village life, eating at local cafes.

Important ethical issues concerned being transparency about my research and giving interviewees opportunities to also ask questions. At the start of each interview I would introduce the research in Swahili and take 5-10 minutes to discuss the research and answer any questions that participants might have. Reciprocal information sharing with the interviewee was important to establish a level of trust and mutual disclosure to support open, uninhibited dialogue. Interviews were always conducted either at the participants home or in a location suggested by the participant before hand, such as at their farm if they had work to attend to. This was done to show proper respect to how busy interviewees might be and to make them feel comfortable by speaking in familiar surroundings. Attribution of potentially contentious quotes have been kept deliberately ambiguous to protect the identities of study participants.

Chapter two: Measuring human wellbeing: a protocol for selecting local indicators

Manuscript published:

Loveridge, R., Sallu, S.M., Pesha, I.J. and Marshall, A.R., (2020). Measuring human wellbeing: A protocol for selecting local indicators. *Environmental Science & Policy*, *114*, 461-469.

2.1 Abstract

Improving human wellbeing is a major focus of international environmental and sustainable development policy. However, clearly defined measures of wellbeing are needed as an empirical base for the formulation and evaluation of policies. Despite conceptual progress towards agreement of universally relevant dimensions of wellbeing, consensus is still lacking on how to translate these dimensions into locally appropriate indicators to measure wellbeing in different contexts. This paper focuses on three interrelated challenges associated with this knowledge gap: (1) navigating trade-offs between complexity versus simplicity of concept; (2) integrating top-down and bottom-up perspectives; (3) ensuring a cost-effective and flexible approach suitable for different policy contexts. We contribute to filling this gap by developing a step-by-step Wellbeing Indicator Selection Protocol (WISP) for measuring wellbeing. The protocol integrates perspectives through an interdisciplinary mixed methods design that includes cross-validation between quantitative approaches of redundancy analysis and statistical modelling and qualitative approaches of focus groups and thematic analysis. In this way we promote a pragmatic approach suitable for a range of social and environmental contexts. We tested WISP in rural Tanzania, identifying 111 candidate wellbeing indicators. This list was simplified to a subset of 19 indicators that retained 91% of measured variation among all wellbeing indicators. The simplified list was representative of both a multidimensional concept of wellbeing and the diversity of opinions sampled. We conclude that the protocol provides practical, statistically validated guidance to support the design of wellbeing assessments, maintaining coherence between universal theory and local realities.

Key words

Human wellbeing, social assessment, Sustainable Development Goals, mixed methods, trade-offs.

2.2 Introduction

Improving human wellbeing has become a major goal of international environmental and sustainable development policy (UNDP, 2015; CBD, 2016). However there remains ongoing debate about how wellbeing should be conceptualised and measured (Dasgupta, 2001; OECD; 2013). Meanwhile these high-level policy goals have largely fallen short in terms of the persistence of extreme poverty, increasing inequality and environmental degradation (Fehling, Nelson, & Venkatapuram, 2013; Allen et al., 2018; McGregor, 2018). Clearly defined, measurable indicators of wellbeing are needed to improve achievement of policy goals by (1) providing an evidence-base to track progress towards a more inclusive society (Hicks et al., 2016; Brende and Bent, 2015; Costanza, Kubiszewski and Giovannini, 2014), and (2) highlighting social issues requiring attention and adaptive action (Brown and Westaway, 2011).

The rising popularity of human wellbeing as a measure of development stems from growing recognition of the failures of economic indicators to adequately represent non-economic aspects of peoples' lives (Klugman, Rodríguez and Choi, 2011; Haq, 1996). For example, education can be a stronger predictor of health than income (Sen, 1999; Herd, Goesling and House, 2007). In contrast, the concept of wellbeing encompasses a broader notion of multidimensional development, building on an understanding of what people need to participate and flourish in society (Max-Neef, 1989; Alkire 2002). Various definitions of wellbeing exist, though none are unanimously accepted (Brown and Westaway, 2011). Here we adopt a definition developed by the Wellbeing in Developing Countries research group, which defines wellbeing as 'a state of being with others, which arises where human needs are met, where one can act meaningfully to pursue one's goals and where one can enjoy a satisfactory quality of life' (Gough and McGregor, 2007).

Three distinct dimensions of wellbeing have been identified, for which there is growing theoretical consensus; objective, subjective and relational wellbeing (Boarini, Kolev and McGregor, 2014). These form the beginnings of a unified theory of wellbeing, with contributions from diverse disciplines of philosophy, psychology, economics, and more recently, the natural sciences (Schleicher et al., 2017). Objective wellbeing is concerned with the material conditions of a person's life, often represented by wealth indicators of poverty (Mcgregor and Sumner, 2010). Subjective wellbeing is concerned with self-evaluation of personal circumstances (Vanhoutte, 2015). Examples of subjective wellbeing measurement include the Satisfaction With Life Scale, a five-question research instrument where respondents self-report their satisfaction with life as a whole (Pavot et al., 1991). Thirdly, relational wellbeing, based on the capabilities approach of economist Amartya Sen (Sen, 1999), concerns the opportunities available to a person, recognising that individual wellbeing is pursued in relation to other people (Gough and McGregor, 2007, Woodhouse et al., 2015).

Progress towards operationalising wellbeing has been made through increasing theoretical convergence towards breaking down these broad conceptual dimensions, into more specific but still universally relevant domains of wellbeing (McGregor, 2018). Alternative lists of domains have been suggested (for a review see King et al., 2014). However, all build on a human, rather than purely economy-centred conception of development and cover similar aspects of peoples' lives, with relabelling of alternative lists largely reflecting the specific purpose, or disciplinary approach (McGregor, 2018). Here we take an interdisciplinary approach to wellbeing for use across social and environmental contexts. We therefore adopt the domains put forward by the Millenium Ecosystems Assessment (MEA, 2005), which explicitly uses a socioecological systems approach and defines five domains: (1) Basic material for a good life - hereafter referred to as material wellbeing, (2) Health, (3) Social relations, (4) Security, (5) Freedom of choice and action (hereafter referred to as freedom; Narayan et al., 2000; Supplementary material).

There is also growing methodological agreement of a general approach for measuring wellbeing. Conceptions of wellbeing are socially constructed and since communities are not homogenous, there is a need to consider how understandings of wellbeing differ between actors and contexts (Martin et al., 2014; Wood et al., 2018). Therefore, participatory methods should be used to include the views of those individuals whose wellbeing is being assessed (Sterling et al., 2017; Camfield, Crivello and Woodhead, 2009). Furthermore, heterogeneity may exist within households (de Lange, Woodhouse and Milner-Gulland, 2016). Therefore individuals should be the unit of measurement, rather than households as a whole (Fry et al., 2015).

Despite these advances towards measuring wellbeing, a remaining knowledge gap concerns how to effectively translate universally relevant wellbeing domains into local indicators (Sterling et al., 2017; Mcgregor, 2018). We refer to 'local indicators' as incorporating context specific values (Caillon et al., 2017; Sterling et al., 2017). Here we focus on three interrelated challenges associated with selecting local indicators, which we refer to as (1) complexity-simplicity, (2) integrating perspectives and (3) practical utility.

Firstly, given the multidimensional nature of wellbeing, thousands of potentially relevant indicators exist (Breslow et al., 2016; Corrigan et al., 2017). Previous studies have identified correlations between different social indicators (Mcgillivray, 1991, S7, Supplementary material). For example, there is a strong correlation between literacy and income (Qizilbash, 2001). The inclusion of highly inter-correlated indicators provides little additional information about variation in wellbeing, suggesting a level of redundancy and the potential to use fewer indicators for concise communication of wellbeing assessments to policymakers. Furthermore, lengthy questionnaires may cause respondent fatigue (Ben-Nun, 2008) which has ethical and data quality implications. Yet, oversimplification risks losing the rich description intended by the wellbeing concept. We refer to this as the 'complexity-simplicity problem'.

We suggest that introducing the use of statistical approaches for variable reduction may help to navigate the complexity-simplicity challenge. Breslow et al., (2016) identify the need to select parsimonious sets of indicators for wellbeing assessment i.e. reducing the number of indicators without loss of the complexity required to adequately describe wellbeing. However, we are not aware of any wellbeing indicator selection methods that utilise statistical approaches to guide the process of reducing the number of indicators. Introducing the use of statistical methods provides several benefits (Murtaugh, 2009). The removal of numerically correlated indicators creates an orthogonal (uncorrelated) set of indicators (Crawley, 2007). Orthogonality among indicators is a fundamental assumption of statistical analysis and required to avoid erroneous results of any subsequent analysis of wellbeing data (Zuur, et al., 2010). Furthermore, statistical methods can be exactly repeated between sites, minimising the introduction of human bias and supporting comparison between wellbeing assessments.

A second challenge is to determine how best to integrate top-down perspectives (i.e. from wellbeing theory) and bottom-up perspectives (i.e. local knowledge of study participants; Boarini et al., 2014). We refer to this as the 'integrating perspectives challenge', which can be understood in terms of how contrasting disciplines seek to maximize different aspects of scientific validity. Top-down perspectives are common in the natural sciences and quantitative social sciences and tend to prioritise 'external validity', i.e. the ability to generalise findings to different contexts and populations (Campbell and Stanley, 1963). For example, top-down selection of wellbeing indicators may take place through a combination of literature review and expert opinion (Biedenweg, Stiles and Wellman, 2016; Breslow et al., 2017). This approach promotes external validity through strong relation to theory, but may marginalise the perspectives of those people whose wellbeing is to be assessed, thereby lacking local relevance (Grillo and Stirrat, 1997; Woodhouse et al., 2016).

Conversely, bottom-up perspectives emphasise the need for contextual understanding and 'ecological validity', defined by Yue (2012) in relation to case study research as the extent to which the researchers' findings reflect the lived experience of those whom the researchers are studying. Ecological validity ensures that local relevance is retained, promoting rather than marginalising the needs of study participants (Howard et al., 2016). Efforts to prioritise bottom-up perspectives in conceptualizations of wellbeing have been undertaken through anthropological and in-depth qualitative research approaches (Beauchamp et al., 2018; Woodhouse and McCabe, 2018). However, if an exclusively, bottom-up perspective is followed, some important issues may go unreported due to the adaptive preferences of survey respondents (Sen, 1999; Mitra et al., 2013).

The Basic Necessities Survey is a quantitative social assessment tool that builds on this bottom-up perspective, prioritizing locally defined indicators by combining focus group consultations followed by a household questionnaire (Davies, 2007). An issue with this approach is that it does not organize indicators in relation to a conceptual framework (Schreckenberg et al., 2010). This risks overlooking subjective indicators that are less easily articulated through participatory discussions, thereby invalidating conclusions about the overall wellbeing of respondents if one dimension of wellbeing is missed (Woodhouse et al., 2015). These tensions between the strengths and weaknesses of top-down and bottom-up perspectives should be carefully considered (Poteete et al., 2010). Each perspective illuminates important aspects of wellbeing, but in doing so, prioritises contrasting forms of validity, which need to be integrated to gain a well-rounded understanding of human wellbeing (Figure 2.1).

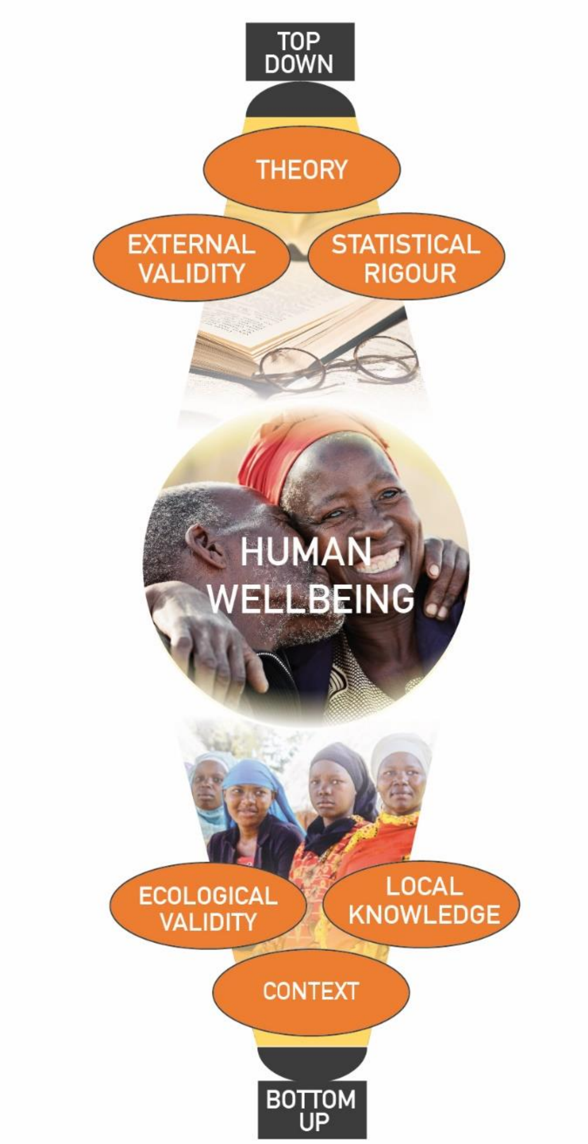


Figure 2.1. Measuring human wellbeing requires the integration of contrasting perspectives. Top down perspectives tend to prioritise external validity i.e. generalisability of findings in relation to accepted theory, while bottom-up perspectives prioritise ecological validity i.e. the experience of study participants.

The third challenge of practical utility, is that the process of selecting local indicators must be cost-effective and adaptable in order to mainstream the process into different policy contexts (Rasmussen et al., 2017).

Here, we contribute a Wellbeing Indicator Selection Protocol (WISP) that aims to operationalise measurement of human wellbeing in different contexts. The protocol provides a generalised, step-by-step method to help researchers and practitioners translate universal wellbeing domains into locally appropriate indicators. To address the complexity-simplicity challenge, we introduce the use of statistical methods to remove redundant indicators. To address the integrating perspectives challenge, the protocol employs a mixed methods design to balance external and ecological validity (Figure 2.2). To assess the practical utility of the protocol, we provide an example of its use in rural Tanzania (Supplementary material S1). We critically evaluate the protocol's effectiveness to address these three challenges.

WISP is intended to be used in the scoping phase of projects operating at landscape, or regional sub-national scales to support the design and testing of wellbeing questionnaires prior to implementation of the survey instrument. Potential applications include exploratory use to identify local priorities in order to supporting policy formulation. The protocol can also be used to support context specific wellbeing impact evaluations of conservation and development projects (for an overview of wellbeing impact design considerations see Woodhouse et al., 2015). We highlight common design considerations for wellbeing assessments and discuss implications for the protocol's wider use.

2.3 Materials and Methods

2.3.1 Generalised overview of WISP

Sample selection. Before undertaking a wellbeing assessment, the diversity of community actors present within the intended study area should be identified and particular consideration given to ensure participation of marginalised groups in a way that is culturally sensitive in the local context (Franks and Small, 2016). A minimum of two contrasting sites (e.g. villages) should be visited in order to sample variation across the study area. Selected sites should be representative of the study area and encompass variation in key socio-economic and environmental variables of relevance to the local context (PEN, 2007). Common criteria for consideration include economic drivers of wellbeing, such as proximity to local markets (Helliwell and Putnam, 2005), environmental drivers, such as topography influencing farming and other livelihood practices (Boarini, Kolev and McGregor, 2014).

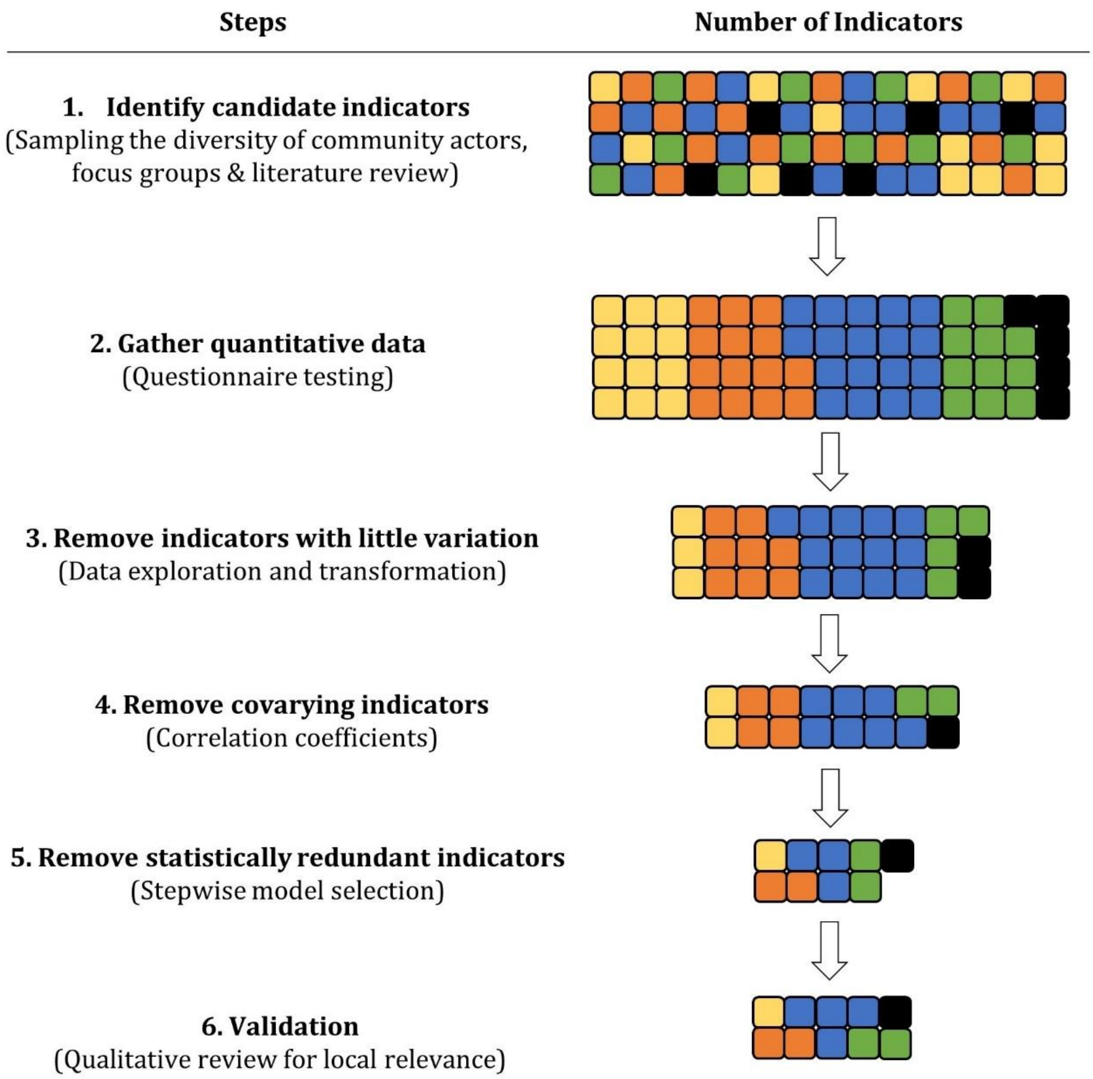


Figure 2.2. Steps for implementing the Wellbeing Indicator Selection Protocol (WISP). Boxes represent indicators coloured to represent the five domains of wellbeing.

Step 1. WISP uses a stratified random sampling design to identify an unbiased sample of local actors within villages (for an overview of sampling approaches see Angelsen et al., 2011). Stratification should use participatory wealth ranking or other criteria relevant to the specific context of the study, such as gender, age or livelihood (Supplementary material S1).

Exploratory focus groups are undertaken to identify candidate wellbeing indicators, with each focus group comprising a single community actor group (e.g. divided by gender) to reduce within-group variation, thereby encouraging uninhibited discussion and cross-validation of ideas between participants (Kitzinger, 1994; Macnaghten and Myers, 2011).

It is important to frame focus group discussions around a sufficiently broad conception of wellbeing and be careful about how this is communicated when translating between languages (OECD, 2013). An open questioning style should be used to facilitate participants to develop a locally understood conception of wellbeing, thereby promoting ecological validity (Supplementary material S2). Thematic analysis of focus group transcripts is then used to identify candidate wellbeing indicators in relation to the five domains of wellbeing (Supplementary material S2, S3), noting indicators that are specific to a particular village or community actor group and local priority indicators, which we define as those indicators discussed in all focus groups. If less than five indicators are suggested per wellbeing domain then additional indicators should be added from relevant frameworks (Supplementary material S1).

Step 2. All identified candidate indicators are used to develop a quantitative instrument (questionnaire), which is implemented with a stratified random sample of respondents (Supplementary material S4; Creswell et al., 2004). This is done to trial the wellbeing questionnaire and gain sample data of how candidate indicators vary across the study area.

Step 3. The spread of responses for each indicator is assessed to eliminate indicators with zero or uneven spread, that would give no helpful information on the variation of wellbeing present within communities. For an overview of data exploration in relation to common statistical problems see Zuur et al. (2010).

Step 4. A holistic wellbeing index (HWI) is calculated to represent all indicators in a single, standardised index following principles of the Human Development Index (UNDP, 2017; Eq.1).

$$\mathrm{HWI} = \frac{\mathrm{Material}(\bar{x}) + \mathrm{Health}(\bar{x}) + \mathrm{Security}(\bar{x}) + \mathrm{Social\ relations}(\bar{x}) + \mathrm{Freedom}(\bar{x})}{5}$$

Eq.2.1. Holistic Wellbeing Index (HWI), where $\bar{x}$ is the mean value of standardised indicators from each wellbeing domain.

HWI is used as a continuous response variable to inform further reduction in the number of candidate indicators, using high covariance between indicators to infer statistical redundancy. We define high covariance as Pearson correlation coefficient (r) ≥ |0.7|, and/or Variance Inflation Factors (VIF) ≥ 3 across all indicators (Zuur, Ieno and Elphick, 2010; Dormann et al., 2013). In the event of high covariance, the indicator that has the strongest relationship with HWI is retained.

Step 5. Statistical modelling using an information-theoretic approach (Burnham and Anderson, 2002; Burnham, Anderson and Huyvaert, 2011) is employed to achieve statistical parsimony, i.e. reduction of indicators without loss of the complexity needed to adequately describe wellbeing. The uncorrelated indicators from step four are used as predictor variables of the HWI response variable in a Generalised Linear Model (GLM) with Gaussian error; suitable for continuous variables with approximately normal distribution. Then to reduce the number of indicators to only those making the strongest contributions to overall wellbeing, we used backwards-forwards stepwise model selection (Venables and Ripley, 2002; Murtaugh, 2009). Stepwise selection is based on Akaike's Information Criterion (AIC; Akaike, 1973) to avoid consequences of frequentist approaches such as F statistics (Whittingham et al., 2006).

Step 6. Finally, it is important that the process of selecting a reduced set of indicators is not blindly automated without critical review and validation checks (Burnham, Anderson and Huyvaert, 2011). The reduced indicators are checked to ensure that each wellbeing domain contains at least two indicators to promote external validity consistent with wellbeing theory and that local priority indicators identified in step 1 are retained to promote ecological validity.

Once the indicator selection process has been completed, the retained indicators can be used either individually, aggregated to indices of each wellbeing domain, or provide an overall wellbeing index in support of different policy applications using the HWI equation listed above.

2.3.2 Study region

WISP was tested in Tanzania within a protected area dominated landscape of 10,000 km^2 in Morogoro and Iringa regions where landscape planning interventions aimed to deliver improvements in wellbeing (SAGCOT, 2016; Figure 2.3). The protocol was used to develop a context specific wellbeing questionnaire to be implemented in a further 20 villages in order to evaluate multidimensional wellbeing impacts of protected areas in the landscape using a site matching design (for an overview of statistical matching see Schleicher et al., 2020).

Two contrasting villages were selected to test WISP. Mang'ula B (Kilombero district, Morogoro region, elevation 306 m) had a population density at 23.6 people per km^2, annual population growth of 2.29%, close to the national average of 2.7% and was located adjacent to a road. In contrast, the mountain community of Udekwa (Kilolo district, Iringa region, elevation 1,611 m) had a population density one tenth of that in Mang'ula at 2.4 people per km^2, with slower annual population growth of 0.72% (Tanzania National Bureau of Statistics, 2012) and poor road access. Two focus groups stratified by gender were undertaken in each village to identify candidate indicators. Participatory wealth ranking with village leaders was used to identify a random sample of 90 questionnaire respondents stratified by gender and socioeconomic status (Supplementary material S1).

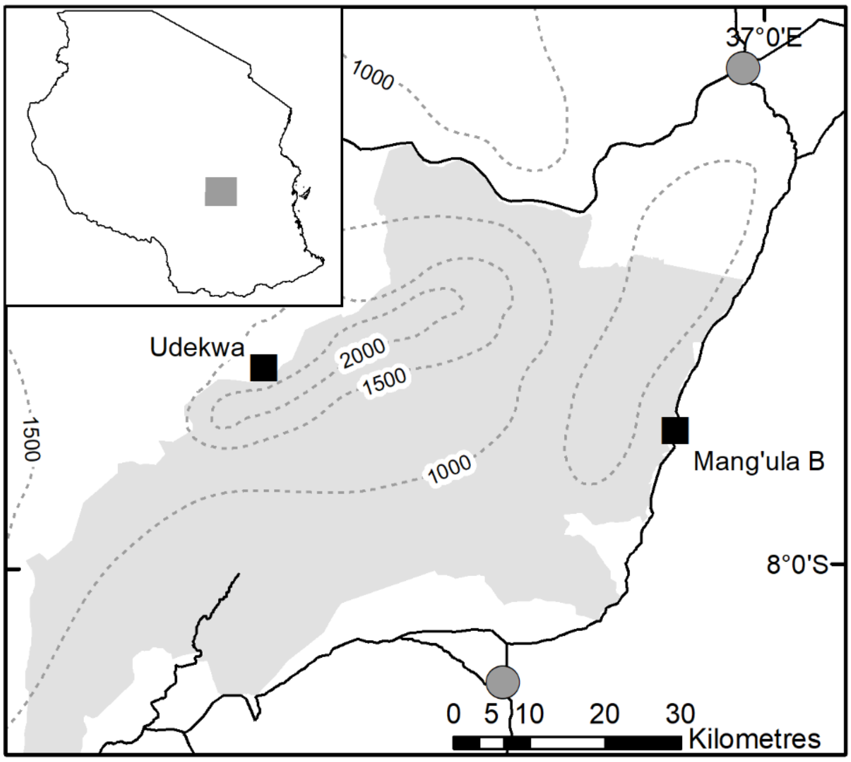


Figure 2.3. Study region in Tanzania (grey rectangle on small map) detailing study villages (black squares) in relation major towns (grey circles), major roads (black lines), protected areas (light grey polygons) and elevation (meters above sea level; dotted contour lines).

2.3.3 Analysis

We undertook analyses to understand how well WISP overcame the three challenges outlined in the introduction. To analyse the complexity-simplicity challenge, we evaluated the value of introducing statistical modelling (step five) by comparing the two lists of indicators selected by steps four and five, treating indicator lists as equivalent if within two AIC (Burnham and Anderson, 2002). The potential for further simplification beyond the final indicator list was evaluated by further indicator removal. This was done by plotting the sequential loss of deviance explained following consecutive indicator removal of the least contributory indicator (Crawley, 2007). These analyses served as both numeric and visual tools for evaluating how conservative the stepwise model selection process was in terms of simplifying the number of indicators in relation to loss of explained variation.

To analyse the integrating perspectives challenge, we assessed how well the final indicator list retained site and gender specific indicators identified during the thematic analysis of step one (Silverman, 2011).

To analyse the practical utility challenge, we retrospectively assessed the minimum number of questionnaire replicates needed to reach our statistical conclusions using the 'pwr' power analysis package in R (Cohen, 1988, Campely, 2020). For step 4, we evaluated the sample size needed to detect a correlation between indicators at a correlation coefficient of 0.7. For step 5, we evaluated the sample size required to achieve the same effect size as the GLM in the Tanzania case study (Cohen, 1998; Campely, 2020).

2.4 Results

From step one, 111 candidate wellbeing indicators were identified by focus groups and literature review. These were grouped into 62 questions included in the wellbeing questionnaire (step 2). Removal of indicators with little variation reduced this list to 56 indicators (step 3; Supplementary material S6). Removal of correlated indicators reduced the list further to 30 indicators (step 4, Supplementary material S5). Statistical modelling (step 5) then reduced the list to 17 indicators. The qualitative validation step reintroduced two indicators, resulting in a final list of 19 indicators (Table 2.1).

Table 2.1. Final list of wellbeing indicators. Likert scales are all on a scale from very low to very high unless stated. [1] Continuous to categorical transformation,

[2] log transformation, [3] Site-specific indicator, [4] Gender-specific indicator. %D = percentage deviance explained in HWI by each indicator in the final GLM. r^2 = proportion of variation explained by Pearson correlation with HWI.

Indicator	Description	Mean	Range	Variation explained %D (r^2)
Material				
Financial savings [1]	Ordinal categories (0, 1 – 99,999, >100,000 TZS)	1.05	0-2	0.54 (0.45)
Household wall materials	Ordinal categories (1=mud, 2=mud bricks 3=concrete bricks, 4=plastered)	2.38	1-4	0.10 (0.08)
Household assets	Integer from 1-7 of total of household assets composed of electricity, solar light, television, radio, phone, plough, tractor	2.87	1-7	0.02 (0.17)
Banking	Use of formal banking facilities (yes/no)	0.31	0-1	0.20 (0.12)
Water access [1]	Ordinal categories for if need to walk to collect water (Never, only in the dry season, in both dry and wet season)	1.02	0-2	0.48 (0.05)
Land [2]	Total area owned (acres)	8.02	0-104	0.34 (0.14)
Livestock [1]	Ordinal categories for most valuable livestock owned (none, chickens, goats/pigs, cattle)	1.20	0-4	< 0.01 (0.18)
Health				
Sickness [2]	Number of days too unwell to work in last year	10.90	0-150	3.31 (0.06)
Health insurance [3]	Binary response (yes/no)	0.31	0-1	2.54 (0.26)
Social relations				
Lending of resources [4]	Binary response stating whether money or land was lent in last year (yes/no)	0.65	0-1	1.75 (0.27)
Recognition in the village [4]	Perception of how much voice heard in community decision making. Likert scale	2.96	1-5	0.35 (0.18)
Security				

Provision for dependents	Likert scale indicating perceived ability to provide for dependents	3.76	1-5	2.06 (0.33)
Provision for self in old age	Likert scale indicating perceived ability to provide for oneself in old age	3.31	1-5	1.81 (0.42)
Number of livelihoods	Total of different livelihood activities undertaken by the interviewee	4.14	1-7	2.07 (0.24)
Theft security	Likert scale indicating perception of security from theft	3.25	1-5	1.65 (0.07)
Freedom				
Livelihood satisfaction	Likert scale indicating satisfaction with livelihood opportunities	2.57	1-5	0.48 (0.05)
Forest access [3]	Likert scale indicating satisfaction with access to forest resources	1.70	1-5	0.41 (0.06)
Education	Ordinal categories for highest level completed (from no formal education to university)	4.50	1-8	0.34 (0.25)
Overall quality of life	Likert scale indicating overall life satisfaction considering all questions asked	3.24	1-5	0.89 (0.22)

2.4.1. Complexity-simplicity

The final indicator list (Table 2.1) explained 91 % deviance in the holistic wellbeing index. Stepwise GLM reduction of the indicator list from step four (30 indicators) to step five (17 indicators) led to reduction in AIC from -397 to -412 (ΔAIC = 15), with only marginal loss of deviance (95% to 91%). Therefore, the reduced indicator list was more parsimonious. Sequential removal of indicators revealed a pattern of increasing loss of deviance explained per indicator removed (Figure 2.4). The loss of deviance explained from 20 to 15 indicators was only 1%, but a 15% loss of deviance explained was observed between 10 and 5 indicators.

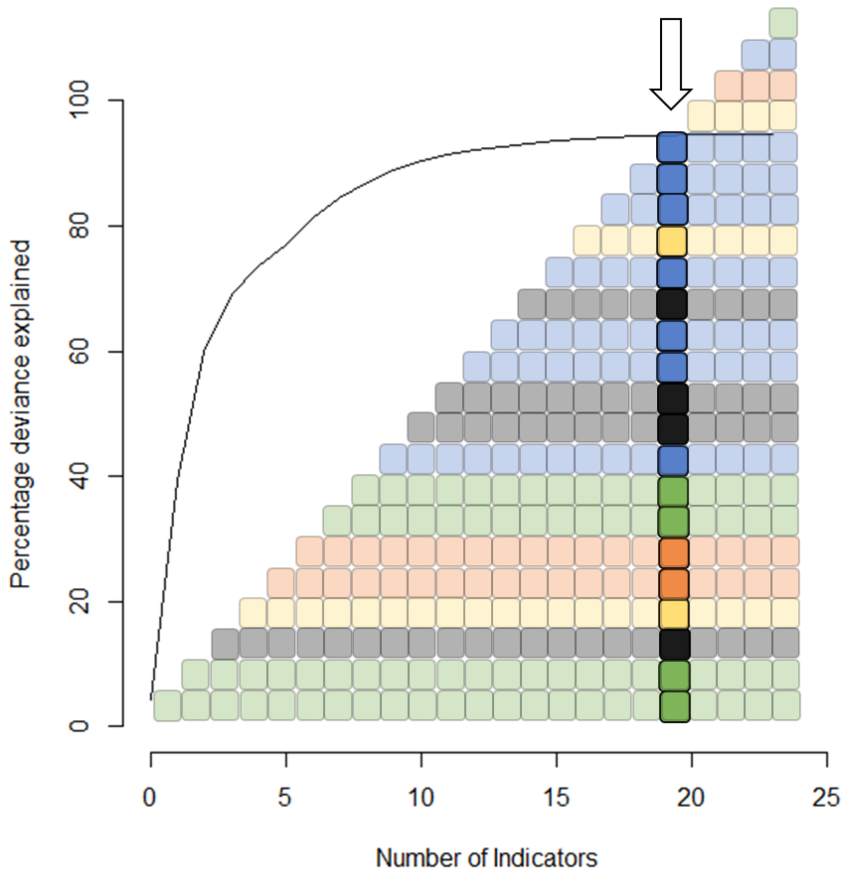


Figure 2.4. Loss of variation (deviance) explained in the holistic wellbeing index by each indicator list, when sequentially removing the indicator explaining the least variation in the GLM (black line). Arrow identifies the final indicator list selected by WISP. Boxes represent an indicator coloured by the five wellbeing domains (blue = material wellbeing, orange = health, green = security, yellow = social relations, black = freedom).

2.4.2. Integrating perspectives

In both the candidate and final indicator lists, material wellbeing had the largest percentage of indicators (Figure 2.5). The final list had a slightly lower percentage of material indicators and slightly higher percentage of security and freedom indicators (Figure 2.5).

The validation checks (step 6) revealed that the social relations domain of wellbeing had been reduced to a single indicator. Therefore, an additional indicator, recognition in the village, was reinserted to improve balance between domains. We identified one disparity between prioritisation of indicators through quantitative analyses versus qualitative assessment of local priorities. Livestock ownership was identified as a local priority indicator. However, this indicator was removed by statistical stepwise selection (step 5). To ensure WISP integrated top-down and bottom-up perspectives, livestock ownership was reinserted in the finalised list (step 6).

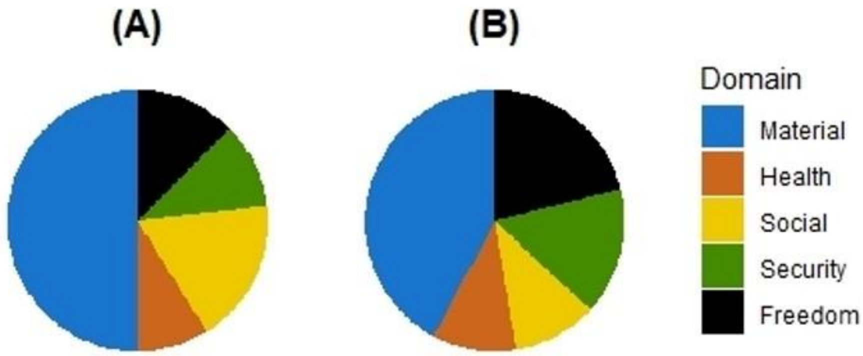


Figure 2.5. Percentage representation of the five wellbeing domains in the candidate (A) and final (B) list of wellbeing indicators.

Whether candidate indicators were considered universally important across the study area, or site, or gender-specific, depended on the wellbeing domain. Indicators relating to the material domain of wellbeing were strongly corroborated between focus groups. Plant-based agriculture was the dominant livelihood at both sites and so the area of agricultural land owned was identified early on in all focus groups. While other material indicators, such as household building materials and livestock ownership were also universally identified in all focus groups. Some gender-specific differences were noted in terms of the social relations domain of wellbeing. All-female focus groups in both villages identified the importance of mutual reliance within communities, defined as the ability to lend resources. However, this informal interdependence was not discussed in all-male focus groups. Instead discussion highlighted the importance of recognition from peers in the village indicated by a felt sense that their voice was heard in village meetings. These gender-specific indicators were included in the final indicator list (Table 2.1).

Finally, some differences were noted between villages in relation to the location and extent of remoteness and self-reliance versus connectedness of villages to urban centres. In Udekwa, the village located further from major transport routes and urban centres, candidate indicators within the health and freedom domains of wellbeing included knowledge and access to local medicines and producing enough food to eat. However, in Mang'ula B, the village located close to a major road with direct transport links to urban centres, health insurance and access to formal banking facilities were also discussed by this village. Site-specific indicators were retained in the final indicator list (Table 2.1).

2.4.3 Practical utility

Power analyses showed that a sample size of 13 questionnaire repeats would be needed to evaluate correlations between indicators in step 4 and 33 replicates to provide sufficient power to undertake statistical modelling in step 5.

2.5 Discussion

2.5.1. Complexity-simplicity

WISP resulted in relatively little loss of information concerning variation in wellbeing, while significantly reducing the number of indicators. Reintroduction of local priority indicators contributed a small amount of statistical redundancy, exemplified by the minimal reduction in deviance explained when additional indicators were removed beyond the final list (Figure. 4). However it is vital to characterise wellbeing in accordance with place-based values to avoid unintended harmful consequences of policies for local residents (Sterling et al., 2017). We therefore suggest that WISP remains sufficiently conservative to retain a rich decription of wellbeing that balances the trade-off between complex local realities and statistical parsimony.

A comparable alternative social assessment approach is the Basic Necessities Survey (BNS), which in previous studies has identified between 20 and 25 local indicators (Schreckenberg et al., 2010; Davies, 2007). BNS creates an index of poverty, providing a narrower conception of wellbeing focused on the material domain. We therefore conclude that in terms of the complexity-simplicity problem, the protocol performed well in relation to BNS, creating a more concise list of indicators, yet more representative of multidimensional wellbeing.

2.5.2. Integrating perspectives

Mixed methods provide an opportunity to identify and address tensions between qualitative and quantitative methods (Denscombe, 2008). In our study the statistically led simplification step reduced the number of candidate indicators and in the process removed livestock ownership, which had been frequently mentioned in focus group discussions. This highlights a tension between external and ecological validity and emphasises the importance of integrating perspectives in order to navigate this trade-off (Figure 1). Indeed, we suggest that there is an inherent tension in translating a felt sense of wellbeing into numeric values with potential to compromise ecological validity through over reliance on quantitative approaches. However in our Tanzanian example, the protocol helped to reconcile this tension through step six, which ensured the final indicator list was aligned to local priorities.

Future users might consider adapting the protocol steps and indicator inclusion criteria depending on the intended application. For example, if used to evaluate the impact of a specific intervention, such as a water security program, it might be important to use focus groups to explore locally relevant indicators for the intervention, such as clean water access and prioritise retention of intervention-specific indicators (Jensen and Wu, 2018). Alternatively, if used to evaluate change in wellbeing through time (Sayer, Campbell and Petheram, 2007), then researchers might choose to be more conservative in retaining indicators that have little variation at the time of the first survey (step 3), but for which variation is expected to increase, either as a result of increasing inequality (Martin et al., 2014), or as a result of the intervention targeting a subsection of the population. Where there is doubt, we recommend using clear hypotheses to justify additional inclusions and using locally stated priorities and ecological validity as a guiding principle to determine inclusion.

Our observation that the material domain of wellbeing comprised more indicators than other domains concurs with observations from other developing and developed countries. Namely, material wellbeing may be distributed among a number of different sources (DFID, 2000, Goodwin, 2003). In our study, the dominant sources of material wellbeing were financial, land and livestock. This pattern of spreading material wellbeing among a number of capital sources can be interpreted as a strategy for enhancing the resilience of individuals; the ability to cope with and overcome shocks (Folke et al., 2002; Walker et al., 2006). As such, we suggest that rich descriptions of material wellbeing that include multiple indicators are also applicable for evaluating the related concept of resilience (Gunderson and Holling, 2002; Hoque, Quinn and Sallu, 2017).

Various cultural, socio-economic and livelihood characteristics of individuals influence which sources of material wellbeing are invested in (Sunderlin et al., 2005; Miller and Hajjar, 2019). For example, pastoralists may invest far more in livestock compared to individuals whose livelihood depends more on crop-based agriculture and invest more in land. Therefore, to accurately compare material wellbeing in heterogenous communities we suggest that a larger number of indicators may be needed for this domain than others to provide an accurate summative measure that accounts for differential capital investment patterns.

An alternative explanation for the large number of material indicators relates to the methods used in this study. Contrasting qualitative and quantitative methods are better suited towards identifying different social phenomena (Braun and Clarke, 2013; Bull et al., 2015). For example, ethnographic approaches are tailored to the identification of in-depth personal narratives and socially constructed themes (Atkinson and Silverman, 1997). In contrast, WISP uses a more rapid approach to identifying candidate wellbeing indicators . As a result, the less tangible aspects of wellbeing, such as social relations, were relatively under-represented among candidate indicators. Instead objective indicators that were more easily observable and articulated were more represented (Schreckenberg et al., 2010). Future adoption of ethnographic approaches to complement the protocol might facilitate exploration of the less tangible aspects of wellbeing. Another approach to promote broader representation of indicators would be to structure focus group discussion topics around the five wellbeing domains. However, we preferred a more open questioning style to encourage study participants to lead discussions, rather than be confined by wellbeing theory, thereby promoting ecological validity.

2.5.3. Practical utility

The introduction of a statistical techniques to the process of selecting wellbeing indicators may cause a technical challenge for researchers with more qualitative backgrounds. However, there is also potential for greater expansion of this element of WISP. Future studies might consider multi-model averaging approaches (Burnham and Anderson, 2002), to determine additional more subtle contributions to wellbeing from indicators dropped from our models. However, step 4 and Figure 4 show that any changes to the indicators selected would add only marginal difference to the variation explained in wellbeing. Therefore we prefer our more simple and accessible statistical approach.

Decisions regarding appropriate sample sizes of villages, the diversity of actors and questionnaire replicates in future studies will depend in part on a practical trade-off between exhaustive sampling and resource constraints. Here we sampled two villages, though we stress that this figure should be used as a guide only and larger scale studies may require additional sampling. The Poverty Environment Network guidance suggests that questionnaire pre-testing should include seven draft questionnaire trials before commencing the main survey (PEN, 2007). However, in our Tanzanian example we estimated that 33 questionnaire repeats were required for statistical analyses. Therefore we recommend a conservative minimum of 40 questionnaires be undertaken in future applications of the protocol to allow for context specific differences in wellbeing indicators. We suggest that increased pre-testing investment is a necessary consequence of moving away from simpler conceptions of wellbeing or poverty, towards robust measurement of a more complex conception of multidimensional wellbeing. As the number of indicators increases, there will be greater potential for correlation and violations to statistical assumptions. Consequently, the introduced orthogonality checks are necessary to promote external validity and robust analysis. An additional benefit of investing in questionnaire simplification at the beginning of a wellbeing assessment is that this shortens the questionnaire; in our case study to less than a third of its original length. This is more efficient of time, resources and reduces respondent fatigue during survey implementation (Trochim, 2006).

2.6. Conclusions

We have demonstrated that WISP makes progress in addressing three interrelated challenges to measuring wellbeing in different local contexts. We therefore recommend the protocol as practical and statistically validated step-by-step guidance to support the design of multidimensional wellbeing assessments, maintaining coherence between universal theory and local realities. In this way, the protocol contributes to a research agenda seeking to support policy makers in advancing a holistic notion of social progress. Future contributions to this field might explore how to integrate local and national scale wellbeing assessments. Also the integration of local perspectives with actors operating at larger scales, such as national policy makers in order to advance transparent and equitable policy decision-making.

2.7. Acknowledgements

Research was funded by the Economic and Social Research Council. Many thanks to Elizabeth Kombo, Herman Michael Lyatuu and Revocatus Lauriean Rwezaula whose help as research assistant and logistics support, respectively, made a huge difference to the project. Thanks to Andy Wright (madebyawdesign.com) for figure design support. We are grateful for suggestions from Dr Peter Howley and Dr Emily Woodhouse that aided the refinement of this study and two anonymous reviewers.

Chapter three: Certified community forests positively impact human wellbeing and conservation effectiveness and improve the performance of nearby national protected areas

Manuscript published:

Loveridge, R., Sallu, S.M., Pfeifer, M., Oldekop, J.A., Mgaya, M., da Silva, D.A., Latham, J., Platts, P.J. and Marshall, A.R. (2021). Certified community forests positively impact human wellbeing and conservation effectiveness and improve the performance of nearby national protected areas. *Conservation Letters*, p.e12831

Authors

Robin Loveridge[1,2*], Susannah M Sallu[3], Marion Pfeifer[4], Johan A Oldekop[5], Mercy Mgaya[6], Daniel A da Silva[7], Julia Latham[8], Philip J Platts[1], Andrew R Marshall[1,9,10]

Author affiliations

[1] Department of Environment and Geography, University of York, York, UK

[2] The Biodiversity Consultancy, Cambridge, UK

[3] School of Earth and Environment, University of Leeds, Leeds, UK4

[4] School of Natural and Environmental Sciences, TROPS lab, Newcastle University, UK

[5] Global Development Institute, The University of Manchester, UK

[6] Department of Geography, University of Dar es Salaam, Tanzania

[7] Regional University of Blumenau, Blumenau, São Paulo St, Brazil

[8] Conservation Insight, UK

[9] Forest Research Institute, University of the Sunshine Coast, Australia

[10] Flamingo Land Ltd., Kirby Misperton, North Yorkshire, UK

Key words: community forest, conservation effectiveness, environmental justice, forest certification, governance, impact evaluation, integrated landscape management, matching, protected area, wellbeing

3.1 Abstract

Community Forests (CFs) aim to improve human wellbeing and conservation effectiveness, though their performance remains contested. A recent innovation in protected area (PA) governance is to combine CFs with forest certification. We assess (1) the impact of certified CFs on wellbeing and conservation effectiveness; (2) gender inequality and elite capture; (3) interaction effects with neighbouring national PAs. We used a novel approach that integrates field data consisting of locally identified indicators representative of multidimensional wellbeing, with remotely sensed data on conservation effectiveness and statistical matching to improve causal inference. We found that CFs positively impacted wellbeing, conservation effectiveness and reduced gender inequality, though did not reduce elite capture. We also detected positive interaction effects between certified CFs and neighbouring national PAs. Our findings suggest that locating contrasting local and national PA governance approaches next to each other may help to maximise wellbeing and conservation benefits within complex multi-use landscapes.

3.2 Introduction

Protected Areas (PAs) are a cornerstone of efforts to conserve forests and attention is also increasingly being paid to how PAs can improve the wellbeing of adjacent communities (Naidoo et al., 2019b). In response to these twin challenges, diverse and increasingly complex PA governance arrangements have been trialled, though their effectiveness remains contested (Gavin et al., 2018). Rigorous impact evaluations are needed to understand and improve PA performance (Baylis et al., 2016). However, robust evaluations remain rare and tend not to fully accounting for diverse wellbeing impacts (Woodhouse, et al., 2015). Evaluations generally prioritise externally defined, economic indicators. Yet, these often ignore local priorities and non-economic wellbeing impacts (de Lange et al., 2016), such as how limiting forest access may impact the ability of communities to cope in times of scarcity (Atela et al., 2015). To more comprehensively account for PA impacts, more locally-relevant measures are needed (Sterling et al., 2017).

Over the past 40 years there have been three distinct trends in forest PA governance with complexity of governance arrangements increasing over time (Figure 1). Initially, formal PAs tended to be centrally governed, excluding local participation. However recognition of the negative impacts of this approach on neighbouring communities led towards decentralisation and the establishment of Community Forests (CFs; Agrawal et al., 2008). The premise underlying this trend is that community participation will synergistically improve both the living standards of local communities and biodiversity conservation (Persha et al., 2011). Initially, this focused on legal devolution of forest tenure. However in much of Sub-Saharan Africa and across the tropics, land tenure remained with central governments with local communities gaining only access rights and receiving few economic benefits (Sunderlin, 2006). Alternatively, power was vested in local elites, resulting in increased inequality and exclusion of women from governance institutions (Agarwal, 2009; Magessa, et al., 2020), while also failing to halt forest degradation (Rasolofoson et al., 2015).

To overcome these shortcomings a second generation of CFs emerged that integrated market-based approaches to promote revenue generation and aim for equitable benefit distribution (Figure 3.2). Though the effectiveness of market-based approaches to achieve distributional equity has been limited (Pascual et al., 2014). Here we focus on the combination of CFs with Forest Stewardship Council (FSC) certification, which is the most widespread certification scheme in the tropics and aims to incentivise sustainable forest use and offset financial costs by attracting additional timber traders (Burivalova et al., 2017). Gender equality is also promoted through requirements for wage equality and equal employment opportunities (FSC, 2018). Despite the potential for certified CFs to deliver synergistic benefits for human wellbeing and conservation, to date no impact evaluation has reported a combined assessment of multidimensional human wellbeing and conservation effectiveness of CFs with forest certification.

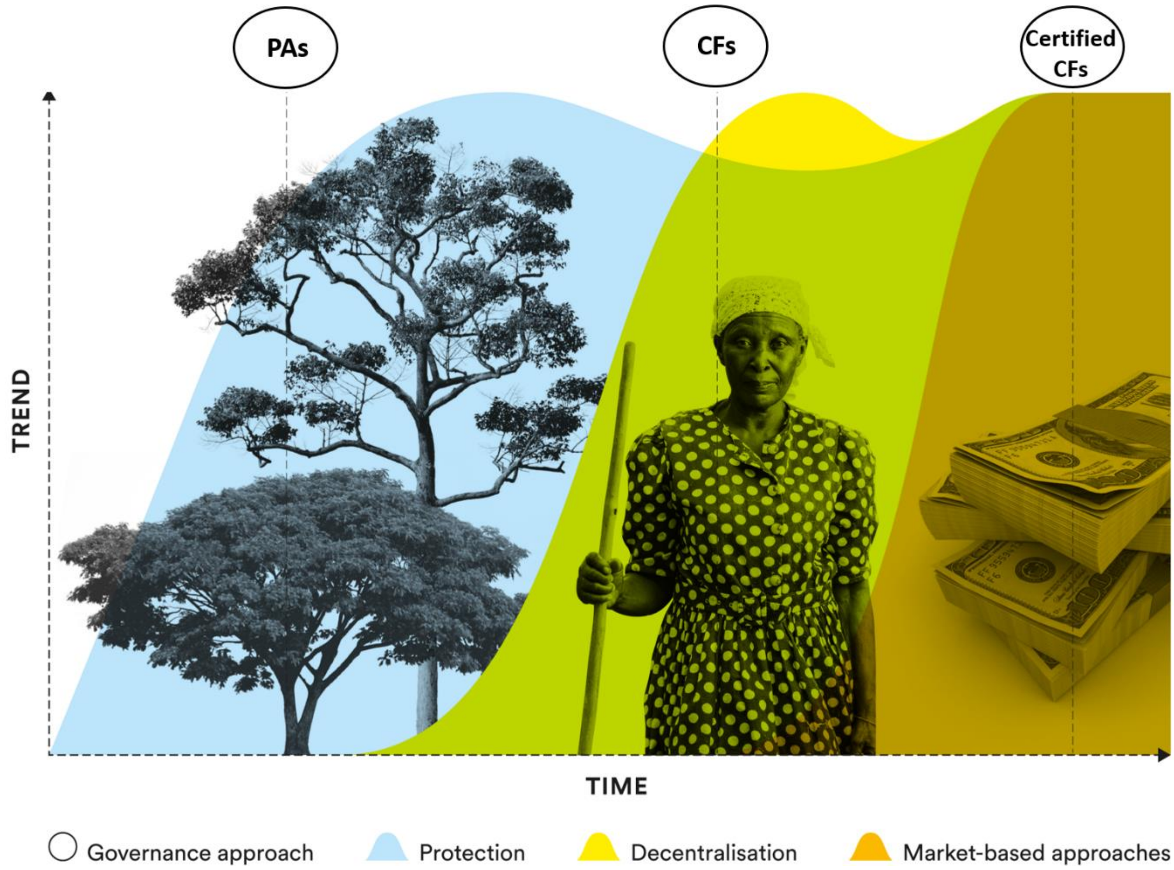


Figure 3.1. Major trends in forest protected area governance through time showing the emergence of community forests (CFs) and certified CFs alongside strictly protected areas (PAs). Trend lines and colour layering show how new trends have added to the governance arrangements of earlier trends creating additional governance complexity.

The expansion and increasing connectivity between PAs (Santini et al., 2016), combined with complex governance arrangements creates a need for sophisticated impact evaluation designs that account for spatial and governance interactions (Barton et al., 2017; Baylis et al., 2016). Interactions, *sensu* spillovers, can have a positive effect on conservation outcomes as neighbouring PAs may buffer each other through complementary patrol effort that benefits both PAs (DeFries et al., 2005). Alternatively, partitioning of forest resources between national and local actors and may strengthen the resource claims of marginalised actors (Kaimowitz, 2003), resolving conflicts between local and national actors and promoting sustainable use. Alternatively interactions may be negative, *sensu* 'leakage', displacing deforestation to neighbouring areas, or increasing conflicts which can have a dramatic effect on observed performance (Joppa & Pfaff, 2010). Yet most studies on PA effectiveness focus on single governance approaches and very few explore the relative effects of co-occuring interventions (Sims & Alix-Garcia, 2017).

We aim to determine whether recent trends in PA governance can overcome past shortcomings of PAs to benefit both forest conservation and human wellbeing. We use a PA dominated landscape typical of complex PA governance arrangements in the developing tropics to assess (1) the impact of certified CFs on multidimensional human wellbeing and conservation effectiveness, (2) how certified CFs impact elite capture of benefits and gender inequality, (3) interaction effects with neighbouring PAs. We use a novel approach for measuring a locally-grounded conception of multidimensional human wellbeing to provide a comprehensive accounting of diverse PA impacts. We integrate this fine-scale field data with remotely-sensed measures of forest conservation and statistical matching techniques to enhance causal inference. Finally, our impact evaluation design incorporates both the layering of PA governance approaches and interactions between PAs. Therefore, our evaluation makes progress in representing the complex reality of modern protected area governance.

Tanzania provides a good test case of complex PA governance arrangements. PAs cover 38% of the country, of which 2.5 million ha have been transferred from central government to local communities in the form of CFs (World Bank, 2018). In 2009 communities in Tanzania became the first in Africa to obtain a FSC certification for CFs (FSC, 2018). However, the national government, like governments elsewhere in Africa and Asia (Basnyat et al., 2018), is exploring curtailing local governance. An assessment of the performance of recent innovations in CF governance is therefore urgently needed to inform the future of CFs.

3.3 Methods

3.3.1 Site information

Our focus is on a PA dominated landscape in southeast Tanzania where certified CFs have been established on village land in close proximity to existing Protection National Forest Reserves (NFRs; Figure 3.2). NFRs are administered by the national government and timber extraction is not permitted. In contrast, CFs allow communities to extract timber and non-timber products and retain timber revenues. The national FSC standard requires additional management oversight in the form of a no harvest zone covering 10% of the CF and an externally audited five-year review of sustainable harvesting plans (FSC, 2018). Communities are supported to meet these requirements by district government and an NGO, Mpingo Conservation & Development Initiative (MCDI).

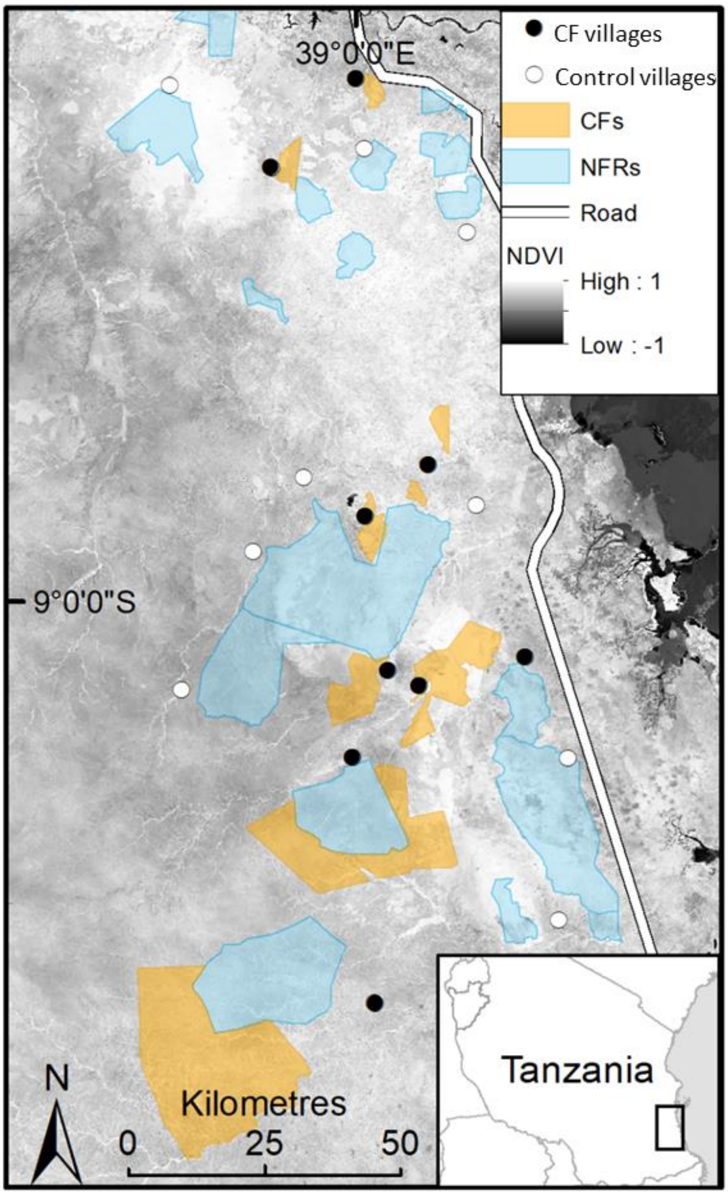


Figure 3.2. Study landscape in Lindi and Pwani regions of Tanzania detailing the 19 identified study villages and 27 protected areas. The main highway stretching north to south along the east coast linking the capital city of Dar es Salaam to the regional capital of Lindi.

3.3.2 Response variables

We measured multidimensional human wellbeing through a Human Wellbeing Index (HWI) using the Wellbeing Indictor Selection Protocol to identify locally appropriate indicators through consultation with target communities (Loveridge et al., 2020). Data were collected through a field questionnaire survey in 2019. The survey was composed of 25 indicators representative of five domains of wellbeing (1) material, (2) health, (3) social relations, (4) security; and (5) freedom of choice and action (Narayan et al., 2000). An index of each domain was calculated as the mean of indicators from each domain standardised between 0-1 and the HWI was then calculated as the mean of domain indices, thereby weighting wellbeing domains equally following principles of the Human Development Index. The unit of analysis was individual questionnaire respondents, since human wellbeing contains subjective components that cannot be generalised across households (Supporting Information).

We calculated change in Normalised Difference Vegetation Index (NDVI) between 2014 and 2019 at a resolution of 30 x 30 m pixels across the study area as a standardised proxy measure of forest conservation effectiveness (Supporting information). NDVI correlates with ground vegetation biomass and productivity under low to medium vegetation density conditions such as the Miombo woodlands studied here (Oindo & Skidmore, 2010; Pfeifer et al., 2016). We define positive conservation outcomes as all cases where NDVI change was positive as this provides evidence of forest recovery.

To assess social interaction effects between PAs, specifically the presence village conflicts with neighbouring NFRs, we included a Likert scale question in the wellbeing survey, with responses varying from 1 (there are big conflicts) to 5 (the relationship is very good).

3.3.3 Research design

PAs tend to be located in remote areas, systematically different from the wider landscape, which has potential to bias impact estimates. We used a quasi-experimental approach to estimate the impact of certified CFs on human wellbeing and NDVI change. Quasi-experimental methods increase the causal inference from observational data by emulating randomised controlled trials. We used statistical matching in the 'Matchit' package in R (Ho et al., 2007) to emulate randomisation by matching treatment units to control units with similar distributions of confounding variables. The key assumption helping to establish causal inference in statistical matching is that once treatment and control units have been matched, treatment allocation is close to random (Stuart, 2010). Control units thus represent the counterfactual situation, i.e. what would have happened in the absence of the intervention (Schleicher et al., 2020). We then define impact as the difference in human wellbeing and conservation effectiveness between matched treatment and control units.

In the study landscape all CFs adhere to both CF legislation and FSC certification requirements and were within 10 km of NFRs. Therefore, we assessed the impact of the certified CFs by matching 9 villages that had certified CFs established for at least five years (range: 2009-2014) to 10 control villages without CFs or forest certification, but also within 10 km of NFRs (Table 1). Due to the small sample size and absence of fine-scale baseline data on wellbeing prior to undertaking the study, we implemented a two-stage matching process to ensure identification of an appropriately matched sample. First, village matching was based on confounding biophysical and socio-economic variables (Supporting information). The quality of identified matches was then verified by an expert panel of local actors (Mitchell et al., 2018; Supporting information). Stratified random sampling within matched villages based on gender, local elite status (identified as being a village government representative) and wealth category was then used to select and interview 955 people. We assessed the impact of certified CFs on conservation effectiveness by a further round of matching that built on the village-scale matching to select treatment pixels from certified CFs and control pixels from within the village land of matched control villages.

Previous assessments suggest that within a threshold of 10km, PAs exert social and ecological impacts (Naidoo et al., 2019b). The identification strategy allowed us to test for interactions between certified CFs and NFRs because we specified that all treatment and control villages were within 10 km of NFRs (Figure 2; Supporting Information). We were able to assess (1) the impact of certified CFs on conflicts between communities and NFR managers by comparing matched villages with and without CFs; (2) the conservation impact of certified CFs on NFRs by comparing matched samples of NFR pixels within 10 km of CFs to NFR pixels more than 10km from CFs (Table 3.1).

Table 3.1. Summary of matching comparisons

Impact test (response)	Matching unit (sample size)	Treatment	Control
Certified CFs on human wellbeing (HWI of individual respondents)	Villages (19)	Villages which have established a CF and maintained FSC certification for at least five years, and are within 10 km of NFRs	Villages without CFs or forest certification, but are within 10 km of NFRs
Certified CFs on relations with NFR managers (relations as perceived by individual respondents)	Villages (19)	Villages which have established a CF and maintained FSC certification for at least five years, and are within 10 km of NFRs	Villages without CFs or forest certification, but are within 10 km of NFRs
Certified CFs (NDVI change)	Forest pixels (8,154)	Pixels of certified CFs	Pixels from village land of matched villages
Certified CFs on NFRs (NDVI change)	Forest pixels (8,050)	Pixels of NFRs less than 10 km from certified CFs	Pixels of NFRs more than 10 km from certified CFs

3.3.4 Analysis

We used linear mixed effects models to test for differences in HWI, each wellbeing domain and NDVI change between matched treatment and control groups (Supporting information). To test whether certified CFs impacted gender inequality and inequality between local elites and non-elites, we included fixed effects for (1) treatment, (2) gender, (3) villager elite status and (4) interaction terms between treatment and both gender and elite status. To control for spatial autocorrelation, we included random effects for village identity. To control for any residual imbalances in the distributions of confounding variables between treatment and control groups we compared models with and without orthogonal sets of confounding variables (Ho, et al., 2007). For wellbeing analyses we included data on confounding variables based on household location collected by GPS during questionnaire surveys. This was important to account for any differences in the distributions of confounding variables between treatment and control groups caused by the small sample size and coarse spatial scale of matching between villages.

We used the MuMIn package in R to undertake model averaging to select preferred models (Supporting Information). To further explore conservation interaction effects we used mixed effects models of NDVI change with separate samples of CFs and NFRs, including distance to the contrasting PA governance approach as a proxy variable indicating exposure to interaction effects. To assess social interaction effects, we used ordinal logistic regression to model the conflict response variable, incorporating village nesting (Supporting information).

3.4 Results

Certified CFs located close to NFRs had a positive impact on human wellbeing (estimate = 0.026, *SE* = 0.012, *P* = 0.035) and NDVI change (estimate = 0.047 *SE* = 0.024, *P* = 0.047; Figure 3.3). Across the study area NDVI remained stable between 2014 (mean NDVI = 0.66, *SD* = 0.11) and 2019 (mean NDVI = 0.67, *SD* = 0.13). Negative NDVI change occurred close to village centres and NDVI change was positively associated with increasing distance from villages (estimate = 0.036, *SE* = 0.002, *P* = <0.001).

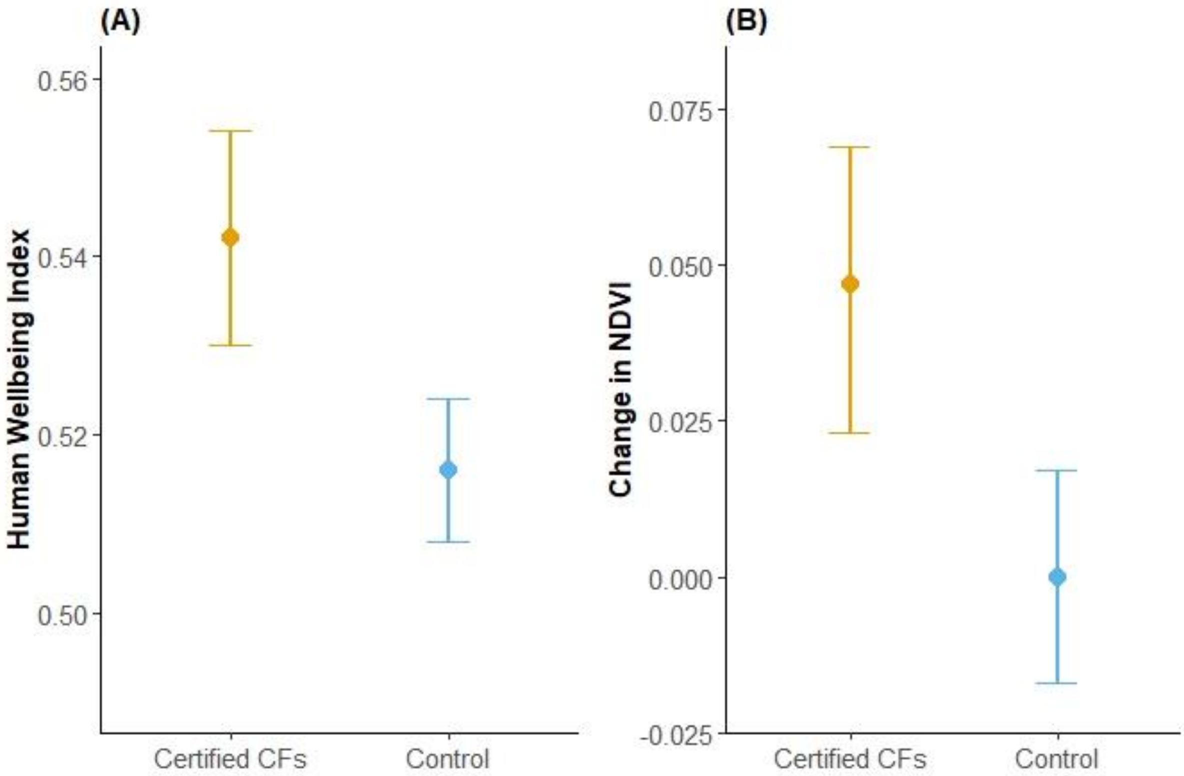


Figure 3.3. Impact of Certified CFs on (A) wellbeing and (B) change in NDVI. Error bars show model standard error estimates.

Across the study area, men experienced higher wellbeing than women (Fig. 3.4A). However, gender inequality in wellbeing was reduced in villages with certified CFs (Figure 3.4A). Certified CFs did not influence the difference in wellbeing between village elites and non-elites and the impact of certified CFs was not uniform across wellbeing domains. The health, security and freedom domains of wellbeing were improved (Figure 3.4B). Specifically, certified CF communities had higher mean indicator scores for access to clean water, food security and education facilities (Supporting information Table S8). But domains of material wellbeing and social relations were not different between villages with and without certified CFs.

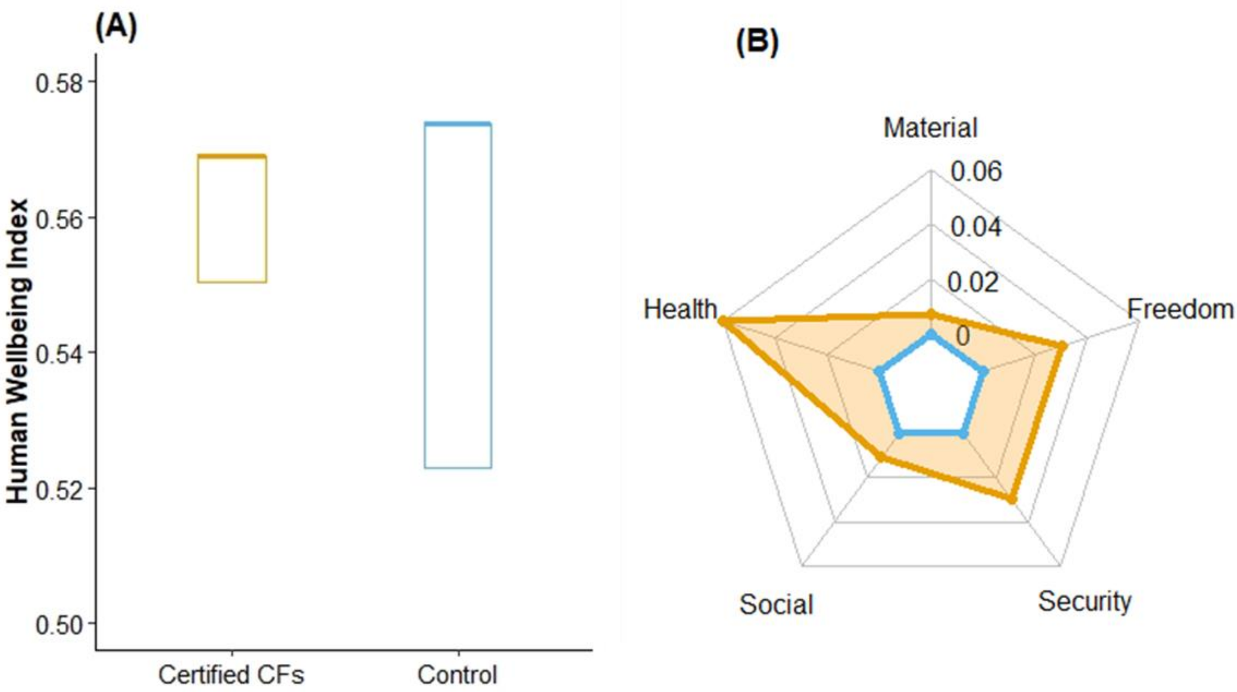


Figure 3.4. (A) Difference in gender inequality between villages with certified CFs and villages without CFs. Thick and thin horizontal lines represent estimated wellbeing of men and women respectively; Estimated change in gender inequality = - 0.03, *SE* = 0.01, *P* = 0.008. (B) Estimated difference in the wellbeing scores for the five domains of wellbeing between Certified CFs (orange) and villages without CFs (blue); Estimated impact of CFs on Material wellbeing = 0.007, *SE* = 0.017; *P* = 0.697; Health = 0.060, *SE* = 0.025, *P* = 0.015; Social relations = 0.011, *SE* = 0.035, *P* = 0.744; Security = 0.030, *SE* = 0.013, *P* = 0.024; Freedom = 0.030, *SE* = 0.013, *P* = 0.024.

NFRs within 10 km of certified CFs had positive NDVI change compared to matched NFRs more than 10 km from CFs (Figure 5A). In both NFRs and Certified CFs, NDVI change was more positive closer to the contrasting governance approach and negatively associated with increasing distance (Figure 5B and C). Communities with Certified CFs positively impacted community relations with neighbouring NFR managers by reducing conflicts (estimate = 1.212, *SE* = 0.204, *P* = <0.001).

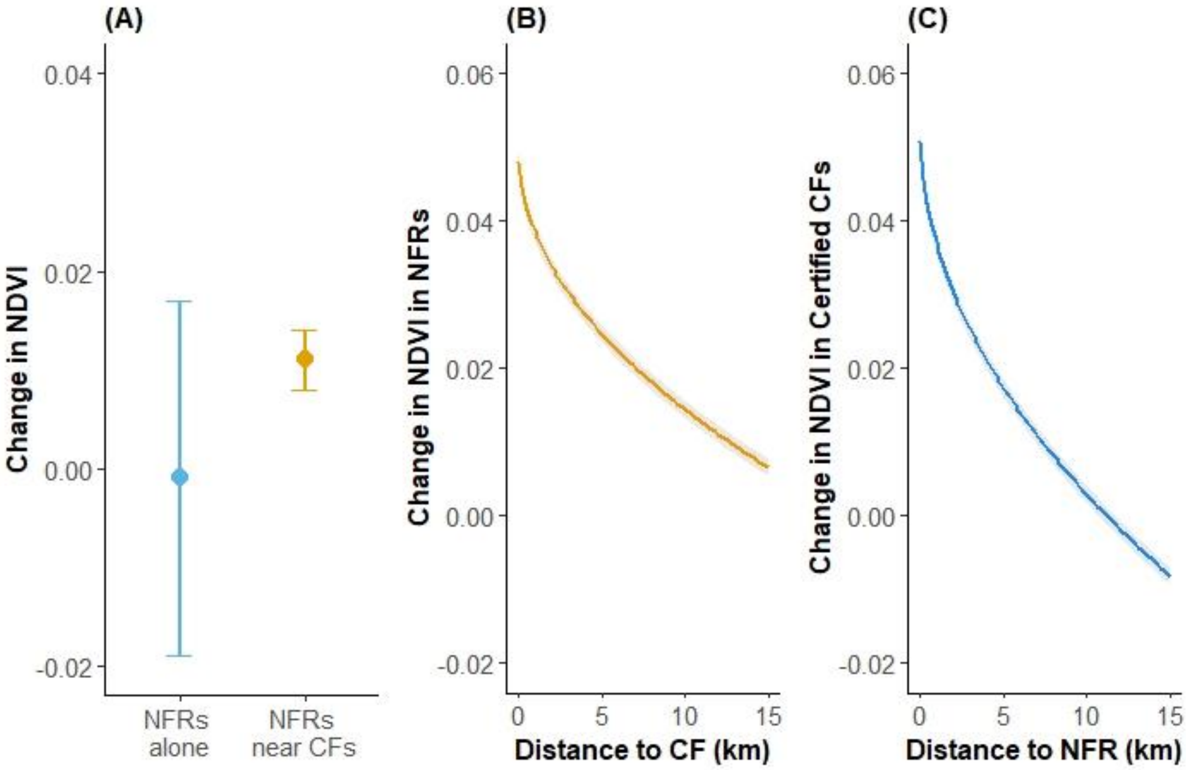


Figure 3.5. (A) The impact on NDVI change of having certified CFs within 10 km of NFRs = 0.011, *SE* = 0.002, *P* = <0.001. How NDVI change in (B) NFRs is affected by distance to CFs = - 0.013, *SE* = <0.001, *P* < 0.001), and (C) in CFs by distance to NFRs = - 0.015, *SE* = < 0.001, *P* < 0.001) with grey shading showing model estimated standard errors (Supporting Information Tables S3).

3.5 Discussion

Our findings suggest that certified CFs (1) positively impact both human wellbeing and conservation effectiveness, (2) reduce gender inequality, though not elite capture, and (3) improve conservation and social interaction effects with neighbouring NFRs.

We found evidence of heterogenous social impacts with health, security and freedom domains of wellbeing positively impacted, but no observed impacts on the material and social relations domains. This is in accordance with global analyses of CFs, which suggest that trade-offs between outcomes are more common than uniform benefits (Hajjar et al., 2020), and shared community benefits are more common than individual financial benefits

(Burivalova et al., 2017). In the CFs studied, a hybrid form of revenue distribution was employed that includes both egalitarian and merit-based distribution (McDermott, et al., 2013). The largest proportion of funds (50% egalitarian distribution) is spent on village development projects that all villagers should benefit from, such as improvements in village health and education provision. Whereas 40% of timber revenue is paid directly to forest managers (merit-based; FSC, 2019) and the remaining paid to NGO and governmental agencies for technical support. This form of benefit distribution is distinct from CF programmes in other countries such as Nepal and Kenya, where timber revenue is distributed through direct financial disbursements to households (Braber et al., 2018; Walpole & Leader-Williams, 2001). This divergence in governance and observed wellbeing impacts likely represents a trade-off between individual, material wellbeing and community, non-material wellbeing. While the observed nonmaterial benefits highlight the value of using broadly framed social impact evaluations to capture diverse impacts of PAs on human wellbeing (Woodhouse et al., 2015).

Unlike findings from CFs elsewhere in Tanzania and globally (Magessa, Wynne-jones and Hockley, 2020; Hajjar et al., 2020), we did not find evidence of increasing inequality between local elites and non-elites. However, we found that inequality was not reduced by certified CFs, suggesting that a greater emphasis on equitable benefit sharing may be needed if CFs are to contribute to overcoming structural inequalities. We found evidence of increased gender equality, likely due to the form of CF governance studied here. Unlike other regions of Tanzania, CFs were augmented by FSC certification. FSC transparency requirements enable greater public scrutiny, supporting democratic decision-making and fairer distribution of benefits and gender equality is specifically promoted through gender-specific requirements for participation in training and employment opportunities (FSC, 2018; Martin et al., 2019). The cost of delivering these equity benefits is that certification programmes have significant documentation requirements that present technical capacity challenges for non-specialists (Burivalova et al., 2017). To overcome the capacity gap, in this case an NGO, MCDI supports community institutions and this long-term governance capacity building is likely a necessary ingredient for success.

The study of PA interaction effects is in its infancy (Baylis et al., 2016; Sims & Alix-Garcia, 2017). Our results indicate the presence of a buffering effect between PAs, as conservation effectiveness was improved by locating contrasting governance approaches of certified CFs and NFRs close to each other. The positive impact of Certified CFs on community relations with NFR managers provides additional support for the presence of synergistic effects between national and local PA governance. We suggest that moving beyond impact evaluations of single governance approaches operating in isolation, to recognising the contribution of multiple governance approaches operating in concert represents an important future direction for impact evaluations. Such advances have a wide range of applications including understanding the impact of PA zonation and the cumulative impacts of mining operations in the growing field of integrated landscape management (Sayer et al., 2013).

Overall, we suggest that discussions concerning recentralisation of CFs should be reconsidered in the light of the observed positive impacts. Our study supports calls to look beyond simplistic win-win framings of PAs in favour of seeking to understand complex PA impacts in order to better serve diverse and particularly marginalised actors (Agarwal, 2009; Baylis et al., 2016). Specific policy implications of our study concern the potential to utilise the observed synergistic effects between PAs. Locating autonomously governed national PAs and CFs next to each other might be a more impactful policy option than power-sharing approaches between national and local actors in the same PA, which have so far yielded poor results (Persha & Meshack, 2015; Keane et al., 2020). We propose that the next trend in the evolving governance of PAs might focus on novel configurations of national and local PA governance approaches to develop coherent PA networks that optimise conservation and wellbeing benefits within complex multi-use landscapes.

3.6 Acknowledgements

We are grateful to field assistants Petro Nnyti, Ignatus Pesha, Lilian Fredy, Lasima Nzao, all interviewees, Jasper Makala and all staff at MCDI, Nicole Gross-Camp, Henry Travers, Tanzania Forest Service, Kilwa District Government Forest Office, Njabha Lyatura, Nike Doggart, Charles Meshack. The research was funded by the UK Economic and Social Research Council award ES/J500215/1, conducted under COSTECH permit 2019-42-NA-2019-33. Marshall was funded by Australian Research Council Future Fellowship, FT170100279.

Chapter four: Pathways to win-wins or trade-offs? How certified community forests impact forest restoration and human wellbeing

Manuscript accepted for publication in Philisophical Transactions B.

Authors

Robin Loveridge [1,2]; Andrew R Marshall [1,3,4]; Marion Pfeifer[5]; Steven Ruston[5]; Petro Nyinti [6]; Lilian Fredy [7]; Susannah M. Sallu [8]

Affiliations

[1] Department of Environment and Geography, University of York, York, UK

[2] The Biodiversity Consultancy, Cambridge, UK

[3] Forest Research Institute, University of the Sunshine Coast, Australia

[4] Flamingo Land Ltd., Kirby Misperton, North Yorkshire, UK

[5] School of Natural and Environmental Sciences, TROPS lab, Newcastle University, UK

[6] Institute of Adult Education Dar es salaam, Tanzania

[7] Sokoine University of Agriculture, Morogoro, Tanzania

[8] School of Earth and Environment, University of Leeds, Leeds, UK

Key words: pathways, protected areas, forest certification, restoration, community forests, wellbeing

4.1 Abstract

Certified community forests combine local governance with forest certification and aim to serve multiple objectives including forest protection, restoration, human wellbeing and equitable governance. However the causal pathways by which they impact these objectives remain poorly understood. The ability of protected area impact evaluations to identify complex pathways is limited by a narrow focus on top-down theoretical, quantitative perspectives and inadequate consideration of local context. We used a novel mixed-methods research design that integrates the perspectives of multiple actors to develop a generalised conceptual model of the causal pathways for certified community forests. We tested the model using a combination of statistical matching, structural equation modelling and qualitative analyses for an agroforestry landscape in Tanzania. We found certified community forests positively impacted human wellbeing, equitable governance and forest restoration. Equitable governance had the largest impact on wellbeing, followed by crop yield and forest resource availability. Timber revenues varied widely between villages and the average effect of financial benefits did not impact wellbeing due to the immature stage of the certified timber market. We identified positive interactions and trade-offs between conservation and agriculture. Our findings suggest that no simple solution exists for meeting multiple objectives. However, developing understanding of the pathways linking social and conservation outcomes can help identify opportunities to promote synergies and mitigate negative impacts to reconcile competing objectives.

4.2 Introduction

Protected areas (PAs) have traditionally been established in areas of biodiversity importance with the objectives of protecting and restoring nature. However, over the last 50 years the roles of PAs have expanded to include human wellbeing and equitable governance objectives (Watson et al., 2014; Schreckenberg et al., 2016). Some contend that these objectives are mutually supporting, with 'win-win' outcomes possible for people and nature (Wells and McShane, 2004). However, others consider these objectives to be competing (Watson et al., 2014). For example, farming is the dominant livelihood of rural communities in low income countries and a key determinant of human wellbeing (Diao, Hazell and Thurlow, 2010). While agricultural expansion is also the primary driver of forest loss in the tropics and a major driver of degradation within PAs (Gibbs et al., 2010). To date trade-offs between objectives are more common than win-wins (Oldekop et al., 2016). Therefore, there is a need to re-evaluate the core assumptions underpinning win-win framings of PA governance (Ferraro and Hanauer, 2015), to understand if and how these multiple objectives can be reconciled.

Certified Community Forests (CFs) represent a new generation of PAs seeking to meet the expanding role of PAs, by combining two recent trends in forest PA governance: (i) decentralisation - transferring governance responsibility from central governments to local actors in efforts to enhance equitable governance; (ii) forest certification to increase the financial benefits and equitable benefit sharing to incentivise sustainable management (Agrawal, Chhatre and Hardin, 2008). Forest landscape restoration and other conservation approaches have also undergone parallel evolutions in governance to integrate social objectives. These new trends create additional complexity as international forest certification requirements are superimposed on local governance arrangements (Hajjar and Oldekop, 2018). Furthermore, diverse forms of certified CFs have evolved as governance arrangements are adapted to align with local cultural norms in efforts to create forms of governance that are considered locally legitimate (Koning, 2011). Perceptions of fair and equitable governance are increasingly associated with positive wellbeing outcomes and effective conservation (53). To understand and improve effective PA governance to serve multiple objectives, greater understanding is needed of how complex PA governance arrangements impact multiple objectives (Newton et al., 2016), and the local contextual factors that determine success and failure (Salerno et al., 2020). Yet few rigorous impact evaluations of certified CFs exist (Burivalova et al., 2017).

Analysis of causal pathways can be used to explain the complex processes by which PAs impact outcomes (Ferraro and Hanauer, 2015). Win-win outcomes require that forest PA governance leads to simultaneously positive impacts on forests and human wellbeing (Figure 1). Furthermore, win-win framings depend on the core assumption that human wellbeing is strongly linked to the natural environment (Stephenson et al., 2015). Therefore, good environmental governance will benefit people, with positive interacting pathways between social and ecological outcomes (Figure 1). Specific interaction pathways might include (i) forests to people – effective PA governance increasing the abundance of those forest products and ecosystem benefits (monetary and non-monetary) used by adjacent communities, resulting in improvements in human wellbeing, (ii) people to forests - positive attitudes towards conservation result in improved conservation outcomes (Oldekop et al., 2016). These interaction pathways influence how people perceive and respond to change (Pascual et al., 2014) and may positively reinforce positive outcomes (Dawson et al., 2017).

Figure 1. Conceptual model of generalised pathways between protected area (PA) governance, forests, and human wellbeing.

However, the presence of negative interactions between social and ecological outcomes would result in trade-offs between PA objectives. For example, if conservation governance is perceived as being unfair, this could lead to conflicts and local resistance (Holmes, 2007), as occurred in Kilimanjaro, Tanzania, where retaliatory killings of elephants have been recorded as a form of protest against PA conservation (Mariki, Svarstad and Benjaminsen, 2015). The presence of negative interaction pathways would constrain the ability of conservation interventions to achieve win-wins.

Box 1. Key concepts

Pathways. We draw on the impact evaluation, international development, sustainable forest livelihoods literature to refer specifically to a causal pathway as a hypothesised means by which PA governance impacts conservation or social outcomes (Sunderlin et al., 2005; Ferraro and Hanauer, 2015; Bernstein and Cashore, 2012). Each pathway is composed of a conservation intervention, one or more mechanisms causally linked to the intervention and a human wellbeing or conservation impact.

Mechanisms are the intermediate outcomes, linked along a hypothesised causal pathway connecting PA governance to a human wellbeing or conservation effectiveness impact (Ferraro and Hanauer, 2014a).

Conceptual models also referred to as a 'theory of change', aggregate the main pathways for describing important relationships in the socio-ecological system (Bernstein and Cashore, 2012; Mckinnon et al., 2016). Conceptual models are well suited to testing core assumptions of PA governance by articulating implicit assumptions into explicit testable pathways (Woodhouse et al., 2015; Baylis et al., 2016; Ferraro and Hanauer, 2015).

Advances in the field of impact evaluation have seen the mainstreaming of statistical matching to exclude alternative explanations and attribute observed differences to the intervention (Gertler et al., 2016). However, this approach has limited ability to assess the pathways by which PAs cause impacts because of a focus on (i) ends, rather than means, (ii) a narrow suite of pre-determined hypotheses, rather than open-ended consideration of multiple contextual drivers. Firstly, evaluations tend to use single response variables to focus on long-term impacts (Salerno et al., 2020), but the mechanisms through which these impacts are achieved remain largely untested, though for an exception see (Ferraro and Hanauer, 2015). For example, improving equitable resource governance is one mechanism by which PAs can improve human wellbeing (Dawson, Martin and Danielsen, 2018), however the relationship between equitable governance and wellbeing is rarely assessed (Galvin, Beeton and Luizza, 2018). Use of a conceptual model provides a powerful alternative approach for unpacking complex impacts into causally linked short and long-term outcomes (Woodhouse et al., 2015).

Secondly, impact evaluations tend to test *a priori* hypotheses of how PAs impact wellbeing or conservation using exclusively quantitative approaches. This top-down framing of a study system excludes local perspectives and has been described as creating a 'bottleneck' in dialogue (Evans, Murphy and Jong, 2014), which may miss locally relevant, unanticipated and alternative explanations for PA impacts (Rode, Gómez-baggethun and Krause, 2014; Turnhout, 2014). Greater integration of the perspectives of local actors in the design of impact evaluations has potential to provide more comprehensive understanding of governance challenges and provide novel insights.

We aim to advance methods to evaluate the success of conservation and restoration interventions. We identify and test the causal pathways by which certified CFs impact human wellbeing and forest restoration for a case study ~~example~~ in Tanzania. Specifically, we evaluate win-win assumptions of PA governance by testing pathways of (i) equitable governance, (ii) financial benefits, (iii) interaction effects, and (iv) trade-offs.

We advance on existing evaluation methodologies by (i) combining statistical matching with a conceptual model (ii) integrating top-down theoretical perspectives with bottom-up perspectives of local actors to promote inclusive consideration of alternative explanations from marginalised actors. We thereby contribute to two key knowledge gaps of the United Nations Decade on Restoration: (i) methods for designing interventions, and monitoring restoration success; (ii) linkages between the health of ecosystems and the flow of services to communities (United Nations, 2020).

4.3 Methodology

4.3.1 Study site and sampling

Tanzania provides an excellent test case of the challenges to reconcile forest restoration and human wellbeing objectives as national development policies aim to expand both agricultural and PA land-uses, while CFs are often established on forests already considered to be degraded and economically marginal (Nelson, 2011). We focus on a PA dominated landscape in Eastern Tanzania, where certified CFs have an established history and community governance is supported by district government and an NGO, Mpingo Conservation & Development Initiative (MCDI).

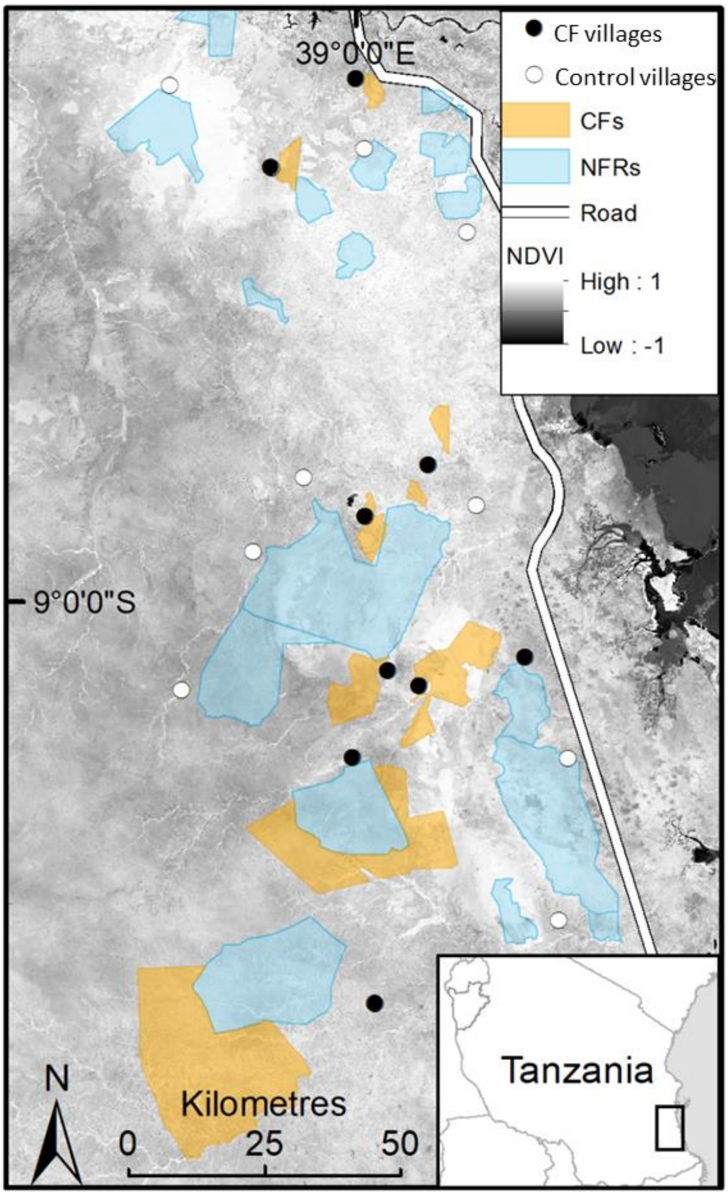


Figure 2. Study landscape in Lindi and Pwani regions of Tanzania detailing the 9 certified CF villages and adjacent National forest Reserves (NFRs) and 10 control villages.

4.3.2 Developing a conceptual model

Our conceptual model aimed to integrate actor perspectives with hypothesised pathways derived from conservation science literature (Figure 3). We consulted national, regional and local actors involved in forest PAs and forest certification, in total undertaking 30 focus groups and 34 key informant interviews between 2018 and 2019 (supplementary material). Questions concerned how PAs influenced conservation and human wellbeing and PA governance processes and challenges.

A thematic analysis of transcripts was then undertaken to identify actor perspectives of the main pathways linking certified CFs with conservation and human wellbeing impacts by identifying logical causal statements that were similar between independent consultation sessions. Identified pathways were then converted into a connected sequence of indicators linking CF governance via one or more causally linked mechanisms to human wellbeing and conservation impacts (supplementary material). The overall conceptual model was then composed of these main pathways. Indicators may be connected to more than one pathway if the causal logic suggested interactions between indicators from different pathways.

All indicators in the conceptual model were included in a quantitative questionnaire to collect data on community perceptions of all indicators, which were then used to test the model. Our conceptual model emphasises measuring 'bottom-up' community perceptions, rather than externally measured data sources because perceptions are important drivers of local behaviour and success of conservation interventions (Bennett, 2016).

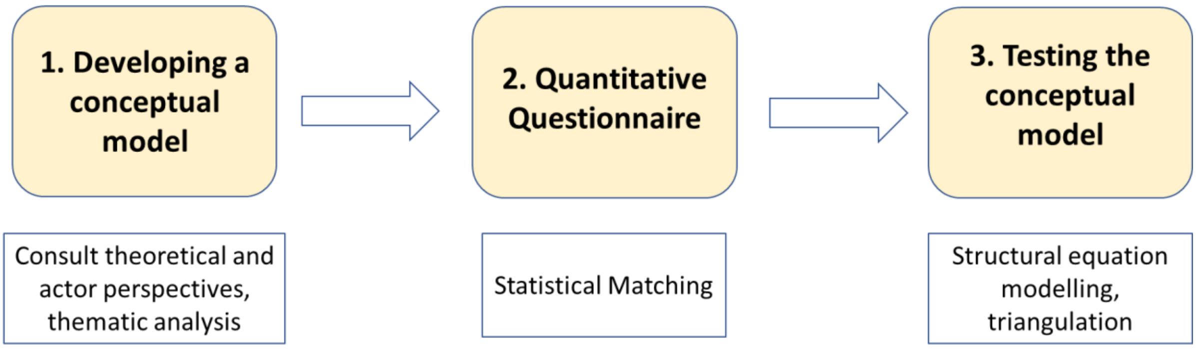


Figure 3. Steps (above) and methods (below) used to develop and test the conceptual model.

4.3.3 Matched sample selection

To improve the causal inference of our study we used statistical matching to compare certified CFs to control villages that represent the counterfactual situation– what would have happened in the absence of the intervention (Schleicher et al., 2020). This approach helps to exclude alternative explanations and attribute observed differences to the PA (Gertler et al., 2016). We therefore define impact as the difference in response indicators between the sampled certified CF villages and matched control villages. We defined treatment as certified CFs that have been established and participated in forest certification for at least five years. Therefore, all certified CFs had completed at least one five-year management cycle of harvesting, revenue disbursement and external assessment by an FSC auditor, providing assurance that the required governance processes were being implemented and the villages represented 'de facto' examples of this governance approach. All 9 villages that met these criteria were matched to 10 control villages using nearest neighbour matching implemented in the 'Matchit' package in R version 3.5.3 to select treatment and control villages with balanced distributions of a suite of confounding socio-environmental variables (supplementary material). The statistically selected control villages were then reviewed by an expert panel of national actors to confirm the appropriateness of empirically determined matches from the perspective of national actors (supplementary material).

The questionnaire was undertaken with 955 people from the 9 villages with certified CFs and 10 matched control villages, with at least 50 respondents per village stratified by gender, local elite status (village government representative, 0/1) and wealth category (supplementary material). The questionnaire was undertaken at the scale of individuals rather than whole households since the concept of wellbeing contains subjective elements which cannot be generalised across households (Woodhouse et al., 2015).

4.3.4 Response indicators

Human wellbeing

Multidimensional human wellbeing provides a comprehensive measure of social impacts. Wellbeing indicators were identified following the Wellbeing Indicator Selection Protocol (Loveridge et al., 2020) by a subset of 8 focus groups with separate groups of women, men, elites and non-elites in target communities. During the focus groups locally relevant indicators were agreed for five broad domains of wellbeing; (i) material wellbeing, (ii) health, (iii) social relations, (iv) security, (v) freedom of choice and action (Boarini, Kolev and McGregor, 2014). Through stepwise reduction of questionnaire data we produced a final list of 25 orthogonal indicators representative of the five domains of wellbeing. A Human Wellbeing Index (HWI) was then calculated as the mean score of each domain mean to provide a single response indicator (supplementary material).

Forest restoration

Miombo woodlands were the dominant forest type in the study landscape. Normalised Difference Vegetation Index (NDVI) correlates with ground vegetation biomass and productivity under low to medium vegetation density conditions such as the Miombo woodlands (Pfeifer et al., 2016). NDVI change therefore provides an appropriate proxy measure of forest recovery. We calculated NDVI change between 2014 and 2019 as a proxy measure of conservation effectiveness on village land (supplementary material). NDVI change was calculated for each survey respondent as the mean of 30 x 30 m pixels within a 10 km radius of the respondent's house as determined by GPS during interview. In this way the conservation effectiveness indicator concerns sustainable forest management on village land surrounding the respondent's home.

Equitable governance

Equitable governance concerns notions of fairness (Schreckenberg et al., 2016), in relation to three dimensions: (i) Distribution - the fair distribution of benefits and costs; (ii) Recognition – respect for the rights and values of different actors; (iii) Procedure – the fair participation of actors in decision making, relating to transparency of information, accountability of managers and equitable dispute resolution. Guided by the IUCN and IIED good PA governance assessment methodologies (Franks and Booker, 2018; Borrini-feyerabend, Dudley and Jaeger, 2013; Schreckenberg, Franks and Martin, 2016), we included questions on governance challenges in exploratory consultations. Based on actor perspectives of governance challenges we included locally relevant equitable governance indicators of all three dimensions of equitable governance in the quantitative questionnaire (supplementary material). An overall governance equity index was generated by confirmatory factor analysis (CFA) of candidate governance indicators from the three dimensions of governance equity.

4.3.5 Testing the conceptual model

We tested the conceptual model by structural equation modelling using PiecewiseSEM in R (36). Latent variables were first modelled by confirmatory factor analysis in the Lavaan package in R (52), then the latent variable constructs were passed into the PiecewiseSEM package for modelling. This approach is justified because PiecewiseSEM does not use a global covariance matrix and so the covariance components for the individual latent variables are not required for solution of the overall path model. Furthermore, we were particularly interested in adjusting for covariation due to village to make the analysis more generic and latent variables are not easily derived in a mixed effect framework. We then fitted (generalised) linear mixed effects models using PiecewiseSEM to account for the hierarchal data structure of interview respondents being nested within villages. For binary response variables we used binomial error distribution in generalised linear mixed effects models.

We included socio-economic indicators for local elite status and gender in models to account for any systematic perception biases between actors. To account for residual imbalances in the distributions of confounding variables between treatment and control groups we included orthogonal sets of confounding variables as predictors of response variables (37). We then performed model simplification to achieve statistical parsimony i.e. the minimum complexity necessary to describe key relationships (Crawley, 2007). For each linear mixed effects model in turn, we sequentially removed the least explanatory predictor variable from the full model if its deletion caused a reduction in Akaike's Information Criterion.

For continuous variables we report standardized path coefficients, which estimate the expected change in the response variable (e.g. wellbeing) as a function of the change in the explanatory variable (e.g. equitable governance), in units of standard deviation. For categorical variables (e.g. governance treatment vs control group), we report the model-estimated means for each factor level (Lefcheck, 2016). For all response variables we report the marginal r^2 values (the variance explained by fixed effects).

We assessed the overall model fit by Shipley's test of d-separation, accepted when Fisher's C statistic is higher than a significance level ($p < 0.05$). Finally, we critically reviewed the conceptual model through triangulation with the qualitative data to assess whether the identified pathways and trends were representative of different village cases and actors sampled. This served as a verification check to ensure inclusive representation of pluralistic perspectives, particularly potentially marginalised or minority actors whose perspectives might otherwise be masked by reporting normative trends.

4.4 Results

We identified five main pathways linking governance of certified CFs to human wellbeing and forest restoration (Figure 4). First, an equitable governance pathway (orange, Figure 4), indicating that certified CFs were hypothesised to improve equitable governance, which would in turn positively impact human wellbeing. The most stated wish by community actors was that CF benefits be fairly distributed, e.g.

'One of our expectations was to make sure the community forest benefits all villagers.' – Community member, village 5.

Testing of our quantitative questionnaire showed that certified CFs positively impacted equitable governance (control estimate = -0.11, certified CF estimate = 0.25), with mean scores of all equitable governance indicators (participation, trust, satisfaction) higher in villages with certified CFs than control villages. In turn, equitable governance positively impacted human wellbeing both directly (standardised estimate = 0.18) and indirectly (Figure 3). The indirect pathway occurred because equitable governance predicted perceived improvements in shared community development benefits (standardised estimate = 0.11). Timber revenue in certified CFs was most commonly spent on improvements to (i) healthcare facilities e.g. solar lights, (ii) education e.g. school meals and (iii) water infrastructure e.g. for drinking and irrigation. Irrigation was particularly important because the dominant livelihood activity was small-scale cultivated agriculture. These infrastructure improvements had a positive impact on ecosystem service benefits represented by perceived soil fertility, firewood and water access (standardised estimate = 0.18). Ecosystem benefits positively impacted crop yield (standardised estimate = 0.11), which in turn had a positive impact on human wellbeing (standardised estimate = 0.14; Figure 3).

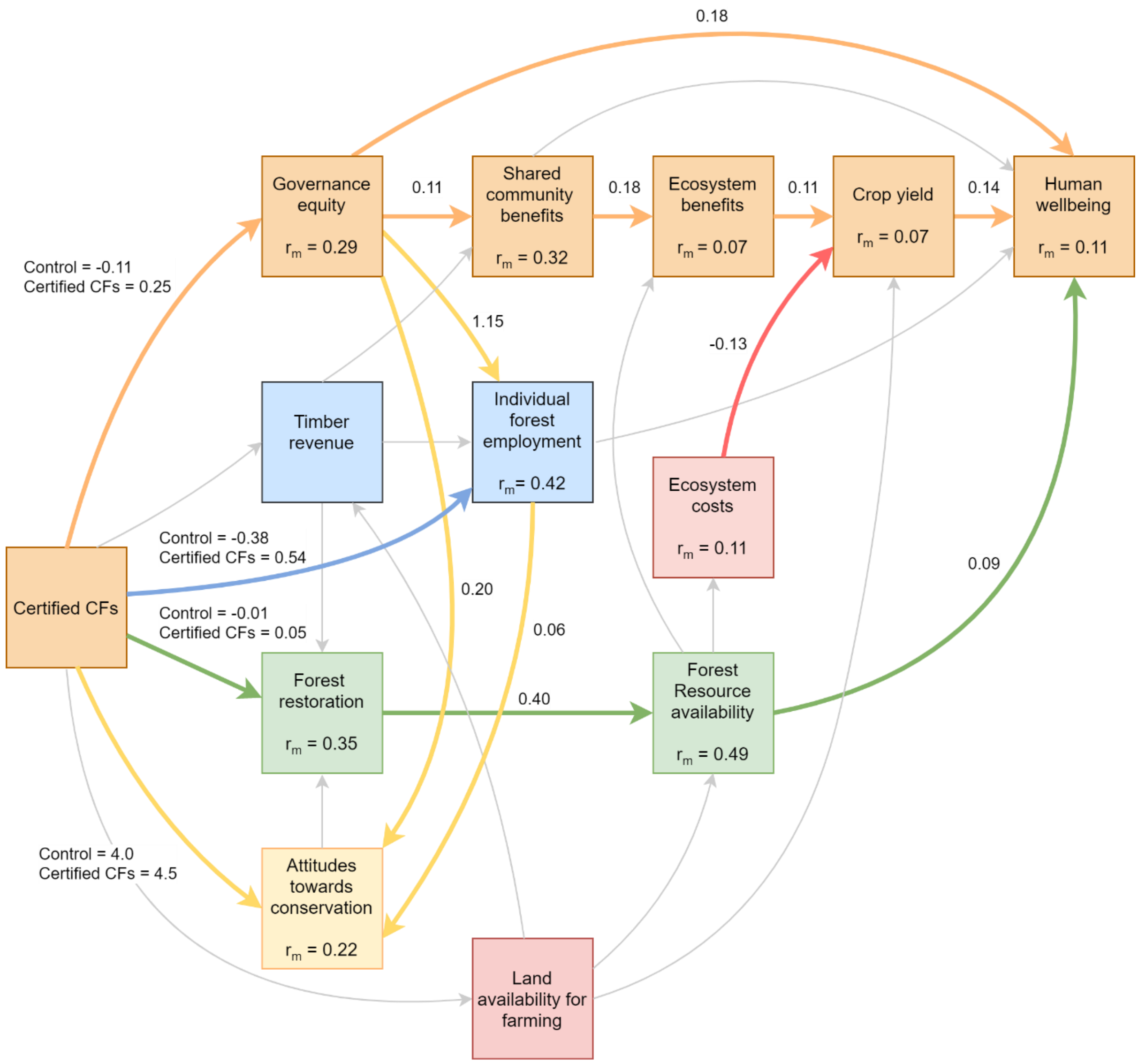


Figure 4. Conceptual model showing theoretical win-win pathways and locally perceived pathways. Colours highlight different pathways; orange = governance equity pathway, blue = financial pathway, green = conservation pathway; yellow = social interaction pathways, red = trade-off pathways. Grey arrows show paths removed from the simplified final structural equation model. Numbers show standardised path estimates for continuous variables and model estimates for binary variables for retained paths. Rm = marginal r^2.

Second, a financial benefits pathway (blue, Figure 4), whereby certified CFs are hypothesised to increase timber revenue, providing paid forest employment opportunities and direct financial benefits for individual community members, as well as contributing to community development projects. However, in our case study example the financial pathway did not impact forest restoration or wellbeing. Certified CFs did not impact timber revenue and although certified CFs impacted forest employment (control estimate = -0.38, certified CF estimate = 0.54), forest employment did not impact wellbeing. Qualitative data corroborated these quantitative findings, with concerns raised in four of the six certified CF villages consulted during exploratory focus groups, for example:

'We have a lot of forest here and we work to manage it. But we have a limited number of customers and so the villagers here have not yet felt the actual benefit of this forest. But if we could get many customers to buy our timber then every member of this village could realize the importance of managing the forest' - Village Natural Resource Committee Secretary, village 11.

However poor economic performance was not the case in all villages. Annual timber revenue was highly variable between certified CFs (mean of 12.75 USD per person; range 1.17 – 80). The size of certified CFs was also highly variable (mean of 14,320 ha; range 920 - 64,550), with larger CFs generating more revenue. Some differentiation in timber revenue spending was observed between villages; the highest timber producing village able to undertake additional, larger-scale and more diverse community development projects, including: (i) building new village government offices, (ii) installing primary school sanitation facilities, (iii) improved village healthcare provision targeting facilities for pregnant women and disabled patients and health insurance for village natural resource committee members, (iv) building a village-run guesthouse, (v) payment for forest patrols, planning meetings and patrol equipment, (vi) payment for professional forestry and governance training services from the supporting NGO MCDI and the district government forest office. An FSC certified sawmill factory run by a sustainable timber production company called Sound and Fair had also been established in this village in 2017 to further up-scale timber production and revenue generation, providing additional employment opportunities.

Third, a conservation pathway (green, Figure 4) showed that Certified CFs positively impacted forest restoration (control estimate = -0.01, certified CF estimate = 0.05)), which in turn positively impacted perceived availability of forest resources (standardised estimate = 0.40) and availability of forest resources had a positive impact on human wellbeing (standardised estimate = 0.09). An indirect conservation pathway was also hypothesised, whereby availability of forest resources would impact provision of ecosystem benefits, with a knock-on positive impact on crop yield. However, forest resource availability did not predict ecosystem benefits in our case study.

Fourth, positive interaction pathways were hypothesised from social to conservation outcomes (yellow, Figure 4) with positive social outcomes driving improved attitudes towards conservation and forest restoration. In our case study, both governance equity (standardised estimate = 0.20) and forest employment (standardised estimate = 0.06) had a positive impact on attitudes towards conservation. However quantitative analyses showed that attitudes towards conservation did not have an impact on forest restoration. Likewise, timber revenue did not predict forest restoration, despite timber sales being used in part to fund forest management:

'We are funding our own forest management activities. We pay even from our own village basket for meetings and patrols' - Village Natural Resource Committee member, village 2.

Fifth, trade-off pathways were hypothesised between conservation and agriculture (red, Figure 4). Focus groups in communities identified the widespread perception that crop damage from wild animals was a serious problem and that this was worse near forests (ecosystem costs). In our case study, ecosystem costs had a negative impact on reported crop yield (standardised estimate = -0.13). However, ecosystem costs was not predicted by our measure of forest resource availability despite anecdotal evidence from key informant interviews associating healthy forests with an increase in crop damage. A further trade-off pathway was identified in three of the villages where expansion of PAs was causing concerns about land shortages for farming:

'The government extended the national forest reserve boundary and so we have been left with a small area for farming. That land, it was very fertile, it was supporting us to have high production and we had a lot of food surplus. But now we have little food because we harvest very little' – Community member, village 1.

Larger certified CFs were also suggested by MCDI to have more economic potential for timber revenue, creating a trade-off between land-uses. However, testing of our conceptual model did not show an impact of certified CFs on perceived land availability, or knock-on impacts of perceived land availability for farming on crop yield or timber revenue.

4.5 Discussion

We found evidence of both win-win and trade-off pathways from certified CF governance to forest recovery and human wellbeing. Certified CFs positively impacted (i) equitable governance and (ii) forest recovery and both pathways positively impacted wellbeing, supporting win-win assumptions that positive social and conservation outcomes can occur together and are causally linked. However, additional hypothesised win-win pathways of (iii) financial benefits from certified CFs and (iv) improved attitudes towards conservation did not impact either wellbeing or forest restoration, suggesting that the importance of these pathways was limited in our case study. The limitation of the economic pathway linking forest governance to wellbeing may be due to the FSC timber market being at an early stage of development and operating sub-optimally. While some villages were able to generate significant FSC timber revenue, which was used to deliver integrated programmes to improve human wellbeing, concerns were raised by other villages about the challenges of accessing timber markets. Nevertheless, these findings agree with other research suggesting that equitable governance can be a more important driver of successful conservation than financial incentives (53). Finally, trade-off pathways between conservation and agriculture, and the importance of agriculture for wellbeing provide contrary evidence that conservation and wellbeing objectives may be competing. These findings show that no simple solution exists for meeting multiple objectives in PA governance.

By disaggregating the impacts of PA governance into multiple pathways it is possible to identify which aspects of an intervention are performing well and which aspects are failing. In our case study, the governance equity component of certified CFs had a positive impact, in contrast to other CF programmes in Tanzania (Magessa, Wynne-jones and Hockley, 2020). However, forest employment did not have an impact on wellbeing. This suggests that the hypothesised financial benefits aspect of certified CFs was underperforming, like other CF programmes in Tanzania (Gross-camp, 2017) and market-based approaches across the tropics more generally (Calvet-Mir et al., 2015).

By comparing quantitative and qualitative findings it is possible to explore how variation in contextual factors can lead to alternative outcomes. Not all CFs showed poor financial performance. The village which had established the largest CF, was also generating the most timber revenue, which was spent on diverse community projects, including additional income generation schemes such as a village-run guesthouse. The economic potential of this village had also attracted the establishment of an FSC certified sawmill, in contrast to other villages which struggled to attract timber buyers. This more economically successful village, contrary to the general trend identified by modelling analyses, suggests that improved financial performance may require villages to dedicate significant land area to CFs and that strategies are needed to improve engagement with timber markets.

The positive interaction pathways suggest the potential for virtuous cycles to occur over time (Barsimantov, 2010), where positive social impacts caused by certified CFs would improve attitudes towards conservation, leading to forest recovery, which would drive an upwards spiral of continuous improvements in social and ecological outcomes. However, evidence to support the presence of social interaction pathways was limited, with the generally poor financial performance of CFs constraining community investment in forest management. Testing of this hypothesised virtuous cycle would require time-series data to show change over time and evaluation of feedback mechanisms through more complex, non-recursive structural equation models. For example, testing forest governance impacts wellbeing and then reciprocally how changes in wellbeing might impact forest governance. Exploration of virtuous cycles represents a future research direction with potential to identify win-win pathways that would amplify the benefits of longer-term interventions.

The trade-off pathways, whereby forests provide both ecosystem benefits and costs for agriculture and potential agricultural land-shortages caused by expansion of PAs, suggest that the study landscape represents a microcosm of global challenges to reconcile forest conservation with rural development objectives (Gibbs et al., 2010; Watson et al., 2014). We did not find significant evidence of land-shortage trade-offs. However looking to the future, as waves of forest degradation and land-use change penetrate deeper into rural areas of low income countries (Ahrends et al., 2010), these trade-offs between agriculture and conservation will likely become more acute. The agriculture-conservation trade-off represents a potential crossroads in land-use decision-making that will determine whether the landscape follows the same trajectory of current global trends, prioritising agricultural expansion and forest degradation (Gibbs et al., 2010). Alternatively, theory on transformative change suggests that to tip the system from one state (e.g. agricultural expansion) into another (e.g. sustainable use) requires disruption of the dominant drivers on the system (Gunderson and Holling, 2002). Our conceptual model identifies key points in the system which could be leveraged to promote sustainable land-use choices. Specifically, mitigation strategies would be needed to minimise negative impacts of wild animals on crop production and efficient land-use planning to minimise land shortages, particularly for marginalised groups. Simultaneously, pathways of equitable governance, financial benefits and availability of forest resources, which all positively impacted wellbeing would need to be optimised to offset costs of foregone agricultural activity.

Our research design sort to embrace a complex systems perspective. However, several simplifications were necessary to aid interpretation. To support a statistical comparison, we employed a binary distinction between the governance approaches of certified CFs and control villages. However, we recognise that within this overarching governance grouping, varying governance arrangements exist. To move beyond a coarse binary description of governance approaches, we employed qualitative methods to highlight outlying cases that contrasted with the normative quantitative trends reported. However further exploration of the within-group variation in governance approach would be possible through more in-depth case study research (10). Both human wellbeing and equitable governance are multidimensional concepts. However, these concepts were both consolidated into single indexes for the purposes of this study as a trade-off between complexity and simplicity to limit the number of pathways in the overall conceptual model to aid interpretation of dominant trends. This aggregation approach limits understanding of which specific dimensions of wellbeing and governance are being impacted by the intervention. Disaggregation of impacts between wellbeing domains for this dataset was undertaken in a separate study, which found that certified CFs positively impacted health, security and freedom domains of wellbeing, but not material wellbeing or social relations (51). To strengthen causal inference in the research design a difference-in-differences approach might be employed to assess change between treatment and control groups through time. However, given the inductive selection of wellbeing indictors, such longitudinal data was not available for this study. Finally, although we found evidence that certified CFs positively impacted forest restoration, equitable governance and attitudes towards conservation, we were not able to confirm the intermediary mechanism linked certified CFs to improved forest restoration. This represents an important area of future research.

Our novel methodology illustrates the utility of a mixed methods approach for developing and testing theory of complex systems, with quantitative analyses showing overall trends, while qualitative analyses identifying alternative pathways missed by normative analyses. The integration of multiple actor perspectives provided a more comprehensive and contextualised understanding of pathways that balanced assumed positive and negative impacts of forest governance on people and forest recovery. By integrating views of actors from the global south our methodology makes progress in operationalising calls for a pluralistic perspective of conservation challenges (Pascual et al.), to improve the equity of both the research process to conceptualising challenges and design of effective solutions. As the role of PAs continues to diversify, methodological innovation is needed to understand how complex, positively and negatively interacting pathways can be navigated to promote forest recovery and human wellbeing. Through the combination of qualitative and quantitative methods, it is possible to move beyond a simplistic understanding of intervention performance, towards more in-depth understanding of what works, where and why.

4.6 Acknowledgements

We are grateful for the help of field assistants Ignatus Pesha, Mercy Mgaya, Lasima Nzao, all interviewees and participants, technical guidance from Nicole Gross-Camp, Phil Franks, Abigail Wills, logistical support from Tanzania Forest Service, Kilwa District Government Forest Office, MCDI, AGRISYS Tanzania and two anonymous reviewers. Primary donor: Economic and Social Research Council, ES/J500215/1. Other donors: BBSRC BB/S014586/1 (Pfeifer); ARC FT170100279 (Marshall). COSTECH permit 2019-42-NA-2019-33.

Chapter five: Synthesis & Discussion

'Simple methods for asking complex questions'

5.1 Introduction

In this thesis I aimed to evaluate how innovations in PA governance have led to complex impacts on human wellbeing and conservation effectiveness. In meeting this aim I aspired to advance robust interdisciplinary analysis techniques that integrate different perspectives in order to provide a balanced representation of the situation on-the-ground. The contributions of this thesis are therefore both methodological and theoretical, providing novel insights into complex PA governance challenges. In this final chapter, I first synthesise key insights and contributions from across the preceeding chapters, then provide a summary discussion of the implications and limitations for PA practice, and finally, look forward to new frontiers for research on PA governance.

5.2 Synthesis of key knowledge contributions

My thesis forms a integrated body of work by making progressvely deeper insight into PA governance challenges. In chapter one I identified conceptual, methodological and explanatory data gaps that inhibit the evaluation of win-wins in PA governance, guiding the focus of subsequent chapters. In chapter two I developed a protocol for measuring multidimensional human wellbeing. This protocol was then implemented in chapter three in combination with remote sensing analysis to assess the human wellbeing and conservation impacts of PAs in a complex PA dominated landscape in Tanzania. Then in chapter four I explored why these impacts have occurred by testing the core assumptions and logic underpinning recent innovations in PA governance. By sequentially building on the knowledge and methods developed in earlier chapters I've sought to move beyond a simplistic win-win framing of PA governance towards a more detailed understanding of what works where, for whom and why.

In chapter one, I identified conceptual, methodological and explanatory data gaps that inhibit understanding of social and ecological impacts of PAs. The identified gaps guided the focus of subsequent research chapters. Key gaps concern how wellbeing impacts should be conceptualised and this ambiguity leads to assessments characterising impacts in a variety of ways. This variable approach creates problems for interpreting whether win-wins in PA governance have been achieved, as assessments may measure only a narrow suite of indicators (de Lange, Woodhouse and Milner-Gulland, 2016), which risks missing unintended negative impacts (Woodhouse et al., 2015). Further data gaps concerned 1) methodological limitations of social assessments, 2) impact evaluations prioritising top-down perspectives,

3) few rigorous assessments of certified CFs (Burivalova et al., 2017). Evaluation of PA governance is lagging behind conservation practice and on-the-ground experimentation. Finally, a lack of critical evaluation of the assumed logic underpinning recent innovations in PA governance act as a barrier to more nuanced insights concerning what aspects of PA interventions are working and why, needed to help inform on the ongoing adaptive refinement of PA governance approaches. The following sections detail the contributions of my thesis with respect to these gaps.

In chapter two I suggested an approach for measuring human wellbeing that is mindful of practical resource constraints of practitioners. In chapters two and three, WISP was used to develop comprehensive measures of multidimensional human wellbeing composed of 19 and 25 indicators, respectively. In chapter three I apply WISP in a complex PA dominated landscape and identify diverse PA impacts with unequal impacts on the different domains of wellbeing. Based on the findings of these two chapters I conclude that social assessments which take a wide, multidimensional framing are an important step forward to ensure comprehensive accounting of PA impacts in complex environments where the impacts of PAs cannot be fully predicted prior to investigation.

In chapter three, we found that certified CFs in Tanzania have a positive impact on conservation effectiveness. This is in contrast to other studies of CF programmes in Tanzania (Magessa, Wynne-jones and Hockley, 2020) and globally (Rasolofoson et al., 2015). This has implications for PA policy in Tanzania, suggesting that current government discussions concerning the centralisation of PAs should be reconsidered. Furthermore, we found that locating certified CFs next to strictly protected national PAs increased the effectiveness of both PA governance approaches. This finding has implications for the polarised debates concerning whether centrally or locally governed PAs are more effective (Brockington, 2003). The evidence for the higher efficacy of either centrally or locally governed PAs remains inconclusive (Oldekop et al., 2016). However when considering the complex reality of diverse PA impacts in complex landscapes, I suggest that this coarse distinction and comparisson between PA governance approaches may be an over simplification which is unlikely to be conclusively resolved by global analyses that lack contextual understanding.

In chapter four I identify pathways by which certified CFs impact human wellbeing and conservation effectiveness. Equitable governance was shown to be a stronger predictor of wellbeing than direct economic benefits from timber sales. Exploratory qualitative data collection also identify lack of timber sales as a weakness of certified CFs. Taken together these finding suggest that the economic pathway is not operating as intended. I also identified a potential win-win pathway, whereby certified CFs positively impacted conservation effectivenes, leading to increased availability of forest resources, which in turn predicted human wellbeing. Qualitative evaluation of one particularly successful village also identifies a possible positive feedback pathway, with timber revenue being fed back into conservation management suggesting the potential for a 'virtuous cycle' of simultaneously amplifying conservation and wellbeing benefits. However significant timber revenue generation was only identified in this single village. My analysis also identified a potential trade-off pathway, whereby PA size predicts timber revenue, however PA expansion may create land shortages and negatively impact farming and crop yield, which was shown to be an important driver of wellbeing. The presence of this trade-off pathway is also corroborated by analyses from chapter three, which showed that forest degradation was greater closer to villages (chapter three supporting information table S26, S29), indicating that agricultural expansion from village centres is progressively driving back forested land to more remote areas. These findings suggest that the study landscape is consistent with wider trends of agricultural expansion being a dominant driver of forest loss (Laurence, 2012) and suggest that rural development and forest conservation are directly competing objectives.

5.3 Methodological advances

A significant methodological contribution of this thesis has been to provide blueprints for how top-down and bottom-up perspectives can be integrated to get a more well-rounded understanding of complex PA governance challenges. In chapters one and two I explore how the two perspectives prioritise different forms of scientific validity. Top-down perspectives that focus on theoretical assumptions and quantitative testing prioritise external validity – the ability to generalise findings to other contexts (Campbell and Stanley, 1963). Being able to make generalisable conclusions is vital to support policy engagement and recommendations at spatial scales beyond the study landscape. However, top-down perspectives risk marginalising the importance of local values, context-specific factors and may lack local relevance (Woodhouse et al., 2015). In contrast bottom-up perspectives that emphasise local knowledge and contextual explanations prioritise ecological validity - the extent to which the researchers' findings reflect the lived experience of those whom the researchers are studying, and support internal validity and theory development - the ability of studies to attribute findings to particular drivers of change (Yue, 2012). Ecological validity ensures that local relevance is retained, promoting rather than marginalising the priorities of study participants (Howard et al., 2016). Therefore integration of top-down and bottom up perspectives is of central importance for providing a comprehensive understanding of complex PA challenges. This is highlighted in chapter four, where top-down theoretical assumptions, and bottom-up exploratory focus groups identified contrasting pathways between PA governance, wellbeing and conservation effectiveness.

A related methodological contribution has been to advance efforts to measure complex phenomena by integrating quantitative and qualitative methodologies. Top-down and bottom-up perspectives are intrinsically linked to specific disciplines and methodological approaches. Top-down perspectives tend to be found in natural, and quantitative social sciences such as economics and rely principally on statistical modelling techniques to test hypotheses (McGregor, McKay and Velazco, 2007). Bottom-up perspectives are common in qualitative social sciences such as political ecology and anthropology and depend on qualitative analyses such a document and thematic analyses (e.g. Beauchamp et al., 2018). I believe that quantitative approaches are essential for policy engagement and to track progress against verifiable targets to support adaptive refinement and increased success of conservation interventions. However, it is important to also recognise the limits of numeric approaches. For example, numeric values may not adequately represent the concepts they intend to portray (construct validity), or fail to capture locally important determinants for inherently contextual concepts such as wellbeing (Woodhouse et al., 2015). Therefore, the relevance of quantitative approaches can be enhanced by considering contextual factors to adjust numeric values based on qualitative insights. For example, in chapter two I present the Wellbeing Indicator Selection protocol (WISP) that combines qualitative and quantitative methods to provide a locally relevant quantitative measure of human wellbeing. Through WISP, locally identified indicators of wellbeing are simplified by statistical modelling to create a parsimonious subset of indicators. However this simplified list is adjusted, based on participant priorities to ensure that the final list of indicators retains local relevance. Since publication, WISP has been applied in other contexts such as incorporation into modelling of the interactions between agricultural practices and human wellbeing (Pfeifer, 2020; Milheras et al., in press).

In chapter three, I applied the methodological approach of integrating top-down and bottom-up perspectives to statistical matching. By integrating bottom-up matching criteria into traditionally top-down matching approaches identified more appropriate control villages for comparison with certified CFs and therefore increased the accuracy of impact estimates.

Conceptual models, also known as theories-of-change, are becoming a common-place tool among conservation practitioners for monitoring of project success (Kapos et al., 2009). However empirical testing of the conceptual models themselves, and the underlying logic on which projects are premised is rarely undertaken. Testing of project logic is important because project planning and evaluation without critical reflection and exploration of alternative perspectives and explanations risks continuously perpetuating past mistakes and closes off the possibility of discovering creative new solutions (Pascual et al., 2021; Sheraz, 2021). In chapter four, I progress methods for understanding PA governance challenges by using an interdisciplinary design to combine top-down and bottom-up persptives to identify alternative explanations. Furthermore, I introduce the use of structural equation modelling to directly test the conceptual model, providing one of the first studies that I am aware of to directly test the assummed logic of innoations in PA governance. By transporting existing methods for use in novel research designs I hope to share opportunities for conservation practitioners to bring more analytical rigor to the evaluation and creative problem solving of complex PA governance challenegs.

5.4 Discussion of implications and limitations for PA practice

To transfrom the current global paradigm of increasing agricultural expansion into a trajectory that sustainably conserves tropical forests and allows continued improvements on human wellbeing is a huge challenge (Adams, 2003). In order for certified CFs to play an effective role in supporting this transition to a more sustainable future, CFs will need to create wellbeing benefits that excede the opportunity costs of agricultural expansion. In the study landscape in Tanzania the different pathways by which certified CFs can contribute to win-wins showed mixed success. Equitable governance pathways were observed to generate positive wellbeing outcomes, however financial benefits did not have a significant impact on wellbeing due to the immature stage of the certified timber market. Both equitable governance and economic pathways will likely need to function optimally in order for certified CFs to become a viable long-term land-use option to halt the waves of agricultural exapansion and forest loss eminating from urban centres (Ahrends et al., 2010). To deliver effective certified CF governance that maximises wellbeing benefits will require investment in advanced interdisicplinary expertise on the part of practitioners in the form of 1) a high level of cultural understanding to develop locally appropriate forms of governance aligned with local perceptions of fairness; 2) practical business experitise to develop competitive supply of timber products to national and international markets. This technical burden could be eased by efforts to better align national protected area legislation with international market regulations to streamline and simplify administrative requirements. This would help narrow potential capacity gaps at the local level and make market-based approaches more accesible to local forest managers (Frey, Charnley and Makala, 2020).

Given the diverse social and ecological impacts of PAs identified, landscape-scale conservation, considering impacts beyond PA boundaries, provides a useful framing for understanding and mitigating unintended negative impacts to improve PA efficacy. However landscape-scale thinking must also be integrated with larger-scale processes, such as supportive national policies and international market access, recognising the embededness of of landscape-scale operations within larger scale political and economic social structures. In the face of the rapid pace of socio-environmental change across the tropics, long-term interventions are needed to continuously adapt globalised forms of PA governance to shifting local conditions.

The dichotomous framing of central, versus decentralised governance of PAs in the academic literature has contributed to a highly polarised debate that is unlikely to contribute meaningfully to political agendas, where trade-offs and compromise between competiting objectives is needed. Our results suggest a potential route to reconciling this debate by designing PA networks that situate centrally and decentrally governed PAs adjacent to each other, building on synergistic benefits of contrasting PA governance approaches to maximise positive social and conservation impacts.

At the time of writing, aspiriations for a global goal for nature conservation are beginning to coalesce at protection of 30% of land and ocean (Butchart et al., 2015; Tera Carta, 2020; Defra, 2020). This would require doubling of size of the current terrestrial PA estate (Waldron et al., 2020) and greater penetration of PAs into populated environments and trade-offs with other land-uses. For such an expansion to gain social and political endorsement, PAs must demonstrate an ability to meet both conservation and human wellbeing objectives. A possible landscape-scale model for expansion of existing protected area networks would be to suround strictly protected areas in remote locations with buffer zones of locally governed, sustainable and mixed-use PAs closer to settlements to accommodate multiple human objectives. This would create a flexible gradient of protection and landuse for meeting the needs of both people and planet and buidls on the synergistic interaction effects identified in this thesis.

WISP contributes an approach that PA managers can use to develop robust, site or landscape specific wellbeing assessments within a questionnaire format lasting between 30 minutes and an hour. This dramatically reduces the length of wellbeing questionnaires from the original list of 111 candidate wellbeing indicators identified and so makes progress in operationalising the wellbeing concept for practical use and mainstreaming into social assessments of PAs. Thereby WISP provides a useful tool for conservation practitioners to look beyond simple economic measures of PA impacts and better account for non-economic unintended social impacts of PAs.

Appropriate use of statistical simplification techniques outlined by WISP contributes to improved robustness of wellbeing measurement. However I acknowledge that implementation of these techniques requires significant quantitative capacity that may be a barrier to the methods' wider use among practitioners. This represents a methodological trade-off between rigor and simplicity of use.

A further potential limitation of WISP is that the identified indicators may be temporally specific and become obsolete in the future. This is of particular relevance for longitudinal applications seeking to assess change in wellbeing over time. The need to review and update wellbeing indicators is a natural consequence of the process of development e.g. the rising widespread availability of mobile phones and other technologies and has been shown to be particularly important for methods that focus on material indicators, such as the Basic Necessities Survey (Beauchamp et al., 2018). I acknowledge this limitation and the need to review the appropriateness of identified wellbeing indicators periodically. However I would suggest that non-material indicators, such as self-evaluations of mental health or the quality of social relationships are less prone to becoming obsolete and will change in accordance with shifting cultural norms. Therefore an indicative review period for wellbeing indicators might be every 5-10 years to ensure local relevance. This is beyond the time period of many short-medium interventions and so should not act as a barrier to the method's use.

Finally, my own positionality stemming from my cultural background from the global north and disciplinary origins in the natural sciences can be used to evaluate the balance achieved with respect to top-down and bottom-up perspectives, and quantative and qualtiative methods. The greater integration of local actor perspctives and emphasis of qualitative methods in the methodologcial approach of chapter four, compared with chapter two illustrates the reflexive growth I have undergone through this thesis in an effort to develop a balanced research approach. In particular intergration of local perspectives early on in the research process of chapter four helped to frame the impact of PAs as having complex positive and negative impacts. For example, many local actors emphasised the negative impacts of PAs and considered forests to be reservoirs of pests that damaged their crops. In contrast, thoeretical concepts such as ecosystem services emphasise the benefits that forests provide for human wellbeing. By drawing on both theoretical perspectives and the lived experience of study participants to inform the design of research questions, I was able to reflect the real-world complexity and trade-offs in human-nature relations and reduce any perspective bias in the research design. In this way I was able to progress from a position of being inclusive of local actor perspectives in chapter two, towards more fully co-creating the research design in chapter four by engaging local actors earlier in the research process and incorporating qualitative analyses more centrally in my interpretation of the results.

5.5 Emergent themes and recommendations for future research

Through the journey of this thesis, several themes have been developed, then refined in subsequent chapters. In the following section I synthesise three major themes, reviewing their emergence across the data chapters to crystallise the major insights of this thesis. I take a forward-looking approach to explore what these insights suggest for advancing future research endeavour. The chosen themes relate to different stages of the research process. Theme one relates to how research questions are framed from simplistic reductionist designs to complex systems perspectives. Theme two explores methodological tensions and complementary insights of quantitative and qualitative approaches. Theme three explores the links between conservation practice and research and opportunities for moving towards greater integration of the two in the form of action-oriented research.

5.5.1 Theme one: Framing research questions: from simplicity to complexity to support deeper understanding

The starting point for this thesis was exploration of the win-win concept in relation to PA governance. My literature review highlighted that win-wins are often an oversimplification of social and environmental outcomes that miss unintended negative impacts. The consequence of this narrow framing of PA impacts was illustrated through by social wins characterised in terms of financial returns missing negative impacts on community cohesion or inequalities in impacts (Gross-Camp, 2017; Hajjar et al., 2020), or environmental wins in terms of a strictly defined geographic area missing leakage effects across the wider landscape (Pfaff and Robalino, 2017). These insights from the literature suggested the need for a broader research framing that recognised the complex, interrelated social and environmental impacts of PAs.

Striving to understand the complexity of PA impacts became a central thread running through my thesis, with a progressively deeper exploration of social-ecological complexity explored through the data chapters. Chapter two sort to characterise social outcomes in terms of multidimensional human wellbeing. This approach recognised the need to understand the complexity of the human condition and to properly understand how PAs can serve a human-centred conception of development that accounts for the multiple human needs. Primarily building on the pioneering work of developmental economist Amarta Sen (1999), who characterised development as human flourishing, WISP was designed to support a holistic analysis that moved beyond a narrow focus on financial costs and benefits of PAs to a richer understanding of the multiple potential interactions between people and PA governance. The WISP method also deliberately incorporated local perceptions and values recognising that wellbeing can mean different things for different people in different contexts (Woodhouse et al., 2015; Woodhouse and McCabe, 2018). Chapter three then applied WISP and expanded the research framing by including both social and environmental impacts. Chapter four then took the exploration of complexity one step deeper by seeking to disaggregate the different pathways by which forest governance can impact human wellbeing and forest conservation. By moving from a simplistic understanding of win-wins, to increasingly fine grained insights about the study system provided increasingly actionable knowledge relevant for on-the-ground conservation governance.

The social-ecological systems approach was suggested by Ostrom (2009) as a means for understanding the complex interactions between social and ecological elements of linked social-ecological systems in the context of natural resource governance. In contrast to a reductionist research framing that seeks to isolate particular elements of a system, a holistic systems perspective recognises the multiple interactions between elements of a system in order to better understand the system as a whole (Figure 5.1). Over the data chapters of my thesis, a complex systems research framing has emerged as I sort first to explore the complexity of human wellbeing, then the multiple interacting connections with forest governance and environmental outcomes. This research framing has evolved through my research as an opportune way to better understand the complexity of the study system. This illustrates the necessity of embracing the complexity of socio-ecological systems, since PA impacts are diverse and may be caused by an interrelated constellation of contextual factors. Furthermore, it has implications for understanding how conservation research can better inform conservation practice. Systems perspectives recognise the interactions and feedbacks between system elements (Leslie et al., 2015). By looking explicitly at system interactions provides an analytical framework for exploring the opportunities and constraints for effecting transformative systems change by understanding the system elements with capacity to drive change and those which are limiting it. This combined understanding can be used to develop systematic approaches for effective interventions.

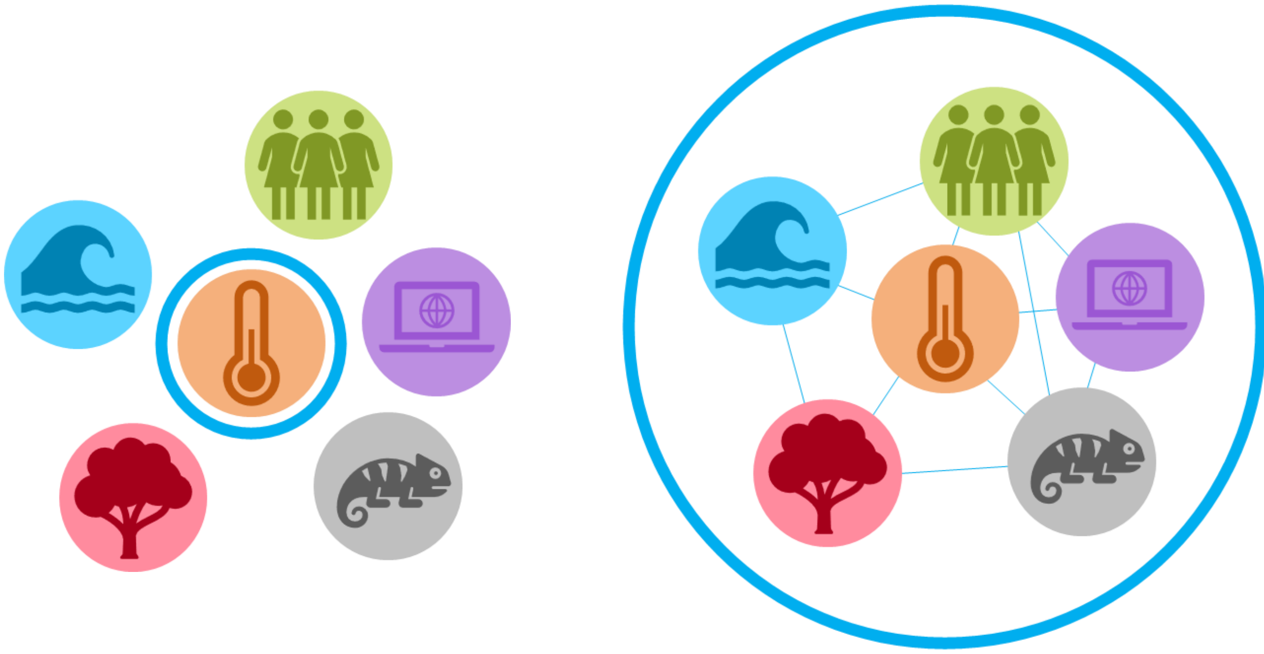


Figure 5.1 Illustration of contrasting reductionist research framing (left) with a holistic, systems research framing that explores the connections between different elements in a study system. Icons represent different system elements. Blue circle highlights the research focus.

PA governance has also evolved from simple governance structures towards greater complexity to meet multiple objectives (Agrawal, Chhatre and Hardin, 2008). Novel research framings and perspectives are needed to keep pace with and provide useful insights into the growing complexity PA governance arrangements. A potentially fruitful research framing for further unpacking PA governance complexity is provided by the emerging field of policy-mix analysis, which focuses on describing the interactions between governance approaches

(OECD, 2007; Flanagan et al., 2010; Barton, 2017). In the field of biodiversity conservation, a policy mix has been defined as 'a combination of policy instruments [governance approaches], which has evolved to influence the quantity and quality of biodiversity conservation' (Ring and Schröter-Schlaack, 2011). By explicitly recognising of the interplay between different governance approaches within research designs allows the exploration of nuanced questions, such as, what is the optimal mix of governance approaches to achieve multiple objectives (Minang and van Noordwijk, 2013). Policy mixes provide one example of integrated landscape analysis, which explores impacts and influences beyond the single intervention of interest (Reed et al., 2020). To properly account for diverse and unpredictable PA impacts, study designs should consider a broad framing of the study system, such as landscape-scale analysis, and consider employing a complex systems perspective to provide policy-relevant insights to inform how PA governance can positively impact the increasingly complex socio-environmental systems in which they are embedded.

The conservation discourse is increasingly moving beyond a strict dichotomy between people and nature to explore how the needs of people and nature can be jointly met within complex, multifunctional landscapes (Nagendra and Ostrom, 2012; Barton et al., 2013). To provide policy-relevant insights for this agenda, research designs that embrace landscape complexity and directly assess trade-offs in land-use decision-making are needed. Examples include modelling of conflicts between conservation and development objectives (Cusack et al., 2020) and how PA placement may impact other land-use options (Robinson, Albers and Williams, 2011). Analytical Approaches such as structural equation modelling provide opportunities for advancing the analysis of complex systems and understanding the interactions between multiple social and environmental objectives, which could not be achieved through more traditional reductionist research designs that rely on single response variables.

5.5.2 Theme two: Doing research: Methodological tensions, synergies and opportunities for complementary insights in a holistic research process

This thesis has demonstrated that both quantitative and qualitative approaches have methodological 'blind-spots'. However complementary methods can be used to shine a light on conservation challenges from contrasting perspectives to provide a more comprehensive understanding than either approach could individually. The methods used here for measuring human wellbeing, statistical matching and developing conceptual models advance approaches for using contextual understanding and local perspectives to translate generalised concepts into representative models of the situation on-the-ground. In this way this thesis has helped progress efforts to provide relevant insights for site-based conservation practice. These advances provide tools to help operationalise calls for more inclusive conceptualisation of conservation challenges through providing approaches for integrating of plural perspectives in decision-making (Pascual et al., 2021). Furthermore tools such as WISP that support the design of metrics that represent local priorities can help to increase the downwards accountability of conservation interventions to local communities and equitability of outcomes (Ribot, et al., 2015; Scheba and Mustalahti, 2015).

Top-down quantitative research designs, such as statistical matching approaches, provide a range of tools for isolating particular causal factors to accurately estimate the size of PA impacts. However they are potentially limited in their ability to explain why the observed impacts exist (Ferraro and Hanauer, 2014b). Exploratory qualitative approaches that unpack socio-ecological complexity are well suited to identifying contextual explanations (Poteete, et al., 2010). Therefore, qualitative approaches provide complementary insights to more top-down quantitative research approaches. Greater integration of quantitative and qualitative approaches and the emergence of conservation science as truly transdisciplinary represents an exciting methodological frontier with potential to elucidate new insights for conservation practice.

The integration of quantitative and qualitative research methods has been an emergent theme of this thesis. Quantitative and qualitative approaches prioritise different forms of scientific validity (see section 1.4 on the comparison between external and ecological validity). This can suggest different ways to approach an analysis and create a tension in the interpretation of research findings. The central theme I have employed for navigating this tension has been to aim to transparently communication where these methodological friction lines lie and to navigate a path between disciplinary specific extremes of analytical approach in an effort to find a pragmatic “middle-way”. This navigation between quantitative and qualitative approaches has occurred across all my data chapters. In chapter two for example, if I had followed a purest statistical simplification approach following the principle of parsimony, would have more stringently reduced the number of wellbeing indicators, thereby removing gender and site-specific indicators. However this would have reduced the ecological validity of the final set of wellbeing indictors, making the final list less recognisably relevant to local actors. As such I sort to prioritise inclusion of priority site and gender specific indicators to balance statistical parsimony and ecological validity. Likewise in chapter three, the statistical site matching approach was complimented with an expert panel review to evaluate the statistically matched villages for qualitative criteria, such as village political stability. The resulting adjustments provided a slightly poorer, though still valid statistical match. While this approach improved qualitative match for other subjective criteria that it was not possible to match for statistically. Finally in chapter four, I assessed the pathways linking forest governance to wellbeing and forest conservation, contrasting those pathways derived by normative statistical analyses of mean affect size, with pathways suggested by key informants and outlying village examples that did not agree with normative trends. In all cases the approach sort to achieve a balance between quantitative and qualitative analytical paradigms. The major implication of this negotiated research approach has been to provide a more well-rounded understanding of the study system. This middle-way approach may lower the scientific rigour from a disciplinary purest perspective. However, this pragmatic approach to methodological innovation suggests that methodological pluralism can provide a means for future research to develop holistic understanding of a study system that meets appropriate levels of rigour for external academic scrutiny and retain local relevance.

The utility of novel methods is maximised by making them widely accessible. In order to mainstream novel methods into widespread PA planning, monitoring and evaluation, it is important therefore to temper growing complexity of methods, with appreciation of the time and technical capacity constraints of PA personnel. Most PA and conservation managers focus at the spatial scales of particular sites or landscapes. In effect, case studies, where contextual understanding is important to explain outcomes (Poteete et al., 2010). This requires approaches that can provide insight into complex real-world scenarios. However, PA managers may not have advanced scientific degrees. Additionally, the majority of conservation professionals come from natural science backgrounds (Stephanson and Mascia, 2014; Sandbrook et al., 2013), have received limited training in social science methods and so non-technical approaches are needed. Therefore, future advances in evaluating PA governance might also focus on providing simple methods for asking complex questions. This would enable PA managers to own the investigative process, rather than rely on academic partnerships for rigorous evaluations.

My thesis has focused on backward-looking approaches to analyse existing interventions. However it should also be noted that the development of new forward-looking, scenario planning methods (e.g. Pereira, 2021; Capitani et al., 2019) to creatively imagine transformational new pathways to more sustainable futures also provide exciting opportunities for methodological innovation to support bridging research-practitioner divides by making the research process more participatory.

5.5.3 Theme three: Closing the research-practitioner divide: a trajectory towards research as action

My positionality at the start of this study was one common to the natural sciences, aiming to undertake objective research about the chosen study system, remaining as an independent observer with externally chosen research questions and methods of enquiry. However, layered on top of this natural science positionality is 1) my background as a conservation practitioner driving my desire to undertake research that informs conservation practice, and 2) my personal ethics of what is just and fair, driving a desire to undertake a research process that is equitable, recognising the knowledge of study participants and valuable contribution that their knowledge can make to the research findings. Exploration of these, perhaps contradictory, stances through my data chapters have led to an evolution in my research process and positionality through this thesis, prompting a trajectory towards action research across the data chapters.

Action research can be understood as the collaborative production of scientifically and socially relevant knowledge, transformative action and new social relations through a participatory process (Wittmayer and Schäpke, 2014). Action research therefore pays careful attention to the process by which research is conducted, giving space for shared learning of both researchers and research participants who are both considered to be capable of being affected by the research process. Rather than being independent from the research subject matter, action-oriented research recognises that the researcher is embedded within, rather than being separate from the systems they observe (Fazey et al., 2018). As such, the interaction between researcher and participant can be conceived as a form of intervention in which research is co-created by an interaction within the study system which might change the behaviour of both the researcher and the participant. The research process can therefore be recognised as a potentially powerful form of intervention whereby both the researchers and participants can 1) affect the research process and 2) be affected by it (Figure 5.2). In this way, action research can be seen as a move away from research that is extractive of the knowledge of study participants, to research that co-creates knowledge through dialogue between researcher and research participants, thereby democratising the research process (Fazey et al., 2020). It’s application to my own research also promotes the vital contribution that community members have made firstly as study participants. But also vitally, as conservation actors whose daily decision-making is an important form of conservation practice that the researcher has opportunity to influence through the research dialogue (Figure 5.2).

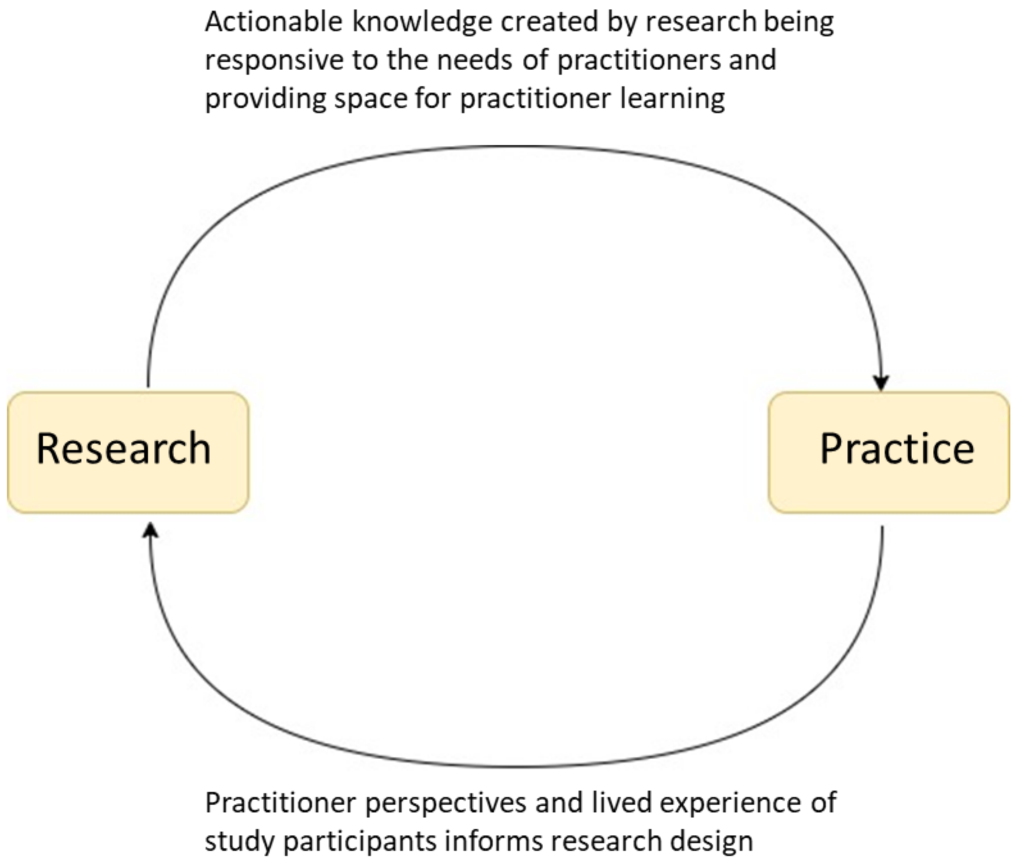


Figure 5.2 Illustration of the interaction between research and practice in action research.

The trajectory of this thesis towards action research has been made incrementally across the data chapters. In chapter two, the methodology for measuring human wellbeing was designed to be 'inclusive' of local actor perspectives of what is important for human wellbeing, with participant perspectives forming an important part of the research design. In chapter three, the research question was selected in response to a growing preference from the government of Tanzania towards recentralising forest governance and rising pressure on NGOs to demonstrate the positive social and environmental impact of decentralised forest governance. In this way the research question was selected to assess an important evidence gap. However the research was a retrospective assessment to understand 'what' the impact was and the results communicated back to study participants in feedback presentations. It thereby represents an example of a first order transformation research, remaining separate from the study system (Fazey et al., 2018). This is in contrast to second order transformation research, which recognises the active role the researcher can play in changing the study system they are embedded in (Fazey et al., 2018), with a more porous interface between research and practice and potential for co-designed research questions concerning 'how' conservation practice could be refined.

In chapter four, the perspectives of study participants were integrated earlier into the research process, with participant perspectives being used to design the complex causal model, which was tested through subsequent quantitative analyses. Causal pathways were derived both from theory and the lived experience of study participants. This more inductive approach thereby gave more equal weighting to the importance of perspectives from conservation practitioners (community members and government and NGO participants) and researchers, with multiple perspectives used to co-create the study's analytical framework (practice to research linkages; Figure 5.2). Integrating multiple perspectives and methodologies into the research process, combined with critical reflection, presents opportunities for making the research process more representative of diverse opinions and less dominated by the worldviews of researchers from global north (Baker, Eichhorn and Griffiths, 2019). Sayer et al., (2013) suggest 'the quality of decision-making is a function of the process by which the decision is reached'. Deliberation and critical review of how both research and conservation actions are designed are thus crucial for the delivery of effective outcomes and representative research. I suggest that more comprehensive consideration of contrasting perspectives, particularly local ways of conceptualising conservation-development challenges, present opportunities to increase the applied value of research by firstly, communicating the needs of diverse local actors. Moreover, by engaging conservation practitioners in dialogue and deliberation creates space and opportunity for practitioners to creatively reflect and redesign more equitable and effective conservation actions. By focusing more on research as a collaborative process between researcher and subject, research approaches can narrow the gap between research and action. Thereby applied research becomes more deeply embedded in local knowledge systems and has greater potential to contribute to the transformational changes required to achieve a more sustainable and just society.

The research process has transformed my own world view, particularly in terms of 1) recognising the central importance of promoting the agency of local actors to be the major conservation practitioners, or stewards vital for achieving effective conservation; 2) the complementarity between research methods for providing well-rounded, holistic insights, 3) the importance of local values, especially, conceptions of fairness in driving collective action and participation in conservation intervention; 4) the importance of paying attention to process, both in terms of understanding conservation governance effectiveness, and the opportunity for recasting research as a potentially powerful form of intervention. However my research has not fully achieved the feedback loop of action research suggested in Figure 5.2. Through the research process I have learnt to be more critically reflexive, to allow the research to inform my thinking (practice to research, Figure 5.2). But the research has likely not been a powerful form of intervention (translating research into practice, Figure 5.2). In the following section I explore the reasons for this and how future research might achieve more successful integration of research and practice.

Firstly, the partner NGO was brought into the investigative process relatively late during the development of research questions and study design. Therefore, the research questions were targeted to answer academic knowledge gaps, rather than being responsive to answering practice-oriented decision-making challenges with practical relevance to the partner NGO. Secondly, in an effort to maintain the neutral, objective positioning of my research relative to the NGO, my fieldwork funding was used to bring in external research assistants, rather than support staff time for the collaborating NGO. Given how overcommitted and time poor the NGO and charity sector at large are, this meant that time for information exchange was given by the NGO practitioners on a voluntary basis, rather than being structured into the project.

To improve the uptake of research findings by the NGO would require firstly, that practitioners are brought into the research design process when research questions are still at a nascent stage of formulation. This would ensure that the research design and final research questions are better aligned to answering the information needs and decision-making challenges faced by the applied partner institution. A second practical way to foster greater information exchange and opportunities for shared learning as part of the research process would be to direct fieldwork funding to supporting the time of personnel from partnering applied institutions to participate in the research process. This would help to legitimise the time invested in the research by practitioners, creating more space for deeper exploration of research matter and co-development of actionable research findings. Finally, To realise this transition towards action-orientated research, the research would need to be explicitly framed as action-oriented from the outset. The research design could then explicitly seek to maximise the role of the researcher as an active agent within the study system by designing the research process around opportunities for engagement with study participants, to foster shared learning and deeper joint-knowledge creation to encourage new understanding to take root and drive future action within the study system. This would require investing in embedding the research process within the structure of applied partner NGOs and other institutions through strong collaboration by creating a shared vision for the research, embracing greater shared control for the setting of research objectives and potentially, greater research direction uncertainty. The positionality of the researcher in this process would require a high degree of critical reflexive practice to decide on the one hand, how to intervene in the study system, while also balancing methodological rigour (Westling et al., 2014). Such a radical departure from the traditional positioning of the research as an objective, independent observer that sits apart from the research subject presents exciting opportunities for harnessing the power of knowledge creation as an effective means of behaviour change and conservation action.

5.6 Conclusion

In this thesis I have presented evidence showing that certified CFs can deliver positive outcomes for people and forest conservation. Though futher refinements, particularly in terms of developing successful economic pathways are still needed. I have demonstrated that no 'silver bullet' exists for delivering win-wins in tropical PA governance as trade-offs between social and environmental objectives were also observed. There is a need to embrace the complexity of PA governance challenges with interventions built on the combination of local understanding and sound science to develop equitable, evidence-based solutions. Quantiative and qualitative approaches represent fundamentally different approaches to scientific enquiry. However methodological innovation that seeks to transparently reconcile these tensions by recognising the complementarity between contrasting approaches can help to advance the discipline of conservation science by providing robust and novel insights. While creating more space for shared learning between researcher and practitioners within the research process would help promote the co-creation of actionable knoweldge and close the gap between research and practice. By integrating the contrasting perspectives of different actors and research methods, this thesis has contributed to advancing understanding of PA governance and a shift away from the panacea of win-win solutions, towards a more in-depth understanding of what works where, for whom and why.

Appendices

Chapter two supporting information

S2.1. Wellbeing Indicator Selection Protocol - detailed case study steps

Step 1. Identify candidate wellbeing indicators

Participatory wealth ranking exercises were undertaken to select individual questionnaire respondents and focus group participants in order to ensure a well stratified sample encompassing the socio-economic variation within communities (World Bank, 2005). We undertook two all-female and two all-male focus groups, to encourage uninhibited discussion and cross-validation of ideas between participants (Kitzinger, 1994; Macnaghten and Myers, 2011) and to explore gender-based variation (Schreckenberg et al., 2010).

All participatory exercises were undertaken in Swahili, the preferred language of study participants. The lead author, a native English speaker with basic Swahili language proficiency facilitated all sessions with translation support from a fluent Swahili speaking research assistant. The wealth ranking exercises were completed with the elected village chairperson and sub-village chairpersons from each village because their official responsibilities assured they had a good overview of the village populations. Through group discussion, participants were asked to describe the characteristics of households within their jurisdiction in terms of social characteristics, education, health, material assets and livelihoods, in order to identify locally relevant indicators of socio-economic status (Pietrykowski, 2016). They were then asked to place each household from their constituency into a particular wealth category (rich, average or poor). Focus groups were then held to determine local perceptions of wellbeing and identify appropriate indicators, in keeping with the understanding that wellbeing should be defined by the communities where wellbeing is being assessed (Woodhouse et al., 2015). Sub-village chairpersons responsible for organising focus groups ensured group composition had representation from each wealth category. Before commencing, free prior informed consent was sought from participants and the sessions were recorded with the participants' permission (Goulet, 1995, Braun and Clarke, 2013). Four focus group sessions were held (two per village) comprising 7-16 people (Morgan, 1988). If resources allow, future users of WISP might also consider additional stratification of focus groups by wealth category, age or ethnicity to reduce within-group variation and further promote participation of marginalised community members.

It is important to frame discussions around a sufficiently broad conception of wellbeing and be careful about how this is communicated when translating between languages (OECD, 2013). Following previous work on wellbeing in Tanzania we used the Swahili phrase 'hali ya maisha', which translates as 'conditions of life' (Gross-Camp, 2017). Focus groups lasted for approximately two to three hours with participants asked to describe important aspects of their wellbeing. The lead author used an open questioning style (Supplementary material Table S2) to encourage discussions to be led by participants and promote ecological validity, only guiding discussion to ensure that the three dimensions of wellbeing were covered (Gough and McGregor, 2007; Woodhouse et al., 2015; 2016).

Audio recordings of focus group discussions were translated from Swahili to English by research assistants. Thematic analysis of focus group transcripts was undertaken by the lead author to code discussions into the five domains of wellbeing based on the stated reasoning of study participants in relation to definitions used in Millenium Ecosystem Assessment (MEA, 2005, Supplementary material table S3). A second round of coding was then used to convert codes into candidate indicators, retaining the semantics used by participants as much as possible (Camfield et al., 2009). Based on candidate indicator codes, a simple content analysis was then undertaken, noting which focus group transcripts contained references to each indicator in order to identify which candidate indicators were site and gender specific (Silverman, 2011). Indictors discussed in all focus groups were defined as local priority indicators. Similar indicators (e.g. species of livestock, types of vehicle, household assets) were grouped together to form single indicators with multiple factor levels where a locally defined order of preference could be established (e.g. vehicle ownership: none – bicycle - motorbike - car/truck).

To ensure each wellbeing domain was well represented, if any domain had less than five indicators, then additional indicators were supplemented from related frameworks. In this case study we used In this case study, we identified that the health domain was under-represented and so one additional indicator was included by reviewing the OECD's How's Life and Sustainable Livelihoods frameworks as they were considered relevant to the context of the case study in Tanzania (OECD, 2013; DFID, 2000).

Step 2. Gather quantitative data

A stratified sampling method was used to represent both the socio-economic spectrum and variation within wealth categories. Individuals were selected at random from village registers, stratifying by sub-village, wealth category and gender to avoid gender bias (step one). Where culturally appropriate, sampling could also be stratified by age. However, in the rural context of this study, cultural norms were such that it was not considered appropriate to interview dependent household members and so only one adult female or male household member was interviewed per household. Interviewees were required to have lived in the village for five years or more to ensure sufficient local knowledge to answer all questions (Franks and Small, 2016). Specifically, we aimed for a minimum of five interviewees per gender, per wealth category, per village, consistent with previous rapid assessment methodologies (Schreckenberg et al., 2010), resulting in 90 interviews from the two villages (Supplementary material Table S4).

Step 3. Remove indicators with little variation

Next the spread of responses for each indicator was assessed, aiming to eliminate indicators with zero or uneven spread, that would give no helpful information on the variation of wellbeing present within communities. This and subsequent steps were completed using the statistical software, R 3.4.2 (R core Team, 2017). For categorical indicators, a minimum threshold of acceptable spread was employed, whereby factor levels with less than 10% sample size of the largest factor level were merged with adjacent factor levels. When categorical indicators failed to meet the minimum threshold for variation, even after merging to a minimum of two levels, they were either removed or where relevant, combined with closely related indicators, e.g. 'ownership of cattle' and 'ownership of pigs' would become 'type of livestock owned'. For continuous variables, histograms were plotted and if the spread of values was skewed, the data were square root or log transformed. For strongly skewed data, e.g. with multiple zero values, the continuous data were converted into categorical data, placing data into factor levels selected to meet the <10% spread threshold.

Step 4. Remove covarying indicators

The Holistic Wellbeing Index (HWI) aimed to represent all indicators in a single standardised index following principles of the Human Development Index (UNDP, 2017). First, values of each indicator with sufficient variation (step 3) were standardised, by expressing each value as a proportion of the maximum value of each indicator. Next, the mean value of standardised indicators was calculated for each wellbeing domain. Finally, the HWI was calculated as the mean of each domain mean. Thereby the index weighted each wellbeing domain equally and each individual had a HWI varying between 0 and 1 (UNDP, 2017; Neugarten et al., 1961). The HWI then provides a continuous variable to inform retention of covarying candidate indicators.

$$\text{Holistic Wellbeing Index} = \frac{\text{Material}(\bar{x}) + \text{Health}(\bar{x}) + \text{Security}(\bar{x}) + \text{Social relations}(\bar{x}) + \text{Freedom}(\bar{x})}{5}$$

High covariance between indicators was defined as Pearson correlation coefficient r ≥ |0.7|, and/or Variance Inflation Factors (VIF) ≥ 3 across all indicators (Zuur, Ieno and Elphick, 2010; Dormann et al., 2013). In the event of high covariation between indicators (Table S5 Supplementary material), we retained the indicator that had the strongest relationship with the HWI assessed by Pearson correlation test. By removing correlated indicators in this way, we produced orthogonal indicator subsets for each wellbeing domain and a multidimensional indicator subset comprising 30 indicators.

Step 5. Remove statistically redundant indicators

The uncorrelated list of indicators was used to model the holistic wellbeing index using a Generalised Linear Model (GLM) with Gaussian error; suitable for continuous variables with approximately normal distribution. We used backwards-forwards stepwise model selection (Venables and Ripley, 2002) based on Akaike's information criterion (AIC) to produce a minimum adequate model (Akaike, 1973). This process aimed to reduce the number of indicators without reducing the goodness of model fit. Diagnostic plots of model residuals were used to check for heteroscedasticity and outliers before accepting the minimum adequate model (Crawley, 2007).

Step 6. Validation

To ensure the stepwise selection process did not bias the reduced list against some wellbeing domains, we set a minimum threshold of two indicators per wellbeing domain. If any wellbeing domain contained less than two indicators, we reinserted additional indicators that had the highest univariate relationship with the holistic wellbeing index. To confirm that the final list of wellbeing indicators retained local relevance, the reduced list was compared with the priorities stated by community members during focus group discussions (step one). If local priority indicators had been removed in step five, they were reinserted in step six to create a final indicator list that promotes ecological validity.

Table S2.2. Wellbeing indictor coding examples from focus group transcripts

Wellbeing dimension & Indictor	**Focus group transcript**
Material wellbeing: Land and Livestock	Village 1, Male focus group. Male 1: In this village we are farmers. So the most important things are the amount of farmland we have and the number of livestock.
Health: Sickness indicator Definition: Number of days too unwell to work in last year	Village 2, female focus group. Woman 1: For me I need to have good health. Facilitator: How do you know that you have good health? Woman 1: First, decrease in having fevers and becoming sick. Facilitator: Please say a bit more about how you know that you are sick? Woman 1: If I become so sick that you can't even work, then it is a problem, like high frequent fever or vomiting. But those small pains, like small headache, they are not a problem.
Security: Livelihood diversity	Village 2, male focus group. Man 2: If you have a problem with your farm, like wild animals come and eat them, it can help if you have another job to make some money, like being a hired labourer for someone else in the village who has a large farm. So if you work there you can just get money or food depending on your agreement with the person who is giving the job. Man 3: Or for example I have a small shop in the village, and my brother, he drives his motorbike to transport people. So if either of us have a problem with our farms we can continue to live, no problem.
Social relations: Borrowing and Lending of land or money	Village 2, female focus group. Woman 4: In this village we are supporting each other. Facilitator: Please can you explain what you mean? Woman 4: Help from neighbours is important and we also help them. Facilitator: Please can you give an example? Woman 4: For example getting salts for cooking. Woman 6: Or if there is a big problem like with our farm, they can lend us some land to farm, or some money for medicine if we are sick.
Social relations: Recognition in the village: Perception of	Village 1, Male focus group. Man 1: It is important to have a good relationship with friends and other people from our village.

how much voice heard in community decision making	Facilitator: How do you know that you have a good relationship with them? For example, we have village assembly here. If the others are respecting me and listen to what I am saying, I know that there is a good relationship in this village.
Freedom of choice and action: Education	Village 1. Female focus group. Woman 3: If I have education I can go and live anywhere I want and do whatever activities that can support my children and then send term to school also.
Freedom of choice: Forest access	Village 1, female focus group. woman 3: I want the forest to be open. Facilitator: What do you mean by saying you want the forest to be open? Woman 3: We want to go to the forest free without any obstacles; we want to go there freely to collect fire wood.

Table S2.3. Definitions of the five wellbeing domains (MEA, 2005)

Domain	**Definition**
Material wellbeing	Adequate livelihoods, sufficient nutritious food, shelter, access to goods.
Health	Strength, feeling well, access to clean air and water.
Security	Personal safety, secure access to resources, security from disasters.
Social relations	Social cohesion, mutual respect, ability to help others.
Freedom	Opportunity to be able to achieve what an individual values doing and being

Table S2.4. Summary of sampling across gender and wealth categories within villages.

	Wealth category & gender						
	Poor		**Medium**		**Rich**		**Total**
Village	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**	
Mang'ula B	9	6	7	9	6	8	45
Udekwa	7	7	7	10	8	6	45
Combined totals	16	13	14	19	14	14	**90**

Table S2.5. Pairwise Pearson correlation coefficients of all 56 indicators with sufficient spread. Yellow 0.3 – 0.49, Orange 0.5 – 0.69, Red ≥ 0.7.

[Please see excel file attached seperately due to large table size beyond margins of A4]

Table S2.6. All 56 indicators with sufficient spread for the assessment of human wellbeing (step three). [1] Indicator supplemented from the literature rather than determined from focus groups. r^2 proportion of variation explained in the holistic wellbeing index.

Indicator	Description (units)	Transformation	r^2	Step excluded
Material				
Household wall materials	Ordinal categories (1=mud, 2=mud bricks 3=concrete bricks, 4=plastered)	NA	0.08	
Household roof materials	Ordinal categories (0=grass roof, 1=corrugated iron roof)	NA	0.01	4
Number of buildings in household compound	No. of buildings owned by household	Combined factor levels 3 and above	0.03	4
Vehicle type owned	Ordinal categories (0= nothing 1= bicycle 2= motorbike	Combined factor levels 2 and 3	0.10	4

	3=car highest score counted)			
Household assets	Integer from 1 to 7 of combined total of identified household assets composed of	Integer from 1-7 of total of household assets composed of electricity, solar light, television, radio, phone, plough, tractor	0.17	
Livestock	Ordinal categories for most valuable livestock species owned (none, chickens, goats/pigs, cattle)	Combined separate indicators for different livestock species	0.18	5
Tractor use	Ordinal categories for use of a tractor in farming activities: 0= no 1= rent 2 = own	Combined factor levels 1 and 2	0.00	4
Banking	Use of formal banking facilities (yes/no)	NA	0.12	
Money received from relatives	Money received from relatives in the last five years. Binary variable	NA	0.01	5
Financial savings	Ordinal categories (0, 1 – 99,999, >100,000 units in Tanzanian Shillings)	Categories made from continuous data 0=0 shillings	0.45	

		1=1 - 99,999 2=>100,000		
Total individual income	Ordinal categories. Annual income in Tanzanian Shillings calculated as the combination of income from disaggregated sources of pastoralism, cultivated agriculture and other paid employment.	Categories made from continuous data 0=0 1=1-9,999 2=10,000-99,999 3=100,000-200,000 4=>200,000	0.37	4
Number of different forest products collected	Summed integer from firewood, timber, thatch, sand, medicinal herbs, wild vegetables, fruits, wild honey	NA	0.13	5
Fruit collection travel time	Time (minutes) taken to walk to collect wild fruits	Square root transformation	0.06	4
Firewood travel time	Time (minutes) taken to walk to collect firewood	Square root transformation	0.02	4
Quality of available forest	Likert scale 1-5 from very degraded to very good condition	1 and 2 combined	0.00	5

Water travel time	Time (minutes) taken to walk to collect water	Converted two separate indicators from wet and dry season to a single categorical variable: 0=Do not need to travel, 1=need to travel in dry season, 2=need to travel in wet and dry season	0.05	
Fallow land area	Area of fallow land farmed (acres)	Converted to binary variable (have/ have not)	0.07	4
Fallow land tenure	Tenure categories for fallow land ownership: 0= no fallow land, 1= rent fallow land, 2 = own fallow land	NA	0.05	4
Agricultural land area	Area of agricultural land used by household (acres)	Log transform	0.07	4
Agricultural land tenure	Tenure categories for agricultural land used by household 0 = none, 1= rented, 2 = owned	Categories 0 and 1 combined	0.01	5

Forest tenure	Ordinal categories. Tenure categories for forest land (plantation or other silvicultural practices or natural) held by household 0 = none, 1= rented, 2 = owned	Converted to binary indicator for ownership of some form of forested land, either plantation forest, agroforestry, natural forest	0.11	4
Total acres	Total area of land used by household	Log	0.14	
Commercial pastoralism	Income from pastoralism activities	Converted to binary variable	0.06	4
Commercial agriculture	Income from commercial cultivated agriculture	Converted to binary variable	0.01	4
Domestic	Binary variable for time spent in care for dependents	NA	0.05	4

Business	Income from ownership of a small business	Converted to binary variable	0.20	4
Beekeeping	Income from beekeeping	Converted to binary variable	0.04	4
Paid labour	Income from paid work such as informal labouring	Converted to binary variable	0.09	4
Health				
Health insurance	Binary indicator for presence of a health insurance policy	NA	0.26	
Sickness	Number of days too unwell to work in last year	log	0.06	
Missed meals	Likert scale 1-4 of how often meals are missed	NA	0.22	4
Dietary diversity[1]	Number of different food groups consumed in the last week	NA	0.06	5

Clean drinking water	Indicator for access to clean drinking water of different sources; Integer with 8 options from piped into home (1) to unprotected river (8)	Converted to binary variable; piped into home or not.	0.02	5
Social relations				
Partner	Binary factor 2 levels	NA	0.02	4
Resident	Number of years residing in this village; 3 level ordered factor: 1) 5-10 years 2) 10-20 years 3) >20 years	NA	0.01	4
Lending	Binary response stating whether money or land was lent in last year (yes/no)	NA	0.27	
Borrowing	Binary response stating whether money or land was borrowed in last year (yes/no)	NA	0.16	4
Household decisions regarding children	Perception of how much voice heard in household decisions concerning children's education. Likert scale 1-5	Categories 1, 2 and 3 combined	0.01	5
Household decisions regarding money	Perception of how much voice heard in household concerning money. Likert scale 1-5	Categories 1, 2 and 3 combined	0.01	4

Participation in community meetings	0= do not attend 1= attend but don't speak at meetings 2= attend and speak at meetings	NA	0.12	4
Perceived recognition in the village	Perception of how much voice heard in community decision making. Likert scale 1- 5	categories 4 and 5 merged	0.18	
Labour support to others	Binary	NA	0.02	5
Community Cohesion	Perceived level of community cohesion and mutual support. Likert scale 1- 5	NA	0.02	5
Security				
Provision for dependents	Likert scale indicating perceived ability to provide for dependents	NA	0.33	
Provision for self in old age	Likert scale indicating perceived ability to provide for oneself in old age	NA	0.42	
Number of livelihoods	Total of different livelihood activities	NA	0.24	
Theft security	Likert scale indicating perception of security from theft	NA	0.07	
Water shortages	Number of months per year	Converted to ordinal categories 0=0 months per	0.00	4

		year 1=1-3 months 2=>3 months		
Savings group	Binary indicator for participation in a group savings scheme	NA	0.00	5
Freedom				
Livelihood satisfaction	Likert scale indicating satisfaction with livelihood opportunities	NA	0.05	
Overall quality of life	The final question. Likert scale indicating overall life satisfaction considering all questions asked	categories 4 and 5 merged	0.22	
Forest access	Likert scale indicating satisfaction with access of forest resources	combined 3,4,5	0.06	
Education	Ordinal categories for highest level completed (no formal education – University)	NA	0.25	
Access to traditional cultural sites	Perceived. Likert scale 1- 5.	merged categories 4 and 5	0.03	4
Access to land for settlement	Perceived. Likert scale 1- 5.	merged categories 4 and 5	0.04	5

Access to agricultural land for cultivation	Perceived. Likert scale 1- 5.	factor levels 4 and 5 merged	0.02	4

S2.7. Correlations between wellbeing domains

We assessed correlations between wellbeing domains in order to evaluate the need for multiple wellbeing domains to adequately explain variation in wellbeing. This is a core assumption underpinning the shift away from traditional social assessments based on a narrower suite of material indicators (Alkire and Foster, 2011). We assessed correlations between wellbeing domains by first calculating the mean value of standardised indicator values for each wellbeing domain. Then pairwise comparisons of Pearson correlation coefficients were used to evaluate variation explained (r^2 values) between domain means.

Weak-to-moderate correlations were observed between wellbeing domains (Table S8). The percentage of variation explained between pairwise domain comparisons was highest between the material and freedom domains of wellbeing (30.7%) and lowest between social relations and health (1.8%).

Table S2.8. Pearson correlation (Proportion of variation explained, r^2) between standardised mean wellbeing indicator values from the five domains of wellbeing.

Health	0.228			
Social relations	0.157	0.018		
Security	0.223	0.178	0.110	
Freedom	0.307	0.182	0.098	0.100
	Material	**Health**	**Social relations**	**Security**

Additionally, we calculated how much variation in the HWI was explained by orthogonal indicator lists made from individual wellbeing domains, compared to the final indicator list that contained indicators from all five domains. Following Burnham and Anderson (2002), we treated indicator lists as equivalent if within two AIC. This test served to evaluate the loss of information that would be incurred if a wellbeing assessment employed a narrower conception of wellbeing using indicators from a single domain compared to all five domains.

The final indicator list explained 91 % deviance in the holistic wellbeing index. Comparison of this set of indicators with orthogonal indicator lists from individual wellbeing domains showed that no list of indicators from a single wellbeing domain could explain equivalent deviance (Table S9). After the preferred list, the best performing unidimensional indicator lists were from the security and material domains of wellbeing, which each explained approximately three quarters of deviance in the holistic wellbeing index (Table S9).

A large part of the challenge of measuring wellbeing relates to its very broad definition encompassing multiple dimensions (Gough, McGregor and Camfield 2007). Some scholars have criticised the need for multidimensional social assessment as unnecessary, by demonstrating interdependence between social indicators (McGillivray, 1991; but also see Qizilbash, 2001). In our case study, material wellbeing most consistently explained variation in the other dimensions. But the proportion of variation explained even between the two most correlated wellbeing domains was less than one third (range 2% - 31%).

Furthermore, the list containing indicators from all wellbeing domains explained more variation in the holistic wellbeing index than any one-dimensional list. Together these findings support calls from the human-cantered development literature that to adequately measure the complexity of wellbeing it is necessary to go beyond simplistic material and economic indicators (Sen 1999; Nussbaum, 2000; OECD, 2013). However, the 31% correlation between material wellbeing and freedom also demonstrates the redundancy between domains. Beauchamp et al. (2018) use a Venn diagram to illustrate this partial overlap of wellbeing domains. Our correlation analysis supports this analogy of partial overlap between domains and emphasises the potential for simplification to remove correlated (overlapping) indicators.

Table S2.9. Relative explanation of the holistic wellbeing index by indicators of alternative wellbeing dimensions, versus the stepwise reduced list of 17 indicators. %D = percentage deviance explained. ΔAIC = the difference in Akaike Information Criterion scores from the best model.

Wellbeing domain	**% D**	**ΔAIC**
All domains	91	0
Material	73	85
Health	60	108
Social relations	55	127
Security	76	68
Freedom	61	108

Chapter three supporting information

S3.1. Response variables

S3.1.1. Human wellbeing response variable

We adopt the definition developed by the Wellbeing in Developing Countries research group, which defines wellbeing as 'a state of being with others, which arises where human needs are met, where one can act meaningfully to pursue one's goals and where one can enjoy a satisfactory quality of life' (Gough and McGregor, 2007). This broad framing contains three dimensions; objective, subjective and relational dimensions (Boarini, Kolev and McGregor, 2014) and five domains (1) Basic material for a good life - hereafter referred to as material wellbeing, (2) Health, (3) Social relations, (4) Security, (5) Freedom of choice and action (hereafter referred to as freedom; Narayan et al., 2000). It is not possible to generalise subjective indicators of wellbeing across households (Franks and Small, 2016). We therefore assessed the wellbeing of individual people, rather than households.

There is theoretical agreement that conceptions of human wellbeing are social constructed and differ from region to region (Martin et al., 2014; Wood et al., 2018). Therefore, participatory methods should be used to include the views of those individuals whose wellbeing is being assessed (Camfield et al., 2009; Sterling et al., 2017). Following Loveridge et al., (2020), we identified 25 locally relevant indicators of wellbeing representative of the five domains of wellbeing through four focus groups in two target communities in the study region (Table S6). The mean of standardised indicators of each domain of wellbeing was used to create domain indices. The mean of all domain indices was then used to calculate a Human Wellbeing Index (HWI) to represent all indicators in a single, standardised index that weights each domain equally following principles of the Human Development Index (UNDP, 2017; Eq.1).

$$\mathrm{HWI} = \frac{\mathrm{Material}(\bar{x}) + \mathrm{Health}(\bar{x}) + \mathrm{Security}(\bar{x}) + \mathrm{Social\ relations}(\bar{x}) + \mathrm{Freedom}(\bar{x})}{5}$$

Eq.1. Human Wellbeing Index (HWI), where $\bar{x}$ is the mean value of standardised indicators from each wellbeing domain (reproduced from Loveridge et al., 2020).

Within each matched treatment and control village, we sampled circa. 50 individual respondents through stratified, random sampling using village registers. Participants were categorised as either elites / non-elites, and by gender, with equal number of female and male respondents (Table S7). Village elite was defined as being a member of either the Village Natural Resource Committee or

Village Council, since these positions provide additional responsibility and decision-making power over forest governance and village budget decision-making, so potential for elite capture of benefits (Zafra-calvo, 2018). Due to the proportion of elites to non-elites in villages, we surveyed 10 elites and 40 non-elites per village and non-elites were further sub-categorised into rich and poor to ensure good socio-economic variation within villages (Schreckenberg et al., 2010). We used a questionnaire completed on digital tablets by fluent Swahili speaking research assistants after a month of field training and review by the lead author (Hartung et al., 2010).

Table S1. Wellbeing indicators. Likert scales are all on a scale of five factor levels from very low to very high unless stated. Transformations: [1] Continuous data transformed to categorical data at specified intervals, [2] log transformation.

Indicator	Description	Mean (CFs, control)	Range
Material			
Financial savings [1]	Ordinal categories (Tanzanian Shillings) 0, 1 – 99,999, 100,000 – 999,999, ≥ 1,000,000	104,000 TZS, 83,000 TZS	0- 2,500,000
Financial debt [1]	Ordinal categories (Tanzanian Shillings) 0, 1 – 199,999, 200,000 - 999,999, ≥ 1,000,000	91,000 TZS, 101,500 TZS	7,000,000
House construction	Ordinal categories walls constructed from (1=wooden poles, 2=mud, 3=mud bricks, 3=concrete bricks, 4=bricks and plastered)	2.29, 2.2	1-5
Vehicle ownership	Ordinal categories 0=none, 1=bike, 2 = motorbike, 3= car	0.88, 0.94	0-3
Assets	Integer from 1-6 of total of locally defined key assets	2.0, 2.1	1-6
Land [2]	Total area owned (acres)	8.4, 8.9	0-245
Livestock	Ordinal categories for most valuable livestock species owned (none, poultry, goats/pigs, cattle)	0.74 ,0.85	0-3
Health			
Access to clean water	Likert scale of cleanliness of drinking water	3.22, 3.07	1-5
Sickness [1]	Binary response any days too unwell to work in last year (0 = Y, 1 = N)	0.80, 0.88	0-1
Health insurance	Binary response (0 = N, 1 = Y)	0.29, 0.25	0-1
Personal dietary diversity	Consumption of key food groups in last week (Leroy et al., 2015)	9.5, 9.4	1-15
Mental health	Likert scale on level of personal worry and anxiety	2.76, 2.51	1-5
Social relations			
Borrowing of resources	Binary response stating whether money or land was borrowed in last year (yes/no)	0.43, 0.45	0-1
Lending of resources	Binary response stating whether money or land was lent in last year (yes/no)	0.41, 0.39	0-1
Social connectedness: Friendship assistance	Likert scale indicating if assistance is received from friends and neighbours when needed	3.50, 3.61	1-5

Community cohesion	Likert scale indicating the strength of relationships within the village based on level of observed conflicts	4.11, 4.17	1-5
Security			
Risk of theft	Likert scale perception of security from theft	3.86, 3.61	1-5
Risk of violence	Likert scale perceived security from personal violence	4.51, 4.54	1-5
Personal food security	Food insecurity experience scale in last year, (Cafiero et al, 2017). Scale inverted so high = more secure	3.0, 2.7	0-8
Provision for dependents	Likert scale perceived ability to provide for dependents	3.17, 3.00	1-5
Number of livelihoods	Total number of different livelihood activities	3.31, 3.22	0-9
Freedom			
Education	Ordinal categories for highest level completed (no formal education – University)	4.02, 3.84	1-8
Livelihood satisfaction	Likert scale indicating satisfaction with livelihood opportunities	2.93, 2.80	1-5
Marriage	Likert scale indicating freedom to choose marriage	4.24, 4.18	1-5
Overall quality of life	Likert scale indicating perceived quality of life	3.09, 3.04	1-5

Table S2. Summary of wellbeing questionnaire sampling between socio-economic groups. Categorisation of non-elites was undertaken by sub-village chiefs using village registers and then randomly sampled at a ratio of 2:1 poor to rich as trial surveys suggested this was the proportion of village residents in each socio-economic category. In Tanzania villages are sub-divided into 3-5 sub-villages distributed around a central sub-village. To ensure good spatial representation of the sample data at least one day of survey effort was undertaken in each sub-village.

	Non-elite		**Elite**	
	Poor	**Rich**	**Village Natural Resource Committee**	**Village Council**
Female	250	141	42	41
Male	241	148	43	49

S3.1.2 Conservation effectiveness response

NDVI saturates in high vegetation density conditions, making it difficult to detect small-scale impacts of degradation in healthy, high biomass forests (Pfeifer et al. 2016). However it correlates with ground vegetation biomass, and productivity under low to medium vegetation density conditions (Oindo and Skidmore, 2010; Pettorelli et al., 2005; Pfeifer et al., 2016) allowing detection of large-scale degradation in forest ecosystems that feature less dense canopies, like the Miombo woodlands studied here. NDVI was therefore deemed an appropriate proxy measure of conservation effectiveness for this study landscape.

We define positive conservation outcomes as all cases where NDVI change was positive, irrespective of previous trends. We set this to be the case for pixels in which (i) NDVI was declining prior to the study, (ii) NDVI change was constant prior to the study, and (iii) NDVI change was increasing prior to the study. For case (i): we are conservative. It could be argued that slowing down of NDVI loss is already a success. We disagree with this assertion in our analyses, focusing on evidence of forest recovery as the minimum requirement for defining conservation success, rather than simply slowed rates of degradation. For case (iii): one could argue that maintaining NDVI gains is a positive outcome nevertheless.'

We mapped NDVI from Landsat 8 OLI Surface Reflectance data (~30 m resolution) for the year 2014 and 2019 in the Google Earth Engine (Gorelick et al., 2017). For each year, we selected all Landsat 8 OLI Surface Reflectance images with less than 20% cloud cover, filtered out cloud and cloud shadow pixels, and averaged the pixel value of the red and near-infrared bands. From these composite images, we calculated NDVI values of each pixel for 2014 and 2019 as the mean of all the non-cloud cover images for each year. By taking the annual mean we take account of seasonal variation in NDVI. This approach also served to average away any remaining noise after the cloud cover exclusion step to account of stochastic events such as wildfires. We validated NDVI estimates for a sample of known habitat types ranging from low NDVI values in agricultural land to high NDVI values in mature forest using field data records of 50 GPS locations collected by the lead author in 2019. The final response variable, change in NDVI was then computed as:

$$NDVI\ change = \ log\frac{NDVI_{2019}}{NDVI_{2014}}$$

Eqn. 2. Calculation of NDVI change. The log transformation serves to adjust for asymmetric scaling of ratio data.

S3.2. Confounding variables and transformations

Confounding variables were selected based on previous work on the social and ecological impacts of protected areas and discussion with PA practitioners with in-depth knowledge of the study region (Andam, Ferraro, Sims, Healy, & Holland, 2010; Joppa & Pfaff, 2010). Tables S4-S6 lists the confounding variables used in social and environmental matching analyses and justifications for inclusion.

Calculation of confounding variables and data sources

Distance to different geographic features was calculated using ArcMap (ESRI, 2019). The only surfaced road running through the study area was treated as the main transport link connecting villages to major markets in district capitals and regional capital cities. GPS tracklog recording collected by the lead researcher was used to provide accurate spatial data of this road. Urban centres were defined as GPS locations of government offices in both district capitals. Village centres were identified using a combination of official district maps at a scale of 1:100,000, local geographic survey protected area maps provided by Tanzania Forest Service (certification dates of official maps were in 2007) cross-checked with Google Earth images to confirm settlement locations. Protected area boundaries were taken from the world database of protected areas (WDPA, UNEP-WCMC and IUCN, 2019). Some changes to PA boundaries have been made prior to 2014 (the start date of our study), but were not yet updated in the WDPA. These revised boundaries were taken from Tanzania Forest Service protected area maps and cross referenced with GPS locations of boundary markers provided by local NGO Mpingo Conservation & Development Initiative. Village populations were provided by village 2012 census data (TNBS,2012). District boundaries were also provided by the Tanzania National Bureau of Statistics (TNBS, 2012).

Slope and elevation were based on the Shuttle Radar Topographic Mission (SRTM) at 30m resolution with slope calculated using ArcMap's Slope tool. Forest resources were subject to anthropogenic pressure both from local settlements and from remote demand centres. Therefore the pressure on forest resources depends on 1) The amount of forest resources available, 2) the number of people in the surrounding area and 3) how far away those people are from the forest resources (Platts, 2012). The pressure on forest resources indicator is given as the amount of tree cover (Hansen, 2013) in a 10 km radius each pixel divided by anthropogenic pressure. Anthropogenic pressure is calculated as population density (World population database) weighted as a function of distance (Platts, 2012). Mean annual precipitation data were obtained from the WorldClim Global Climate data (1950-2000) provided at 30 arc-seconds resolution (~ 1km^2). Soil variables of soil pH, cation exchange capacity (cmol(c)/kg) and soil clay content (kg/dm³) at 250m^2 resolution were chosen through discussion with soil experts to best represent important divers of agricultural suitability (ISRIC, 2019).

Transformations procedures

Predictor variables with skewed distributions were transformed with either a squared or log transformation to reduce kurtosis and associated problems of outliers disproportionately influencing the analysis (Zuur, Ieno, & Elphick, 2010, Tables S3, S5). Then since the range of values for predictor variables varied widely due to differences in the units of measurement, all continuous predictors were Z transformed by subtracting each value by the mean of that predictor and then dividing by the standard deviation (Kuhn, 2019).

S3.3. Matching

PAs are established non-randomly in space for strategic reasons, such as biodiversity values and remote locations with limited economic potential (Joppa and Pfaff, 2010). Consequently PAs are systematically different from the wider landscape in terms of factors, other than governance practices that are likely to influence wellbeing and conservation effectiveness (Ferraro, Hanauer and Sims, 2011). All matching analyses were performed in R using the "Matchit" package in R version 3.5.4 (R core team, 2020; Ho et al., 2007). Nearest neighbour matching was employed to select one control unit with the closest propensity score to each treatment unit. Combinations of uncorrelated confounding variables were used to select treatment and control groups that maximised the number of confounding variables with post-matching standardized mean differences below 0.25 (Stuart, 2010). We also used a caliper function of 0.25 to set a maximum bound on the difference between treatment and control units (Schleicher et al., 2020).

S3.3.1 Wellbeing matching

We defined treatment as certified CFs that have been established and participated in forest certification for at least five years. In this way all certified CFs had completed at least one five-year management cycle of harvesting, revenue disbursement and external audit, so provide a well-established example this governance approach. All treatment villages were within 10 km of NFRs. We defined control villages as villages within 10 km of NFRs but not within 10 km of CFs. We excluded potential control villages that overlapped a 10 km buffer around the treatment villages to minimise the possibility that control villages were affected by spill-over effects of the treatment sites (Mitchell, 2017). We also excluded control villages with other known NGO interventions such as Payment for Ecosystem Services projects or that had partially begun the CF designation process.

Matching of confounding variables was undertaken at the village scale as this was the smallest scale for which data on confounding variables were available across the study region (Table S3). However in post-matching mixed effects models we were able to use values of confounding variables at the finer spatial scale of individual household locations once these had been collected by GPS during wellbeing questionnaire interviews.

As a final matching validation step we reviewed the list of matched villages with a CF expert panel (Mitchell et al., 2018). This final step is important because various community characteristics influence the location of CFs (Ostrom, 2000), several of which are qualitative criteria for which it is not possible to undertake statistical matching. But are non the less important to account for to achieve accurate matches. In this expert panel validation step, we assessed the quality of matches against three additional key criteria used to select villages suitable for CF establishment, but did not have comprehensive data for the entire study landscape. These were village land area, forest economic potential and local political stability (Table S5). The partial data available were used to assess differences between the distribution of these confounding variables in statistically selected treatment and control villages. If differences were identified, the balance was adjusted through a panel discussion drawing on expert knowledge of the study area to select alternative sites (Martin et al., 2011). To confirm that the adjusted match still adequately balanced the statistically matched confounding variables, plots of the mean standardized residuals were compared with the statistically selected match (Figure S1). Adjustments to the statistical match based on this validation step are listed below.

Table S3. Confounding variables used in village-scale matching and post-matching mixed effects models at the scale of individual household locations. Superscript number ^{1} indicates pairs of variables correlated at $r > 0.7$. ▮ Indicators specific to the village matching analysis and not used in the pixel matching analysis. The total number of villages within the study region was 135. Of these, 62 villages were within 10 km of NFRs and classified as potential control villages and matched to the nine treatment villages. The distance of 10 km conforms with previous assessments of the threshold within which protected areas exert social and ecological impacts (Naidoo et al., 2019b). * indicates square root transformation, + indicates log transformation. Post-matching analyses used confounding socio-environmental variables at the household scale collected by GPS during interviews.

Variable	Justification
Distance to major roads *[1,]	Roads provide access to markets which are key drivers of livelihood opportunities, transport costs and viability of selling produce.
Distance to urban centres[1]	Urban centres provide important markets. Urban centres were defined as the district capitals.
Distance to village centre *	Villages are the smallest government administrative unit in Tanzania and are composed of several sub-villages that may be located 5-10 km from the main village. Village services such as education, health, water and local markets tend to be clustered at the main village centre. Therefore people living in more remote locations within villages may have less access to village services, negatively impacting their wellbeing.
Distance to PA boundary *	People living closer to a PA boundary may be more impacted by the PA than those living further away. Since all villages included as potential matches were within 10km of PA boundaries it was not necessary to include this variable in the initial village matching phase. However this variable was included in subsequent linear mixed model analyses using distance from household location to PA boundary.
Village population ■	The number of people living in each village is likely to influence the provision of local amenities, which will influence wellbeing.
Slope +, elevation *	The dominant livelihood in this region is cultivated agriculture, therefore crop productivity is a key driver of development. Slope and elevation are key determinants of agricultural suitability of land for conversion of forest to other uses (Gentle and Narayan, 2012).
Pressure on forest resources +	Timber and non-timber forest products were an important subsistence and commercial livelihood activity in the study region. Therefore the availability, or more specifically availability per unit of population i.e. pressure on forest resources influences livelihood activities (Platz, 2012). Forest pressure was calculated as forest cover (Hansen, 2014) within a 10 km radius divided by population density (world pop, 2015) weighted so that more remote populations exert less pressure on forest resources using a sigma function of 5 so that relative

	weight reduces to zero at approximately 10 km to account for forest use by local users.
District	District is a key unit of government administration, with different government staff responsible for delivering management in different districts and differing approaches for management arising.
Precipitation	Rainfall will influence the agricultural suitability of land and yield of crops. 30 second resolution ~ $1km^2$
Soil composition	Soil composition influences what crops may suitably be grown, soil fertility and agricultural productivity and potential yields of different soil types with consequent wellbeing benefits.

Table S3.4. Confounding variables matched by expert review specifying variables influencing location of CFs, but for which incomplete data was available across the study area.

Variable	**Justification**
Village land area	At the commencement of the study it was not possible to obtain village land maps for the entire study area since this was a politically sensitive topic and several villages were in land boundary disputes with neighbouring villages and national government. However this was an important variable influencing CF location, as CFs were preferentially located in villages with potential to establish larger CFs. So this variable was included in the qualitative matching process using the partial information available and knowledge of the expert panel to compare and balance the distribution of village area between treatment and control groups.
Forest economic potential	Economic potential is assessed by the NGO in terms of the size of village forest land available and a field assessment to undertak a species inventory. The forest must be well stocked with desirable hardwood species.
Political stability	Local political stability was defined in relation to the village demonstrating a commitment to achieve sustainable forest governance. Several different forms of evidence are used to assess this commitment including a) receiving a letter from the village chairperson requesting support, b) village meetings and a village vote to assess community cohesion and village wide agreement to designate a community forest c) the village having made progress towards completing a village land-use plan.

List of organisations represented in the expert panel review, each with one representative

District Forest Officer, Kilwa district government,

District Forest Officer, Rufiji district government,

Tanzania Forest Service, Kilwa district government

WWF, Forests Programme

MCDI, Forest Certification Manager

Alterations made to the statistically selected match following expert review:

An additional village was added to the control group that had recently been assessed to have high economic potential and a large village land area.

One proposed control village was excluded due to known poor local political stability and so was not considered equivalent to the treatment villages. This village was replaced with a new control village that had recently submitted a request for assistance to MCDI to establish a CF, thereby demonstrating commitment to sustainable forest governance equivalent to treatment villages.

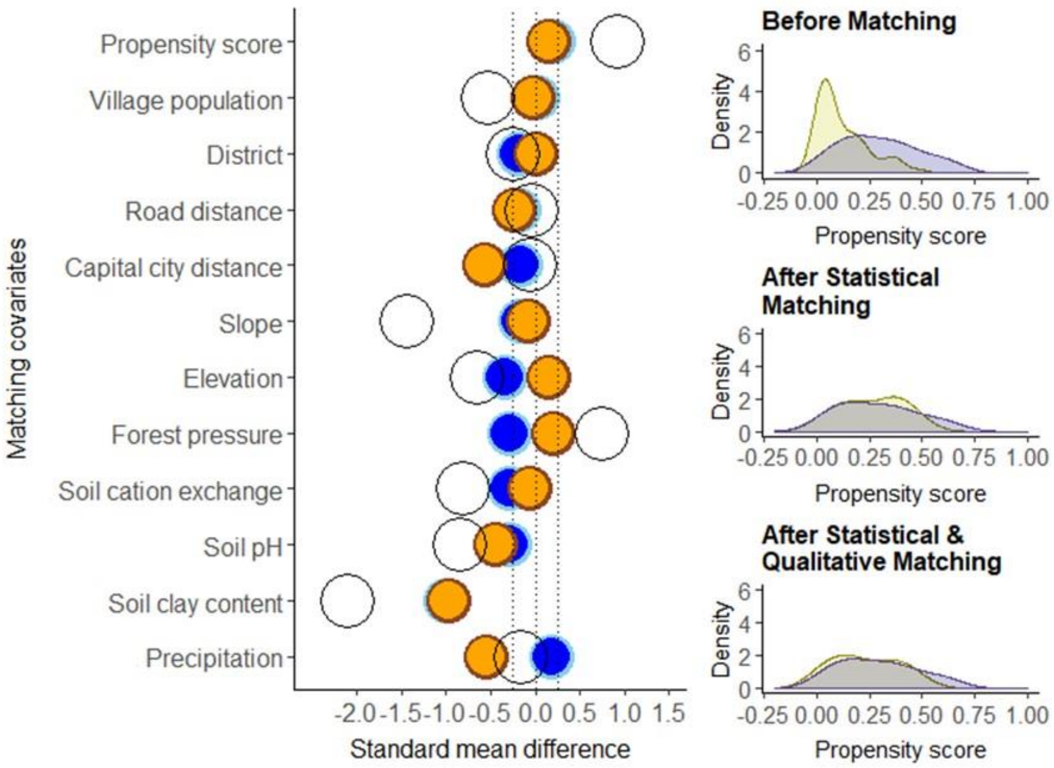


Figure S3.1. Covariate balance of villages selected for undertaking the wellbeing survey before and after matching using villages with Certified CFs certification as treatment and villages without CFs but within 10 km of NFRs as control. (Left) Standardized mean difference for the propensity score and all matching covariates before (open circles), after statistical matching (orange circles) and the finalised match (blue circles) after integrating statistical and qualitative matching by expert panel review. Dotted lines indicate the desired threshold of 0.25 for acceptable difference in standardized means (Stuart 2010). (Right) Propensity score density distribution before matching (top) and after statistical matching (middle) and after integrating statistical and qualitative matching by expert panel review (bottom) with treatment (purple) and control (yellow) groups (overlaps between propensity score distributions are represented in grey). Matching resulted in improved overlap in propensity scores.

S3.4.2 Conservation effectiveness matching

The study area contained 13,000,000 pixels which were randomly sampled to between 150,000 - 300,000 pixels for matching and interaction analyses (Schleicher et al., 2017). This sampling served to reduce spatial autocorrelation and balance competing requirements for statistical power and processing time (Rasolofoson et al., 2015).

CF pixels were matched to unprotected forest pixels to assess the impact of certified CFs in relation to unprotected forest – the counterfactual situation. To ensure that sampling did not bias results, two repeated two additional random samples were taken and matched separately. The direction of relationships of predictor variables remained the same between repeat analyses, confirming that sampling had not influenced the conclusions.

The same NFR vs unprotected comparison was not possible for wellbeing analyses as the social assessments required field visits to individual residents, which was prohibitively expensive to repeat for both PA governance approaches and matched control sites. Since the focus of the paper was on CFs, we prioritised this governance type. To control for potential leakage effects of certified CF conservation displacing degradation to adjacent areas influencing comparisons between treatment and control groups, we excluded unprotected forest that was within a 1 km buffer of PAs from the comparison control group (Andam et al., 2008; Schleicher et al., 2017). Treatment and control units were matched on a suite of confounding variables (Table S6, figures S2-S6).

Table S3.5. Confounding variables used in pixel matching and post matching linear mixed effects models. Superscripts indicate pairs of indicators correlated at $r > 0.7$. ◊ Indicators specific to the pixel matching analysis and not used in the village matching analysis. * indicates square root transformation, + indicates log transformation.

Variable	Justification
Distance to major roads * [1, 2, 6, 9]	Proximity to roads is a key driver of deforestation as this influences accessibility of forests and transportation costs.
Distance to urban centres [4, 9, 10]	Proximity of forests to markets is a major driver of deforestation as this influences transportation costs.
Distance to villages * [7]	Distance of forest pixels to village centres was included as a matching covariate in pixel matching since distance to villages provides a measure of forest accessibility.
Distance to PA boundary *	Distance of pixels from within the PA to the PA boundary provides a measure of forest accessibility and so was included as a matching covariate in matching between protected area types. When matching PAs to unprotected pixels, treatment and control groups were matched on distance to the other PA category, so controlling for spill-over effects.
Slope +, Elevation *	The dominant livelihood in this region is cultivated agriculture, therefore crop productivity is a key driver of development. Slope and elevation are key determinants of agricultural suitability of land for conversion of forest to other uses (Gentle and Narayan, 2012).
Pressure on forest resources + [2, 7, 8]	Forests are subject to anthropogenic pressure both from local settlements and from remote demand centres which influences probability of forest degradation (Platz, 2012).
District [3]	District is a key unit of government administration, with different government staff responsible for delivering forest management in different districts and differing approaches for management arising.
Precipitation [5, 6, 8, 10]	Rainfall will influence the agricultural suitability of land and yield of crops.
Soil composition	Soil composition influences what crops may suitably be grown, soil fertility and agricultural productivity and so influences desirability of land to be cleared for agriculture.
NDVI in 2014 ◊	Since forest with greater biomass may be preferentially harvested, it is important to match forest biomass between treatment and control groups at the start of the time period to be assessed.

X [1], Y [3, 4, 5] coordinates ◊	The unit of pixel matching (30x30m pixels) was much smaller scale than village matching (each village approximately 5-10 km apart) so there is a greater risk of spatial autocorrelation when matching at the scale of pixels. To reduce the risk of spatial autocorrelation we included an extra spatial matching covariate of the pixel X and Y coordinates.

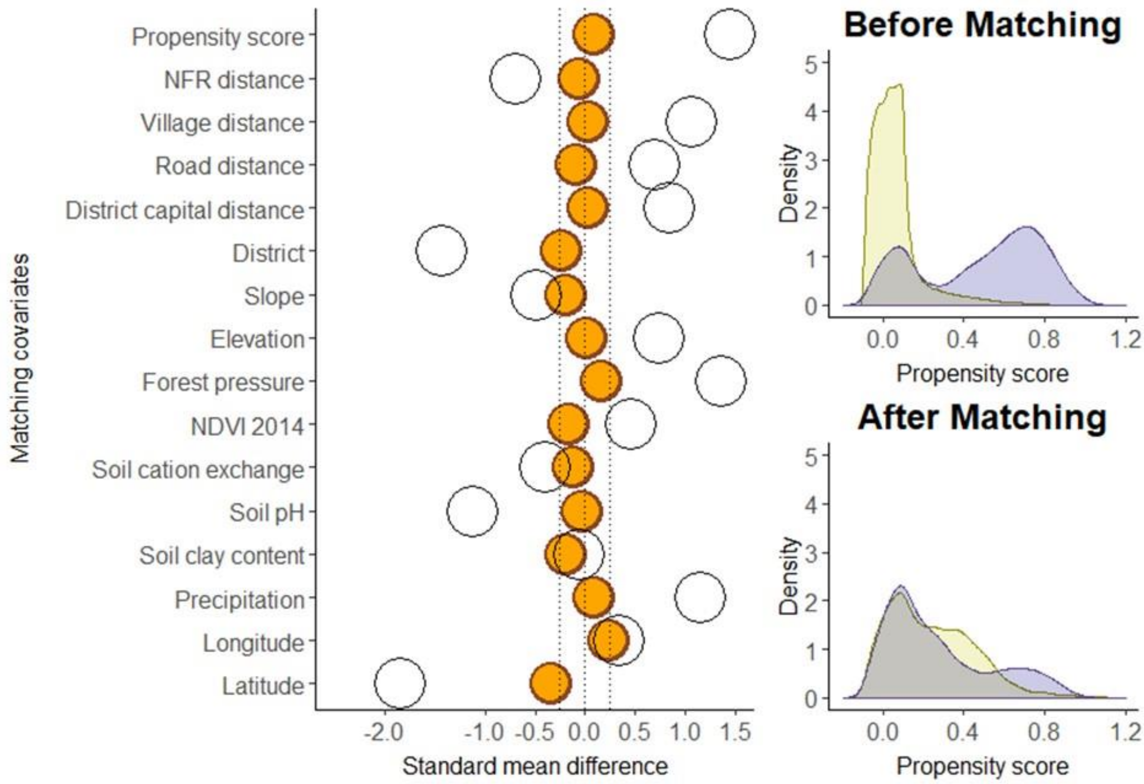


Figure S3.2. Covariate balance for pixels with CFs as treatment and unprotected pixels as control. (Left) Standardized mean difference for the propensity score and all matching covariates before (open circles) and after statistical matching (orange circles). Dotted lines indicate the desired threshold of 0.25 for acceptable difference in standardized means (Stuart 2010). (Right) Propensity score density distribution before matching (top) and after statistical matching (bottom) with treatment (purple) and control (yellow) groups (overlaps between propensity score distributions are represented in grey). Matching resulted in improved overlap in propensity scores.

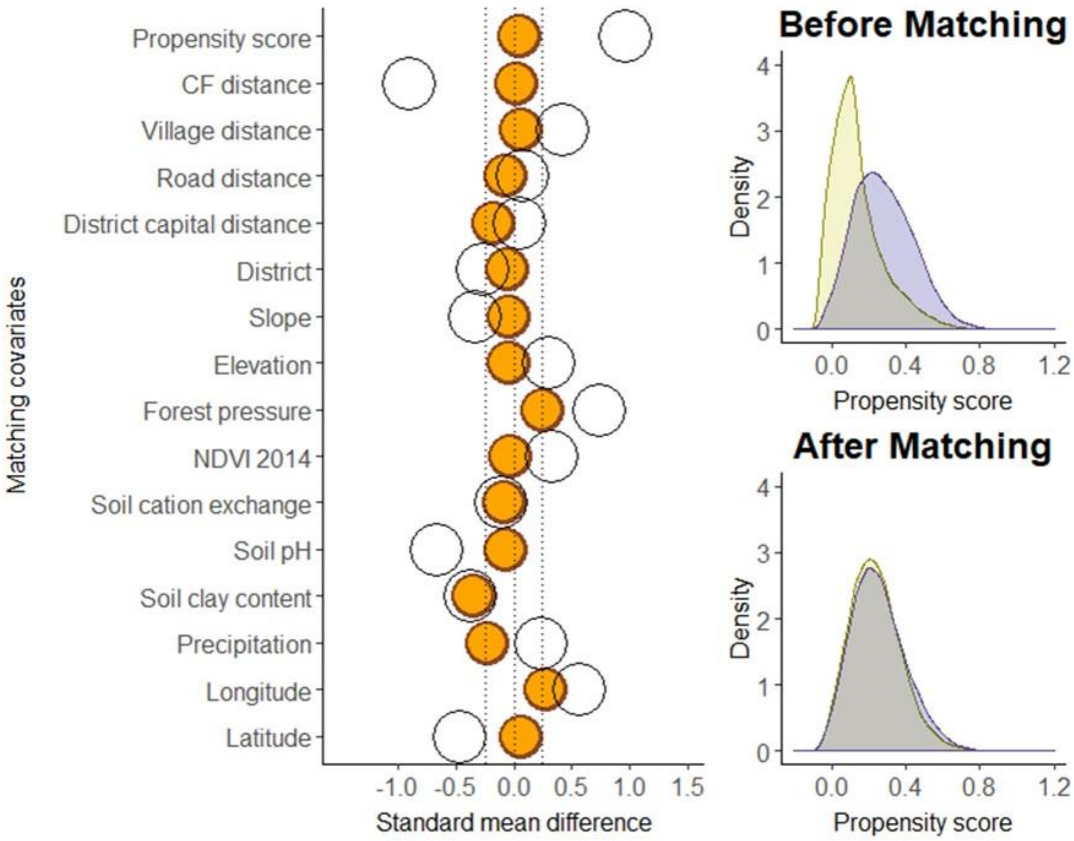


Figure S3.3. Covariate balance for pixels with NFRs as treatment and unprotected pixels as control. (Left) Standardized mean difference for the propensity score and all matching covariates before (open circles) and after statistical matching (orange circles). Dotted lines indicate the desired threshold of 0.25 for acceptable difference in standardized means (Stuart 2010). (Right) Propensity score density distribution before matching (top) and after statistical matching (bottom) with treatment (purple) and control (yellow) groups (overlaps between propensity score distributions are represented in grey). Matching resulted in improved overlap in propensity scores.

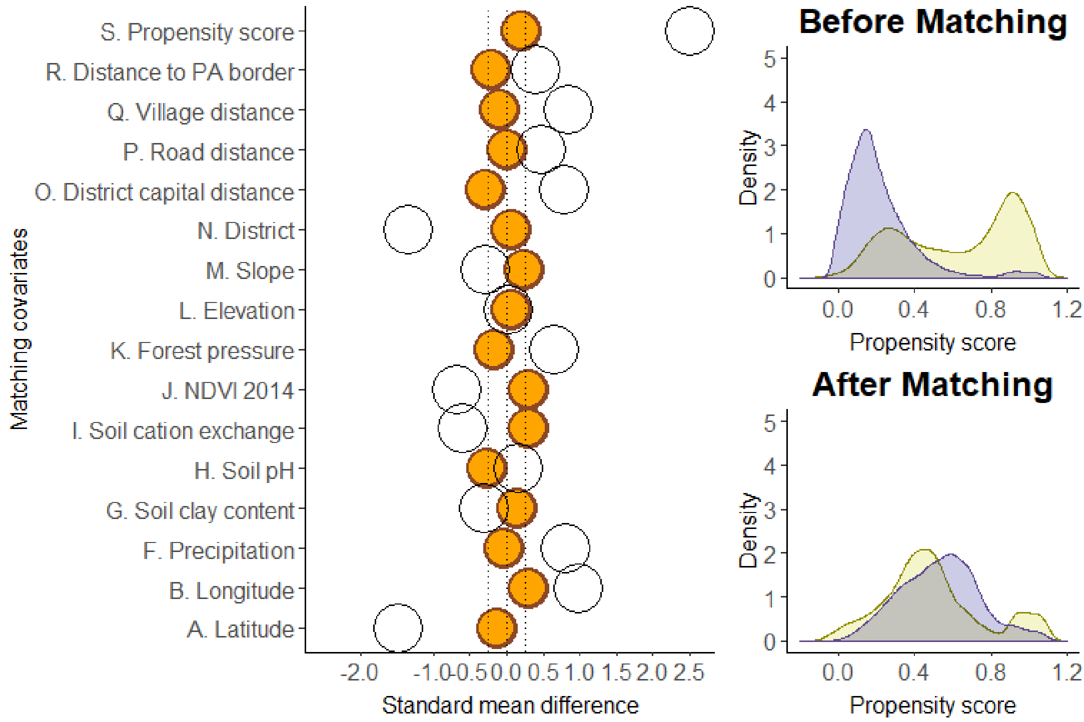


Figure S3.4. Covariate balance for pixels with NFRs within 10 km of Certified CFs as treatment and NFRs beyond 10 km from CFs as control. (Left) Standardized mean difference for the propensity score and all matching covariates before (open circles) and after statistical matching (orange circles). Dotted lines indicate the desired threshold of 0.25 for acceptable difference in standardized means (Stuart 2010). (Right) Propensity score density distribution before matching (top) and after statistical matching (bottom) with NFRs close to CFs (purple) and NFRs alone (yellow) groups (overlaps between propensity score distributions are represented in grey). Matching resulted in improved overlap in propensity scores.

S3.5. Analysis

For each model, we used an information theoretic approach via the 'model.avg' function in the MuMIn package to compare two sub-models (1) a simple model that included only hypothesis testing variables as fixed effects and (2) a more complex model that, in addition to hypothesis testing variables, contained an orthogonal set of confounding socio-environmental variables (Barton, 2019). For each model we report the simple, complex and averaged sub-models (Supporting Information Tables S6-S34). Orthogonality among predictors is a fundamental assumption of statistical analysis and required to avoid erroneous results (Zuur, Ieno, & Elphick, 2010). We specified that variables were highly correlated at r >= 0.7 (Table S4 and S6).

While the inclusion of additional confounding variables has potential to introduce model redundancy, it serves to control for any residual imbalances in the distributions of confounding variables between governance approaches after matching (Ho et al., 2007; Ferraro and Miranda, 2014). Diagnostic plots of model residuals were used to check for heteroscedasticity and outliers (Crawley, 2007). Results report model averages weighted by AICc (Burnham and Anderson 2002). Figures of mixed effects models show estimated marginal effect sizes while accounting for the effect of other variables (Lüdecke, 2018; Fox and Weisberg, 2019).

To check whether pixel sampling influenced model estimates, we undertook a second random draw of pixels and repeated the matching and post-matching analyses on this second sample. The identified significant relationships were replicated in this second sample, verifying that model estimates were not influenced by sampling.

Model estimates of human wellbeing

Table S6. Predictor variables of the Human Wellbeing Index, excluding confounding socio-environmental variables, with conditional r^2 of fixed and random effects = 0.111*, marginal r^2 of fixed effects only = 0.072.

Fixed effects	Estimate	Standard error	T value	P value+
(Intercept)	0.516	0.009	59.455	<0.001

GovernanceCF	0.027	0.013	2.159	0.038
EliteYes	0.041	0.011	3.835	<0.001
GenderMale	0.051	0.008	6.146	<0.001
governance:gender	-0.032	0.012	-2.665	0.008
governance:elite	0.000	0.015	0.023	0.982

Note: *r^2 calculated following Nakagawa et al., (2017) using the performance package in R.

+ P values for mixed effects models calculated using lmerTest package in r (Kuznetsova, Brockhoff and Christensen, 2017).

Table S7. Predictor variables of the Human Wellbeing Index, including all confounding orthogonal socio-environmental variables in post matching analysis, with conditional r^2 of fixed and random effects = 0.104*, marginal r^2 of fixed effects only = 0.100. Confounding variables are all Z transformed.

Fixed effects	Estimate	Standard error	T value	P value
(Intercept)	0.518	0.007	76.150	<0.001
GovernanceCF	0.022	0.011	2.056	0.046
EliteYes	0.040	0.011	3.704	<0.001
genderMale	0.050	0.008	5.987	<0.001
governance:gender	-0.031	0.012	-2.542	0.011
governance:elite	-0.001	0.016	-0.081	0.936
Confounding variables				
Distance to PA	0.003	0.005	0.669	0.507
Distance to roads	-0.013	0.005	-2.493	0.015
Distance to villages	-0.005	0.004	-1.264	0.216

Elevation	0.009	0.004	2.125	0.047
Slope	0.002	0.004	0.391	0.700
Forest Pressure	0.007	0.004	1.893	0.061
Precipitation	0.000	0.004	-0.100	0.922
Soil pH	-0.002	0.004	-0.595	0.553

Table S8. Model averaged estimates of predictor variables of the Human Wellbeing Index, including all confounding orthogonal socio-environmental variables in post matching analysis. Confounding variables are all Z transformed.

Fixed effects	Estimate	Standard error	Z value	P value
(Intercept)	0.516	0.008	60.871	<0.001
GovernanceCF	0.026	0.012	2.111	0.035
EliteYes	0.041	0.011	3.807	<0.001
genderMale	0.051	0.008	6.112	<0.001
governance:gender	-0.032	0.012	2.641	0.008
governance:elite	0.000	0.016	0.007	0.995
Confounding variables				
Distance to PA	0.000	0.002	0.223	0.823
Distance to roads	-0.002	0.005	0.391	0.696
Distance to villages	-0.001	0.003	0.323	0.746
Elevation	0.001	0.004	0.380	0.704
Slope	0.000	0.002	0.144	0.885
Forest Pressure	0.001	0.003	0.370	0.711
Precipitation	0.000	0.002	0.039	0.969

Soil pH	0.000	0.002	0.205	0.838

Individual wellbeing domains

Table S9. Predictor variables of the **Material domain** of wellbeing with overall conditional $r^2 = 0.089$, marginal $r^2 = 0.068$.

Fixed effects	Estimate	Standard error	T value	P value[+]
(Intercept)	0.415	0.011	36.311	<0.001
GovernanceCF	0.019	0.017	1.141	0.260
EliteYes	0.035	0.016	2.130	0.033
GenderMale	0.091	0.013	7.281	<0.001
Governance:Gender	-0.048	0.018	-2.663	0.008
Governance:Elite	-0.017	0.024	-0.732	0.464

Table S10. Predictor variables of the **Material domain** of wellbeing, including all confounding orthogonal socio-environmental variables, with marginal r^2 of fixed effects only = 0.093. Confounding variables are all Z transformed.

Fixed effects	Estimate	Standard error	T value	P value
(Intercept)	0.424	0.010	43.000	<0.001
GovernanceCF	0.003	0.015	0.223	0.824
EliteYes	0.034	0.016	2.050	0.041
GenderMale	0.088	0.013	6.966	<0.001
Governance:Gender	-0.045	0.018	-2.442	0.015
Governance:Elite	-0.018	0.024	-0.758	0.449
Confounding variables				

Distance to PA	0.000	0.007	0.027	0.979
Distance to roads	0.003	0.008	0.335	0.738
Distance to villages	-0.016	0.006	-2.618	0.009
Elevation	0.002	0.006	0.338	0.736
Slope	-0.010	0.006	-1.770	0.077
Forest pressure	0.004	0.006	0.712	0.476
Precipitation	0.014	0.006	2.265	0.024
Soil pH	0.003	0.006	0.494	0.622

Table S11. Model averaged estimates of predictor variables of the material domain of wellbeing, including all confounding orthogonal socio-environmental variables in post matching analysis. Confounding variables are all Z transformed.

Fixed effects	Estimate	Standard error	Z value	P value
(Intercept)	0.422	0.011	39.354	<0.001
GovernanceCF	0.007	0.017	0.390	0.697
EliteYes	0.034	0.016	2.063	0.039
genderMale	0.089	0.013	6.995	<0.001
governance:gender	-0.045	0.018	2.474	0.013
governance:elite	-0.018	0.024	0.752	0.452
Distance to PA	0.000	0.006	0.024	0.981
Distance to roads	0.002	0.007	0.295	0.768
Distance to villages	-0.013	0.008	1.516	0.129
Elevation	0.002	0.005	0.298	0.766
Slope	-0.008	0.006	1.238	0.216
Forest Pressure	0.003	0.005	0.606	0.545

Precipitation	0.011	0.008	1.419	0.156
Soil pH	0.002	0.005	0.430	0.667

Table S12. Predictor variables of the **Health domain** of wellbeing with overall conditional $r^2 = 0.150$, marginal $r^2 = 0.045$.

Fixed effects	Estimate	Standard error	T value	P value[+]
(Intercept)	0.392	0.018	21.517	<0.001
GovernanceCF	0.060	0.026	2.253	0.033
EliteYes	0.011	0.016	0.651	0.515
GenderMale	0.050	0.013	3.966	<0.001
Governance:Gender	-0.062	0.018	-3.413	0.001
Governance:Elite	0.055	0.024	2.339	0.020

Table S13. Predictor variables of the **Health domain** of wellbeing, including all confounding orthogonal socio-environmental variables, with marginal r^2 of fixed effects only = 0.092. Confounding variables are all Z transformed.

Fixed effects	Estimate	Standard error	T value	P value
(Intercept)	0.394	0.015	26.046	<0.001
GovernanceCF	0.060	0.023	2.608	0.015
EliteYes	0.011	0.017	0.653	0.514
GenderMale	0.042	0.013	3.219	0.001
Governance:Gender	-0.053	0.018	-2.894	0.004
Governance:Elite	0.052	0.024	2.178	0.030

Confounding variables				
Distance to PA	-0.002	0.009	-0.229	0.819
Distance to roads	-0.007	0.010	-0.720	0.472
Distance to villages	0.004	0.009	0.488	0.627
Elevation	0.005	0.010	0.487	0.630
Slope	0.001	0.010	0.140	0.889
Forest pressure	0.021	0.007	3.205	0.001
Precipitation	-0.030	0.010	-2.887	0.009
Soil pH	-0.009	0.007	-1.345	0.179

Table S14. Model averaged estimates of predictor variables of the health domain of wellbeing. Confounding variables are all Z transformed.

Fixed effects	Estimate	Standard error	Z value	P value
(Intercept)	0.394	0.017	23.609	<0.001
GovernanceCF	0.060	0.025	2.426	0.015
EliteYes	0.011	0.017	0.651	0.515
genderMale	0.045	0.013	3.369	0.001
governance:gender	-0.058	0.019	3.036	0.002
governance:elite	0.053	0.024	2.241	0.025
Distance to PA	-0.001	0.007	0.167	0.867
Distance to roads	-0.004	0.008	0.479	0.632
Distance to villages	0.002	0.007	0.342	0.732
Elevation	0.003	0.007	0.341	0.733
Slope	0.001	0.007	0.103	0.918

Forest Pressure	0.012	0.012	0.996	0.319
Precipitation	-0.017	0.017	0.976	0.329
Soil pH	-0.005	0.007	0.737	0.461

Table S15. Predictor variables of the **Social relations** domain of wellbeing with conditional r^2 = 0.172, marginal r^2 = 0.082.

Fixed effects	Estimate	Standard error	T value	P value$^+$
(Intercept)	0.523	0.024	22.081	<0.001
GovernanceCF	0.014	0.034	0.414	0.682
EliteYes	0.086	0.022	3.850	<0.001
GenderMale	0.119	0.017	6.920	<0.001
Governance:Gender	-0.027	0.025	-1.078	0.281
Governance:Elite	-0.044	0.032	-1.366	0.172

Table S16. Predictor variables of the **Social relations** domain of wellbeing, including all confounding orthogonal socio-environmental variables, with marginal r^2 of fixed effects only = 0.113. Confounding variables are all Z transformed.

Fixed effects	Estimate	Standard error	T value	P value
(Intercept)	0.532	0.022	24.649	<0.001
GovernanceCF	-0.008	0.033	-0.255	0.801
EliteYes	0.085	0.023	3.774	<0.001
GenderMale	0.127	0.018	7.237	<0.001
Governance:Gender	-0.035	0.025	-1.390	0.165
Governance:Elite	-0.043	0.032	-1.329	0.184

Confounding variables				
Distance to PA	0.023	0.012	1.845	0.066
Distance to roads	-0.024	0.014	-1.735	0.084
Distance to villages	-0.006	0.012	-0.516	0.607
Elevation	0.001	0.014	0.047	0.963
Slope	-0.001	0.014	-0.105	0.917
Forest pressure	-0.004	0.009	-0.383	0.702
Precipitation	0.034	0.015	2.248	0.033
Soil pH	-0.011	0.010	-1.127	0.260

Table S17. Model averaged estimates of predictor variables of the social relations domain of wellbeing. Confounding variables are all Z transformed.

Fixed effects	Estimate	Standard error	Z value	P value
(Intercept)	0.524	0.024	22.168	<0.001
GovernanceCF	0.011	0.035	0.327	0.744
EliteYes	0.086	0.022	3.835	<0.001
genderMale	0.120	0.017	6.864	<0.001
governance:gender	-0.028	0.025	1.109	0.267
governance:elite	-0.044	0.032	1.359	0.174
Confounding variables				
Distance to PA	0.003	0.009	0.324	0.746
Distance to roads	-0.003	0.009	0.319	0.750
Distance to villages	-0.001	0.005	0.163	0.871
Elevation	0.000	0.005	0.016	0.987

Slope	0.000	0.005	0.037	0.971
Forest Pressure	0.000	0.003	0.126	0.900
Precipitation	0.004	0.012	0.338	0.735
Soil pH	-0.001	0.005	0.271	0.786

Table S18. Predictor variables of the **Security domain** of wellbeing with conditional r^2 = 0.196, marginal r^2 = 0.013.

Fixed effects	Estimate	Standard error	T value	P value[+]
(Intercept)	0.523	0.024	22.081	<0.001
GovernanceCF	0.014	0.034	0.414	0.682
EliteYes	0.086	0.022	3.850	<0.001
GenderMale	0.119	0.017	6.920	<0.001
Governance:Gender	-0.027	0.025	-1.078	0.281
Governance:Elite	-0.044	0.032	-1.366	0.172

Table S19. Predictor variables of the **Security domain** of wellbeing, including all confounding orthogonal socio-environmental variables, with marginal r^2 of fixed effects = 0.109. Confounding variables are all Z transformed.

Fixed effects	Estimate	Standard error	T value	P value
(Intercept)	0.546	0.013	41.674	<0.001
GovernanceCF	0.047	0.020	2.342	0.027
EliteYes	0.023	0.014	1.667	0.096
GenderMale	-0.006	0.011	-0.564	0.573
Governance:Gender	0.013	0.016	0.836	0.404
Governance:Elite	0.010	0.020	0.524	0.600

Confounding variables				
Distance to PA	-0.026	0.008	-3.432	0.001
Distance to roads	-0.011	0.008	-1.312	0.191
Distance to villages	0.006	0.007	0.810	0.420
Elevation	0.003	0.008	0.411	0.684
Slope	0.020	0.008	2.413	0.022
Forest pressure	0.008	0.006	1.392	0.165
Precipitation	0.003	0.009	0.299	0.768
Soil pH	0.008	0.006	1.327	0.185

Table S20. Model averaged estimates of predictor variables of the security domain of wellbeing. Confounding variables are all Z transformed.

Fixed effects	Estimate	Standard error	Z value	P value
(Intercept)	0.685	0.009	76.163	<0.001
GovernanceCF	0.030	0.013	2.262	0.024
EliteYes	0.056	0.012	4.469	<0.001
genderMale	-0.004	0.010	0.465	0.642
governance:gender	-0.030	0.014	2.175	0.030
governance:elite	-0.009	0.018	0.473	0.636
Distance to PA	0.000	0.001	0.090	0.929
Distance to roads	0.000	0.001	0.009	0.993
Distance to villages	0.000	0.001	0.090	0.928
Elevation	0.000	0.001	0.020	0.984

Slope	0.000	0.001	0.099	0.921
Forest Pressure	0.000	0.001	0.073	0.942
Precipitation	0.000	0.001	0.096	0.924
Soil pH	0.000	0.001	0.022	0.982

Table S21. Predictor variables of the **Freedom domain** of wellbeing with conditional $r^2 = 0.072$, marginal $r^2 = 0.048$.

Fixed effects	Estimate	Standard error	T value	P value[+]
(Intercept)	0.685	0.009	76.199	<0.001
GovernanceCF	0.030	0.013	2.265	0.029
EliteYes	0.056	0.012	4.475	<0.001
GenderMale	-0.004	0.010	-0.463	0.643
Governance:Gender	-0.030	0.014	-2.181	0.029
Governance:Elite	-0.008	0.018	-0.471	0.638

Table S22. Predictor variables of the **Freedom domain** of wellbeing, including all confounding orthogonal socio-environmental variables, with marginal r^2 of fixed effects only = 0.062. Confounding variables are all Z transformed.

Fixed effects	Estimate	Standard error	T value	P value
(Intercept)	0.685	0.009	80.338	< 0.001
GovernanceCF	0.030	0.013	2.298	0.025
EliteYes	0.056	0.013	4.433	< 0.001
GenderMale	-0.006	0.010	-0.584	0.559
Governance:Gender	-0.029	0.014	-2.032	0.042

Governance:Elite	-0.011	0.018	-0.626	0.532
Confounding variables				
Distance to PA	-0.006	0.006	-0.987	0.326
Distance to roads	0.000	0.006	-0.074	0.941
Distance to villages	-0.005	0.005	-0.993	0.325
Elevation	-0.001	0.005	-0.160	0.874
Slope	0.006	0.005	1.230	0.228
Forest pressure	0.003	0.005	0.705	0.482
Precipitation	-0.006	0.006	-1.135	0.267
Soil pH	0.001	0.005	0.180	0.858

Table S23. Model averaged estimates of predictor variables of the freedom domain of wellbeing. Confounding variables are all Z transformed.

Fixed effects	Estimate	Standard error	Z value	P value
(Intercept)	0.685	0.009	76.163	<0.001
GovernanceCF	0.030	0.013	2.262	0.024
EliteYes	0.056	0.012	4.469	<0.001
genderMale	-0.004	0.010	0.465	0.642
governance:gender	-0.030	0.014	2.175	0.030
governance:elite	-0.009	0.018	0.473	0.636
Confounding variables				
Distance to PA	0.000	0.001	0.090	0.929
Distance to roads	0.000	0.001	0.009	0.993
Distance to villages	0.000	0.001	0.090	0.928

Elevation	0.000	0.001	0.020	0.984
Slope	0.000	0.001	0.099	0.921
Forest Pressure	0.000	0.001	0.073	0.942
Precipitation	0.000	0.001	0.096	0.924
Soil pH	0.000	0.001	0.022	0.982

Forest conservation models

The following models aim to test the difference between matched pairs of governance approaches in NDVI change. Note that fixed effects do not include NDVI in 2014, since this would introduce mathematical coupling with the response variable and could lead to spurious model estimates (Tu and Gilthorpe, 2007). Instead NDVI in 2014 is accounted for in the calculation of the response variable as described in supporting information section S1.2.

In terms of spatial heterogeneity, increasing distance from villages was associated with more positive NDVI change (estimate = 0.036, SE = 0.002, P = <0.001, Table S26). This trend may be explained by village expansion in line with population growth and agricultural expansion from village centres. A further point of interest is that NDVI change was more negative further from roads (estimate = - 0.013, SE = 0.005, P = 0.008, Table S26). At face value this may be a counterintuitive finding. However if we consider the recent historical context of the region, the Rufiji bridge and surfaced road running through the study area was constructed in 2003 and a large amount of deforestation occurred directly adjacent to the roads at that time. So considering the work of Ahrends et al. (Ahrends et al., 2010) on waves of deforestation in Tanzania, we suggest that more recent trends in deforestation that occurred during the study period (2014-2019) would be further from this road, since forest directly adjacent to the road had already been degraded.

Certified CFs vs unprotected

Table S24. Model output of predictor variables of NDVI change between matched CF and unprotected pixels. Conditional r^2 = 0.300, Marginal r^2 = 0.029. Number of matched pixels from each governance approach = 8,154.

Fixed effects	Estimate	Standard error	T value	P value
(Intercept)	0.012	0.020	0.58	0.569
governanceCF	0.039	0.029	1.379	0.185

Table S25. Model output of predictor variables of NDVI change between matched CF and unprotected pixels. Conditional r^2 = 0.317, Marginal r^2 = 0.133. Number of matched pixels from each governance approach = 8,154. Confounding variables are all Z transformed.

Fixed effects	Estimate	Standard error	T value	P value
(Intercept)	0.000	0.017	0.011	0.991
governanceCF	0.047	0.024	1.984	0.063
Distance to NFR	-0.008	0.002	-3.867	<0.001
Distance to roads	-0.013	0.005	-2.636	0.008
Distance to villages	0.036	0.002	17.183	<0.001
Elevation	-0.011	0.002	-5.015	<0.001
Slope	0.005	0.003	2.048	0.041
Precipitation	-0.038	0.006	-6.313	<0.001
Soil pH	-0.643	0.106	-6.059	<0.001

Table S26. Averaged model estimates from multi-model inference of NDVI change between matched CF and unprotected pixels. Confounding variables are all Z transformed.

Fixed effects	Estimate	Standard error	T value	P value
(Intercept)	0.000	0.017	0.011	0.991
governanceCF	0.047	0.024	1.984	0.047

Distance to NFR	-0.008	0.002	-3.867	<0.001
Distance to roads	-0.013	0.005	-2.636	0.008
Distance to villages	0.036	0.002	17.182	<0.001
Elevation	-0.011	0.002	-5.015	<0.001
Slope	0.005	0.003	2.048	0.041
Precipitation	-0.038	0.006	-6.313	<0.001
Soil pH	-0.643	0.106	-6.058	<0.001

PA interaction effects on NDVI change

Table S27. Model output of predictor variables of NDVI change between matched NFR pixels that are 1 = within 10km of Certified CFs vs 0 = more than 10 km from CFs. Conditional r^2 = 0.258, marginal r^2 = 0.048. Number of matched pixels from each governance approach = 8,050.

Fixed effects	Estimate	Standard error	T value	P value
(Intercept)	0.008	0.012	0.696	0.495
Governance1	0.046	0.002	21.727	<0.001

Table S28. Model output of predictor variables of NDVI change between matched NFR pixels that are 1 = within 10km of CFs vs 0 = more than 10 km from CFs. Conditional r^2 = 0.568, marginal r^2 = 0.197. Number of matched pixels from each governance approach = 8,050. Confounding variables are all Z transformed.

Fixed effects	Estimate	Standard error	T value	P value
(Intercept)	-0.001	0.019	-0.051	0.960
Governance1	0.011	0.002	4.835	<0.001
Distance to NFR boundary	-0.003	0.001	-4.314	<0.001
Distance to roads	-0.127	0.004	-35.209	<0.001

Distance to villages	0.020	0.001	13.622	<0.001
Elevation	0.009	0.002	4.441	<0.001
Slope	0.005	0.003	1.726	0.084
Precipitation	0.067	0.005	14.486	<0.001
Soil pH	-0.013	0.001	-11.491	<0.001

Table S29. Averaged model estimates from multi-model inference of NDVI change between matched NFR pixels that are 1 = within 10km of CFs vs 0 = more than 10 km from CFs. Confounding variables are all Z transformed.

Fixed effects	Estimate	Standard error	T value	P value
(Intercept)	-0.001	0.019	0.051	0.959
Governance1	0.011	0.002	4.835	<0.001
Distance to NFR boundary	-0.003	0.001	4.314	<0.001
Distance to roads	-0.127	0.004	35.207	<0.001
Distance to villages	0.020	0.001	13.621	<0.001
Elevation	0.009	0.002	4.440	<0.001
Slope	0.005	0.003	1.726	0.084
Precipitation	0.067	0.005	14.485	<0.001
Soil pH	-0.013	0.001	11.490	<0.001

The effect of proximity between PAs on change in NDVI

CF only- Table S30. Mixed effects model of log NDVI change in Community Forest Reserve pixels as a function of socio environmental predictors based on a random sample of 150,000 pixels from a total pool of 1.4 million pixels from 10 Community Forest Reserves with FSC certification within 10km of matched villages. Conditional $r^2 = 0.462^*$, marginal $r^2 = 0.126$.

Fixed effects	Estimate	Standard Error	T value	P value
(Intercept)	0.035	0.025	1.402	0.190
Distance to NFR	-0.015	<0.001	-36.820	<0.001
Distance to PA boundary	0.021	<0.001	74.740	<0.001
Distance to roads	0.025	0.002	11.278	<0.001
Distance to villages	0.048	0.001	59.855	<0.001
Elevation	0.003	<0.001	5.938	<0.001
Slope	-0.021	0.001	-40.525	<0.001
Precipitation	-0.094	0.002	-47.491	<0.001
Soil pH	-0.004	<0.001	-14.648	<0.001

NFR only Table S31. Mixed effects model of NDVI change in National Forest Reserve pixels as a function of socio environmental predictors based on a random sample of 200,000 pixels from a total pool of 2 million pixels from 15 National Forest Reserves within 10 km of matched villages. Conditional r^2 = 0.443, marginal r^2 = 0.258.

Fixed effects	Estimate	Standard Error	T value	P value
(Intercept)	0.013	0.013	0.988	0.338
Distance to CF	-0.013	<0.001	-35.307	<0.001
Distance to PA boundary	0.003	<0.001	15.175	<0.001
Distance to roads	-0.109	0.001	-132.205	<0.001
Distance to villages	0.021	<0.001	66.275	<0.001
Elevation	-0.008	<0.001	-16.542	<0.001
Slope	-0.001	0.001	-0.673	0.501
Precipitation	0.052	0.001	45.878	<0.001
Soil pH	0.003	<0.001	11.588	<0.001

Impact of certified CFs on community relations with NFR managers

Table S32. Ordinal logistic regression of the impact of having a CF with FSC on community relations with NFR managers. Partial D^2 = 0.005. Partial D^2 was calculated as a measure of explanatory power (equivalent to partial r^2 for linear regressions) as D^2 = (null deviance – model deviance) / null deviance (Guisan and Zimmermann, 2000).

Fixed effects	Estimate	Standard error	T value	P value
GovernanceCF	0.468	0.015	3.215	0.001

Table S33. Ordinal logistic regression of the impact of having a CF with FSC on community relations with NFR managers including confounding socio-environmental variables. Partial D^2 = 0.081.

Fixed effects	Estimate	Standard Error	T value	P value
GovernanceCF	1.212	0.204	5.935	<0.001
Distance to NFR	-0.239	0.455	-0.526	0.598
Distance to roads	1.035	0.430	2.406	0.016
Distance to villages	2.680	0.515	5.199	<0.001
Elevation	-1.258	0.703	-1.788	0.074
Slope	7.582	0.862	8.788	<0.001
Forest Pressure	-3.063	0.703	-4.355	<0.001

Table S34. Averaged model estimates from multi-model inference of ordinal logistic regression of the impact of having a CF with FSC on community relations with NFR managers between simple and confounding socio-environmental variables.

Fixed effects	Estimate	Standard Error	T value	P value
GovernanceCF	1.212	0.204	5.935	<0.001
Distance to NFR	-0.239	0.455	-0.526	0.598
Distance to roads	1.035	0.430	2.406	0.016

Distance to villages	2.680	0.515	5.199	<0.001
Elevation	-1.258	0.703	-1.788	0.074
Slope	7.582	0.862	8.788	<0.001
Forest Pressure	-3.063	0.703	-4.355	<0.001

Chapter four supporting information

S4.1. Exploratory focus groups and interviews

S.4.1.1 Sampling

Table S4.1. List of exploratory interviews and focus groups completed 2018-2019 to identify actor perspectives and locally appropriate indicators.

Location	**Governance approach**	**Focus groups**	**Key informant interviews**	**Year**
Kikole	Certified CF	1 x VNRC (Village Natural Resource Committee)	1 x VNRC	2018
Mchakama	Certified CF	2 x VNRC 1 x Women 1 x Men	8 x subvillage representatives	2019
Makangaga	Matched control village without CF	1 x village committee 1 x women 1 x men 1 x subvillage	6 x subvillage represenatatives 1 x Pindiro National Forest Reserve TFS (Tanzania Forest Service)	2019,
Nanjarinji A	Certified CF	1 x VNRC 1 x Women 1 x Men	4 x subvillage represenatives	2019

Ngea	Certified CF	1 x VNRC	1 x wood trader	2019
Kilwa Masoko	Regional policy context of CF and FSC certification	2 x MCDI	2 x MCDI 1 x District TFS 1 x District Forest Office 1 x Sound & Fair wood trader 1 x WWF donor	2018
Dar es Salaam	National PA policy context	1 x FSC national policy planning meeting	1 x National TFS 1 x TFCG NGO 1 x WWF NGO 1 x Mjumita NGO	2019
Hoteli Tatu	Matched control village without CF	1 x village committee 2 x women 2 x men	1 x livestock keeper	2019,
Nainokwe	Certified CF	1 x livestock keepers 1 x youth female 1 x youth male	1 x livestock keeper	2019
Likawage	Certified CF	2 x livestock keepers 1 x NNRC 2 x women 2 x men	1 x carpenter	2019

S4.1.2 Causal pathways and mechanisms

Specific causal pathways and mechanisms linking certified CFs, human wellbeing and conservation effectiveness impacts were identified by a combination of open-ended questions in focus groups concerning the links between PA governance and wellbeing and conservation effectiveness (Booker and Franks, 2018), and subsequent thematic of consultation transcripts to identify actor perceptions of causal drivers, supplemented by literature review of the logical augments underpinning win-win dialogues in PA governance. Participants gave free, prior, informed consent before consultations began and participant personal information were anonymised to protect their identities. Either Swahili or English was used depending on the preference of participants, with sessions recorded for later translation and analysis.

The exact phrasing of questions varied depending on the group of actors being consulted. However an example of the open-ended questioning style used for a community focus group was:

Open question: 'How does the wellbeing of people in this village relate to the forest?'

Then specifically, 'How does forest reserve [insert name] impact people in this village?' (Booker and Franks, 2018)

And further background contextual information from answers to the following interview questions:

What changes have you seen in this village during your time here?

And the follow up question 'why is the situation like this?'.

Context specific mechanism indicators (Table S3.4) identified during the exploratory qualitative phase of fieldwork were included in the quantitative questionnaire. To limit the complexity and aid interpretation of the conceptual model, related mechanism indicators were grouped into key concepts of ecosystem benefits, costs and shared communal benefits by CFA to create latent variables for these concepts.

Table S4.1.2. Pathways linking certified CFs, forest restoration and human wellbeing identified by literature review and exploratory focus groups and key informant interviews. Table includes both academic references and transcript exerts of the perspectives of different actors from the exploratory phase of fieldwork.

Pathways / hypotheses	**Logical statement**	**Source**
Equitable governance Relations with managers Conservation effectiveness	Equitable governance improves fair distribution of benefits, improves relations with forest managers, leading to improved conservation outcomes	(Martin et al., 2019; Romero et al., 2017)

Financial benefits Improve wellbeing Improve attitude towards conservation Improve conservation	Forest certification improves direct/personal economic benefits from forest management. Financial incentives will improve wellbeing, leading local communities to support conservation and thereby improve conservation effectiveness	(Dietz, Ostrom and Stern, 2003; Persha, Agrawal and Chhatre, 2011b)
Trade-off pathway **Land availability** Cop yield Wellbeing **Land availability** PA size Timber revenue	'The government extended the national forest reserve boundary and so we have been left with a small area for farming. That land, it was very fertile, it was supporting us to have high production and we had a lot of food surplus. But now we have little food because we harvest very little' 'If you are a new person in this sub-village it is difficult to get land for farming even for settlement, because the TFS want us to move from here and they have left us with small area for farming' Facilitator: How to you assess which villages to support to establish a CF?	Focus group with subvillage representatives from a village without a CF MCDI focus group on the challenges of CF governance

	MCDI director of operations: There are many factors, for example we need to wake up about the political stability. For example there are some sites where we started but could not complete the process. Because of the political issues. MCDI project administrator: ‘Having the community forest limits the space for cultivation. When we setup the community forest it limits the common pool resources available and this creates some conflict.‘ MCDI director of operations: ‘Yes, for example within Mchakama, some say “we need forest” others say “we need space cultivation”. Then we decided to make a court to decide; Those who decided wanted conservation and those against – there was a small difference in favour of the CF, but it was not much’. Field officer: Also, if we go to a village and they have already done a land-use plan then we	

	can see how big the land available is, and if the trees have good potential, then we can see if there is good potential in terms of the number and species of trees and land available.	
Win-win feedback pathways **Conservation effectiveness** Forest availability Timber revenue **Economic benefits** support for conservation	Facilitator: Do villages have to produce a land-use plan? MCDI director of operations: 'Yes, but sometimes we will support them if they have enough land and forest. One issue is that in the past perhaps the village was a centre for illegal logging, so then you go to waste money setting up a VLFR where there are few trees. We would like people to do conservation and at the same time take benefit. it is very difficult to talk about conservation without benefit.'	MCDI focus group, discussing the process of CF establishment (continuing the conversation from above)
Revenue from timber sales used to fund patrols Conservation effectiveness	We are funding forest management activities, we pay even from our own village	Village 2, Village Natural Resource Committee (VNRC)

	basket for meetings and patrols	
Equitable Governance + Wellbeing Conservation	Positive perceptions of governance and positive wellbeing impacts improve conservation effectiveness	(Oldekop et al., 2016; Koning, 2011)
Equitable governance Wellbeing + Conservation	CFs positively impact human wellbeing and conservation by improving equitable forest governance	(Dawson, 2017; Oldekop et al., 2016)
Conservation **Ecosystem services** Crop yield Wellbeing	Conservation increases availability of forest resources and ecosystem services which improve human wellbeing. 'when the forest is healthy, we expect to get high rainfall through the year, maintaining soil fertility and water availability' 'The community they think the forest is a bank for them; for soil fertility, for wildlife, for water' 'previously we were harvesting high crop yield such as 20 sacks of maize or any crop and once we were selling them we were getting enough money but now days we are harvesting few crop yield and hence we are getting less	(Millenium Ecosystem Assessment, 2005) Village 11, all – male exploratory focus group National TFS interview, Dar es Salaam Village 11 – all women exploratory focus group

	money so in general life is very tough'	
Ecosystem costs Wellbeing	'wild animal like warthogs, baboon, monkeys, have increased now days since we are protecting that forest and they are making a big disturbance to us, they are raiding on our crops, and hence causing small harvest' 'Crop raiding by warthog and monkey is getting much worse because no one is hunting anymore'	Village Natural Resource Committee (VNRC), village 11 key informant interview with woman respondent in certified CF village in the poor socio-economic category
Community development/ infrastructure Ecosystem services Wellbeing	'the main reason which made us to agree on that idea (to establish a CF) it's because that forest was being used without any general benefit to the villagers, because people were accessing and harvesting forest material without following rules. One of our expectation was to make sure the forest benefit all the village members. We are spending the money from timber sales on some village development projects, such as improvements for the school building, water pumps for drinking and irrigation, solar light for the health centre. So	VNRC, village 5, Exploratory focus group

	people from this village do not need to pay their own money to help these projects and can keep for buying food for their household’	

S.4.2 Quantitative questionnaire sampling

Table S4.2.1: Quantitative questionnaire village locations with a minimum of 50 questionnaires completed per village.

Village	Governance (Certified CF =1 / control = 0)	District	Date completed
Mbwara	0	Rufiji	07/07/2019
Kunguruwe	0	Rufiji	12/08/2019
Chumbi B	0	Rufiji	09/09/2019
Nyamwage A	1	Rufiji	21/07/2019
Tawi	1	Rufiji	31/07/2019
Megeregere	0	Kilwa	04/08/2019
Kikole	1	Kilwa	16/08/2019
Nainokwe	1	Kilwa	25/07/2019
Ngea	1	Kilwa	27/06/2019
Liwiti	1	Kilwa	30/07/2019
Likawage	1	Kilwa	21/07/2019
Nanjarinji A	1	Kilwa	15/06/2019
Zinga Kibaoni	0	Kilwa	24/06/2019
Makangaga	0	Kilwa	10/06/2019

Hoteli Tatu	0	Kilwa	29/08/2019
Naking'ombe	0	Kilwa	01/07/2019
Kisima Mkika	0	Kilwa	01/07/2019
Mchakama	1	Kilwa	27/08/2019
Kinjumbi	0	Kilwa	11/08/2019

Sampling was undertaken within villages to identify a minimum of 50 respondents per village; 10 village elites and 40 non-elites per village to align with the ratio of elites and non-elites present in the whole village population as identified through pilot surveys. Categorisation of elite status was undertaken by sub-village chiefs using village registers and then randomly sampled using a random number generator. In Tanzania villages are sub-divided into 3-5 sub-villages distributed around a central sub-village. To ensure good spatial representation of the sample data, at least one day of survey effort was undertaken in each sub-village.

Table S4.2.2. Summary of questionnaire sampling between gender and village elite categories. Total sample size of 955.

	Non-elite	Elite
Women	391	83
Men	389	92

For structural equation models, only complete questionnaires could be used, discounting any questionnaires with incomplete responses. To verify that discounting of incomplete questionnaires did not make a material difference to the sampling strategy we subset the data by the main village actor sample units.

Table S4.2.3. Summary of complete questionnaire sampling between gender and village elite categories. Total sample size of 671

	Non-elite	Elite
Women	213	71
Men	307	80

S4.3. Response variables

Human wellbeing and forest conservation response variables are described in the supporting information of chapter three.

S4.3.1. Governance equity indicators

Locally appropriate governance equity indicators were identified first by developing an understanding of how CFS units are governed through interviews and focus groups with government representatives and community forest managers from Village Natural Resource Committees (VNRCs). This focused on understanding certified CFs governance arrangements, such as forest access rules and restrictions, bylaws and procedures for participation, forest revenue expenditures and punishments and fines. Any issues relating to the implementation of these governance arrangements were then identified by asking community members in interviews and focus groups about relationships between forest managers and non-managers and follow-up questions asking 'why is the situation like this?' (Table S3.3). Although Franks and Brooker (36) recommend exploring 5 or 6 governance equity issues/indicators, their intended use is largely qualitative. Due to the limitations of multidimensional analysis in CFA, we used only three orthogonal indicators to generate a single latent variable representative of the governance equity concept.

Table S4.3.3. Indicators of governance equity

Indicator	Dimension	Description
Participation	Procedure	Ordinal categories ranging from 0-4 for level of participation in key forest governance decisions: 12.1. In the last 5 years were you involved in meetings to: (a) elect the people responsible for managing your village's forest resources? (Yes/No)

		(b) decide what activities you can do in the forest belonging to this village? (Yes/No) (c) decide where the boundaries of the conserved forest in your village are located? (Yes/No) (d) decide how the revenue from timber harvesting should be spent? (Yes/No)
Satisfaction	Recognition	Likert scale 1-5 12.3 How satisfied are you with the rules about what activities people can be done in [insert name of VLFR/open area] Scale 1-5: 1= dissatisfied, 5 = satisfied
Trust	Distribution	Likert scale 1-5 12.4. Do you trust the people responsible for managing [insert name of CF/ forest of village open area] to implement the rules in a fair way? (Example: Specifically, consider how any conflicts have being resolved, fines given and bylaws implemented, corruption of managers)) Scale 1-5: 1= unfair, 5 = fair

Table S4.3.4. Mechanism indicators included in the quantitative questionnaire

Indicator	Group / latent variable	Description
Land availability	Land availability for farming	This question is about the amount of fertile land available to you. Is there enough fertile land for farming available to you on village land? Scale 1-5:

		1= serious shortage of fertile land, 2= Little shortage of fertile land 3= just the right amount of fertile land available 4 = a little spare fertile land available 5 = a lot extra fertile land available
Land security	Land availability for farming	This question is about the Security of your household's personal land tenure. How worried are you that your farm land may be taken away by other people in this village or other villages, or government or private companies? Scale 1-5 1 = Very worried 2= little worried 3 = OK 4 = quite confident 5 = very confident
Soil fertility	Ecosystem benefits	In the last 5 years, has there been a change in the fertility of soil in your farm(s)? Scale 1-5: 1= a lot less fertile, 2= a little less fertile 3 = the same, 4 = a little more fertile 5 = A lot more fertile
Firewood	Ecosystem benefits	Has your ability to collect firewood changed? 1= harder, 3 = no change, 5 = easier
Water	Ecosystem benefits	In the last 5 years, has there been any change in how reliably you can access water for farming and domestic use (drinking/washing/cooking) Scale 1-5: 1= decrease a lot, 5 = increase a lot
Crop damage from wild animals	Ecosystem cost	In the last five years how much damage has there been to your crops because of wild animals?

		5. More than ¾ 4. Between a ½ and ¾ 3. Between a ¼ and ½ 2. Less than ¼ of the value of all my crops 1. None
Crop damage from insects	Ecosystem cost	In the last five years how much damage has there been to your crops because of insects? Scale 1-5: see options below 5. More than ¾ 4. Between a ½ and ¾ 3. Between a ¼ and ½ 2. Less than ¼ of the value of all my crops 1. None
Paid forest management	Direct personal impacts	Binary response (Y/N) In the last five years were you paid to do any forest management activities (e.g. patrols, monitoring, tree planting, meetings, training)?
Timber revenue		At village scale, annual forest revenue in year 2017-2018 per head of population based on village records.
Health infrastructure	Shared community development benefits	In the last five years, have you noticed any changes in health facilities available in your village? (for example: availability of medicine, facilities at your local dispensary, doctor and nurses, solar lights in hospital) Scale 1-5: 1 = much worse, 3 = no change, 5= much better
Water infrastructure	Shared community development benefits	In the last five years, have you noticed any changes in water infrastructure in this village? (For example: communal tap, well, reservoir, water pump, irrigation) **Scale 1-5:** 1 = much worse**,** 3 = no change, 5= much better
Education infrastructure	Shared community development benefits	In the last five years, have you noticed any changes in education facilities in your village? (For example, facilities for teachers, sanitation (toilets) at school, other education facilities such as quality of school building, books) Scale 1-5: 1 = much worse, 3 = no change, 5= much better

Forest resource availability		Timber and non-timber forest products were important for subsistence and commercial livelihood activities in the study region, including collection of firewood, poles for housing construction and forest fruits and vegetables for consumption. Therefore the availability, or more specifically availability per unit of population i.e. pressure on forest resources influences livelihood activities and wellbeing (Platz, 2012). Forest resource availability was calculated as forest cover in 2019 (Hansen, 2020) within a 10 km radius divided by population density (world pop, 2015), weighted so that more remote populations exert less pressure on forest resources using a sigma function of 5 so that relative weight reduces to zero at approximately 10 km to account for forest use by local users. Forest cover in 2019 was used to align with the year the quantitative survey was undertaken.
Yield		How has the yield of your main cash crop changed in the last 5 years? Note: yield is bags per acre Scale 1-5: 1 = decrease a lot, 3 = the same, 5 = increase a lot
Attitude	Wellbeing to Conservation	Likert scale 1-5 12.7. Do you think the forest owned by your village should continue to be conserved or converted to farmland? Please choose the option that best describes your opinion: Scale 1-5 1= convert all forest to farmland 2= convert most forest to farmland 3= shared equally between farmland and forest 4= most should be kept as forest, but some small parts should be converted to farmland. 5= all forest should be conserved

S4.4 Matching

The matching steps are described in the supporting information of chapter three.

S4.5 Structural Equation modelling

We tested the conceptual model by structural equation modelling using PiecewiseSEM in R (Lefcheck, 2016). Throughout the quantitative testing of our conceptual model we adopt a forward causality research design, testing the effect size of particular causes, rather than reverse causality, which seeks to explain the maximum variation in the effect (e.g. wellbeing) based on all potential causes (Gelman and Imbens, 2013). The former is common to impact evaluation frameworks and has the benefit that by isolating particular effects, impact evaluations can provide more precise estimates of specific causes (Oldekop et al., 2020), though reverse causality is more commonly adopted in ecological modelling. We therefore report conceptual models containing all identified paths, but omitting exogenous confounding socio-ecological variables, as these do not test key hypotheses, but are included to account for confounding variation only (S5.1).

Table S4.5.1. Standardised coefficients, standard errors, p values from the structural equation model. Grey predictors donate confounding (exogenous) socio-ecological variables.

Response	Predictor	Estimate	Std. error	p value	Std. estimate
governance equity	CF	-	-	<0.001	-
governance equity	CF = 0	-0.114	0.039	0.009	-
governance equity	CF = 1	0.248	0.037	<0.001	-
governance equity	gender	-	-	0.36	-
governance equity	gender = Male	0.054	0.030	0.089	-
governance equity	gender = Female	0.080	0.032	0.023	-
governance equity	elite	-	-	<0.001	-
governance equity	elite = 0	-0.123	0.027	<0.001	-
governance equity	elite = 1	0.257	0.037	<0.001	-
wellbeing	governance equity	0.039	0.009	<0.001	0.179
wellbeing	yield	0.013	0.004	<0.001	0.140
wellbeing	forest resource availability	0.010	0.005	0.038	0.088
wellbeing	road distance	-0.012	0.004	0.006	-0.120
wellbeing	elite	-	-	0.021	-
wellbeing	elite = 0	0.558	0.004	<0.001	-
wellbeing	elite = 1	0.578	0.008	<0.001	-
wellbeing	gender	-	-	<0.001	-
wellbeing	gender = Female	0.553	0.006	<0.001	-
wellbeing	gender = Male	0.583	0.006	<0.001	-
forest recovery	CF	-	-	0.019	-
forest recovery	CF = 0	-0.008	0.016	0.606	-

forest recovery	CF = 1	0.044	0.016	0.014	-
forest recovery	distance to NFRs	-0.010	0.001	<0.001	-0.168
forest recovery	village distance	-0.011	0.001	<0.001	-0.165
forest recovery	elevation	0.010	0.001	<0.001	0.157
forest recovery	precipitation	-0.016	0.003	<0.001	-0.245
forest recovery	elite	-	-	0.076	-
forest recovery	elite = 1	0.017	0.011	0.152	-
forest recovery	elite = 0	0.019	0.011	0.118	-
forest recovery	gender	-	-	0.108	-
forest recovery	gender = Male	0.017	0.011	0.148	-
forest recovery	gender = Female	0.019	0.011	0.122	-
ecosystem benefits	communal benefits	0.183	0.038	<0.001	0.176
ecosystem benefits	elevation	0.080	0.031	0.01	0.219
ecosystem benefits	distance to NFRs	0.024	0.022	0.28	0.071
ecosystem benefits	gender	-	-	0.094	-
ecosystem benefits	gender = Male	-0.046	0.053	0.398	-
ecosystem benefits	gender = Female	-0.010	0.053	0.854	-
ecosystem costs	precipitation	-0.093	0.031	0.003	-0.192
ecosystem costs	gender	-	-	<0.001	-
ecosystem costs	gender = Male	-0.110	0.036	0.007	-
ecosystem costs	gender = Female	0.137	0.038	0.002	-
ecosystem costs	elite	-	-	0.117	-
ecosystem costs	elite = 1	-0.016	0.043	0.712	-
ecosystem costs	elite = 0	0.044	0.033	0.199	-

communal benefits	governance equity	0.085	0.026	0.001	0.11
communal benefits	elevation	-0.165	0.028	<0.001	-0.472
communal benefits	road distance	-0.037	0.020	0.067	-0.109
communal benefits	distance to NFRs	0.090	0.021	<0.001	0.278
communal benefits	village distance	-0.041	0.022	0.063	-0.115
communal benefits	precipitation	0.064	0.037	0.083	0.181
communal benefits	gender	-	-	<0.001	-
communal benefits	gender = Male	-0.066	0.038	0.102	-
communal benefits	gender = Female	0.094	0.039	0.027	-
forest resource availability	forest recovery	5.610	1.919	0.003	0.396
forest resource availability	precipitation	0.709	0.147	<0.001	0.751
forest resource availability	road distance	0.944	0.045	<0.00	1.036
forest resource availability	village distance	0.092	0.056	0.1	0.096
forest resource availability	distance to NFRs	0.097	0.053	0.069	0.111
forest resource availability	elevation	0.069	0.074	0.353	0.073
attitude	CF	-	-	0.001	-
attitude	CF = 0	4.008	0.110	<0.001	-
attitude	CF = 1	4.482	0.104	<0.001	-
attitude	governance equity	0.520	0.105	<0.001	0.196
attitude	forest employment	0.134	0.090	0.134	0.056

attitude	precipitation	-0.154	0.077	0.044	-0.123
attitude	road distance	-0.087	0.063	0.165	-0.072
attitude	gender	-	-	<0.001	-
attitude	gender = Female	3.871	0.088	<0.001	-
attitude	gender = Male	4.619	0.082	<0.001	-
yield	ecosystem benefits	0.340	0.134	0.011	0.111
yield	ecosystem costs	-0.309	0.089	<0.001	-0.134
yield	village distance	0.109	0.069	0.115	0.097
yield	road distance	-0.083	0.065	0.205	-0.077
yield	gender	-	-	<0.001	-
yield	gender = Female	2.122	0.109	<0.001	-
yield	gender = Male	2.570	0.105	<0.001	-
yield	elite	-	-	0.18	-
yield	elite = 0	2.288	0.099	<0.001	-
yield	elite = 1	2.405	0.117	<0.001	-
forest employment	CF	-	-	0.075	-
forest employment	CF = 0	-0.382	0.355	0.28	-
forest employment	CF = 1	0.544	0.341	0.11	-
forest employment	governance equity	1.147	0.290	<0.001	-
forest employment	village distance	-0.561	0.234	0.016	-
forest employment	precipitation	0.661	0.241	0.006	-
forest employment	distance to NFRs	0.329	0.187	0.079	-
forest employment	road distance	-0.272	0.180	0.131	-
forest employment	elite	-	-	<0.001	-

forest employment	elite = 0	-1.426	0.239	<0.001	-
forest employment	elite = 1	1.587	0.325	<0.001	-
forest employment	gender	-	-	0.175	-
forest employment	gender = Male	-0.060	0.246	0.808	-
forest employment	gender = Female	0.222	0.260	0.393	-

Table S4.5.2. Goodness of fit of the structural equation model. The goodness of fit can be evaluated by Fisher's C test for directed separation. This tests whether the causal structure specified by the model reflects the data by evaluating if other paths between variables which have been omitted from the model are truly independent (not associated at a statistically significant value such as 0.05). The piecewiseSEM package combines all independence tests into a single p value and the model structure is accepted when p > 0.05 (Lefcheck, 2016)**.**

AIC	Fisher's C	df	p value
335.5	165.5	138	0.06

References

Adams, V. M., Pressey, R. L. and Naidoo, R. (2010). Opportunity costs: Who really pays for conservation? *Biological Conservation*, **143** (2), pp.439–448. [Online]. Available at: doi:10.1016/j.biocon.2009.11.011.

Adedokun, O. A., Childress, A. L. and Burgess, W. D. (2011). Testing Conceptual Frameworks of Nonexperimental Program Evaluation Designs Using Structural Equation Modeling. *American Journal of Evaluation*, **32** (4), pp.480-493. [Online]. Available at: doi:10.1177/1098214011401368.

Agarwal, B. (2009). Gender and forest conservation: The impact of women's participation in community forest governance. *Ecological Economics*, **68** (11), pp.2785–2799. [Online]. Available at: doi:10.1016/j.ecolecon.2009.04.025.

Agarwala, M. et al. (2014). Assessing the Relationship Between Human Well-being and Ecosystem Services: A Review of Frameworks. *Conservation and Society*, **12** (4), p.437. [Online]. Available at: doi:10.4103/0972-4923.155592.

Agrawal, A., Chhatre, A. and Hardin, R. (2008). Changing Governance of the World's Forests. *Science,* 320 (June), pp.1460–1462. [Online]. Available at: doi:10.1126/science.320.5882.1435.

Ahrends, A. et al. (2010). Predictable waves of sequential forest degradation and biodiversity loss spreading from an African city. *Proceedings of the National Academy of Sciences*, **107** (33), pp.14556–14561. [Online]. Available at: doi:10.1073/pnas.0914471107.

Alkire, S. and Bank, T. W. (2002). Dimensions of Human Development. *World Development*, 30 (2), pp.181–205.

Alkire, S. and Foster, J. (2011). Understandings and misunderstandings of multidimensional poverty measurement. *Journal of Economic Inequality*, **9** (2), pp.289–314. [Online]. Available at: doi:10.1007/s10888-011-9181-4.

Allen, C. et al. (2018). Initial progress in implementing the Sustainable Development Goals (SDGs): a review of evidence from countries. *Sustainability Science*, **13** (5), pp.1453–1467. [Online]. Available at: doi:10.1007/s11625-018-0572-3.

Andam, K. S. et al. (2008). Measuring the effectiveness of protected area networks in reducing deforestation. *Proceedings of the national academy of sciences*, **105** (42), pp.15–20.

Andam, K. S. et al. (2010). Protected areas reduced poverty in Costa Rica and Thailand. *Proceedings of the national academy of sciences*, **107**(22), pp.9996-10001. [Online]. Available at: doi:10.1073/pnas.0914177107.

Angelsen, A. Larsen, H. O., Lund, J. F., Smith-Hall, C., Wunder, S (eds). (2011). *Measuring livelihoods and environmental dependence: Methods for research and fieldwork*. Oxford: Earthscan.

Braun, V., and Clarke, V. (2013). Successful Qualitative Research: A Practical Guide for Beginners. SAGE: London.

Angelsen, A. et al. (2017). Learning from REDD+: a response to Fletcher et al. *Conservation Biology*, **31** (3), pp.718–720. [Online]. Available at: doi:10.1111/cobi.12933.

Atela, J. O. et al. (2015). Implementing REDD+ in view of integrated conservation and development projects: Leveraging empirical lessons. *Land Use Policy*, **48**, pp.329–340. [Online]. Available at: doi:10.1016/j.landusepol.2015.06.011.

Baird, T. D. (2014). Conservation and unscripted development: Proximity to park associated with development and financial diversity. *Ecology and Society*, **19** (1). [Online]. Available at: doi:10.5751/ES-06184-190104.

Baker, K., Eichhorn, M. P. and Griffiths, M. (2019). Decolonizing field ecology. *Biotropica*, **51** (3), pp.288–292. [Online]. Available at: doi:10.1111/btp.12663.

Barsimantov, J. A. (2010). Vicious and Virtuous Cycles and the Role of External Non-government Actors in Community Forestry in Oaxaca. *Human ecology*, **38**(1), [Online]. Available at: doi:10.1007/s10745-009-9289-3.

Barton, D. N., Blumentrath, S. and Rusch, G. (2013). Policyscape-A Spatially Explicit Evaluation of Voluntary Conservation in a Policy Mix for Biodiversity Conservation in Norway. *Society and Natural Resources*, **26** (10), pp.1185–1201. [Online]. Available at: doi:10.1080/08941920.2013.799727.

Barton, D. N., Ring, I. and Rusch, G. M. (2017). Policy Mixes: Aligning instruments for biodiversity conservation and ecosystem service provision. *Environmental Policy and Governance*, **27**(5), pp.397-403. [Online]. Available at: doi:10.1002/eet.1779.

Barton, K. (2019). MuMIn: Multi-Model Inference. R package version 1.43.6. https://CRAN.R-project.org/package=MuMIn

Basnyat, B. et al. (2018). Legal-sounding bureaucratic re-centralisation of community forestry in Nepal. *Forest Policy and Economics*, **91**, 5–18. [Online]. Available at: doi:10.1016/j.forpol.2017.08.010.

Bastakoti, R. R. and Davidsen, C. (2017). Optimism, hopes and fears: local perceptions of REDD+ in Nepalese community forests. *International Forestry Review*, **19** (1), pp.1–16. [Online]. Available at: doi:10.1505/146554817820888627.

Baylis, K. et al. (2016). Mainstreaming Impact Evaluation in Nature Conservation. *Conservation Letters*, 9 (1), pp.58–64. [Online]. Available at: doi:10.1111/conl.12180.

Beauchamp, E. et al. (2018). "Living a good life": Conceptualizations of well-being in a conservation context in Cambodia. *Ecology and Society*, **23** (2). [Online]. Available at: doi:10.5751/ES-10049-230228.

Beauchamp, E., Clements, T. and Milner-gulland, E. J. (2018). Assessing Medium-term Impacts of Conservation Interventions on Local Livelihoods in Northern Cambodia. *World Development*, **101**, pp.202–218. [Online]. Available at: doi:10.1016/j.worlddev.2017.08.008.

Bennett, E. M. et al. (2015). Linking biodiversity, ecosystem services, and human well-being: three challenges for designing research for sustainability. *Current Opinion in Environmental Sustainability*, *14*, pp.76–85. [Online]. Available at: doi:10.1016/j.cosust.2015.03.007.

Bennett, N. J. (2016). Using perceptions as evidence to improve conservation and environmental management. *Conservation Biology*, **30** (3), pp.582–592. [Online]. Available at: doi:10.1111/cobi.12681.

Bernstein, S. and Cashore, B. (2012). Complex global governance and domestic policies: Four pathways of influence. *International Affairs*, **88** (3), pp.585–604. [Online]. Available at: doi:10.1111/j.1468-2346.2012.01090.x.

Biedenweg, K., Stiles, K. and Wellman, K. (2016). A holistic framework for identifying human wellbeing indicators for marine policy. *Marine Policy*, **64**, pp.31–37. [Online]. Available at: doi:10.1016/j.marpol.2015.11.002.

Blackman, A., Goff, L. and Rivera Planter, M. (2018). Does eco-certification stem tropical deforestation? Forest Stewardship Council certification in Mexico. *Journal of Environmental Economics and Management*, **89**, pp.306–333. [Online]. Available at: doi:10.1016/j.jeem.2018.04.005.

Blomley, T. et al. (2017). When community forestry meets REDD +: has REDD + helped address implementation barriers to participatory forest management in Tanzania? *Journal of East African Studies*, **11**(3), 549-570. [Online]. Available at: doi:10.1080/17531055.2017.1356623.

Blomley, T. et al. (2019). Scaling up community forest enterprises in Tanzania. *Briefing*, London.

Boarini, R., Kolev, A. and McGregor, A. (2014). *Measuring well-being and progress in countries at different stages of development: Towards a more universal conceptual framework*. OECD Development Center Working Papers, 325 (325), pp.1–59. [Online]. Available at: doi:10.1787/5jxss4hv2d8n-en.

Booker, F. and Franks, P. (2018). *Governance Assessment for Protected and Conserved Areas (GAPA) Methodology manual for GAPA facilitators*. London: IIED. [Online]. Available at: http://pubs.iied.org/17655IIED.

Borrini-feyerabend, G., Dudley, N. and Jaeger, T. (2013). *Governance of Protected Areas. From understanding to action*. Best practice protected area guidelines series No. 20, Gland, Switzerland: IUCN. CBD.

Bowler, D. E. et al. (2012). Does community forest management provide global environmental benefits and improve local welfare? *Frontiers in Ecology and the Environment*, **10**(1), pp.29-36. [Online]. Available at: doi:10.1890/110040.

Braber, B. Den, Evans, Karl, L. and Oldekop, J. (2018). Impact of protected areas on poverty, extreme poverty, and inequality in Nepal. *Conservation Letters*, **11**(6), p.e12576. . [Online]. Available at: doi:10.1111/conl.12576.

Brauman, K. A. et al. (2020). Global trends in nature's contributions to people. *Proceedings of the National Academy of Sciences of the United States of America*, **117** (51), pp.32799–32805. [Online]. Available at: doi:10.1073/pnas.2010473117.

Bray, D. B. et al. (2008). Tropical Deforestation, Community Forests, and Protected Areas in the Maya Forest. *Ecology & Society*, **13** (2).

Brende, B. and Bent, H. H. (2015). Towards evidence-based, quantitative Sustainable Development Goals for 2030. *Lancet*, **385** [Online]. Available at: doi:10.1016/S0140-6736(14)61654-8.

Breslow, S. J. et al. (2017). Evaluating indicators of human well-being for ecosystem-based management. *Ecosystem Health and Sustainability*, **3** (12), pp.1–18. [Online]. Available at: doi:10.1080/20964129.2017.1411767.

Breslow SJ, Sojka B, Barnea R, Basurto X, Carothers C, Charnley S, Coulthard S, Dolšak N, Donatuto J, García-Quijano C, H. C. (2016). Conceptualizing and operationalizing human wellbeing for ecosystem assessment and management. *Environmental Science & Policy*, **66**, pp.250-259. [Online]. Available at: doi:10.1016/j.envsci.2016.06.023.

Brockington, D. (2003). Injustice and conservation - is 'local support' necessary for sustainable protected areas? *Policy Matters*, **62** (25), pp.6–15.

Brown, K. and Westaway, E. (2011). Agency, Capacity, and Resilience to Environmental Change: Lessons from Human and Disasters. *Annual Review of Environment and Resources*, **36,** pp.321–342. [Online]. Available at: doi:10.1146/annurev-environ-052610-092905.

Burivalova, Z. et al. (2017). A Critical Comparison of Conventional, Certified, and Community Management of Tropical Forests for Timber in Terms of Environmental, Economic, and Social Variables. *Conservation Letters*, **10** (1), pp.4–14. [Online]. Available at: doi:10.1111/conl.12244.

Burnham, K., Anderson, D. R. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer, New York.

Burnham, K. P., Anderson, D. R. and Huyvaert, K. P. (2011). AIC model selection and multimodel inference in behavioral ecology: Some background, observations, and comparisons. *Behavioral Ecology and Sociobiology*, **65** (1), pp.23–35. [Online]. Available at: doi:10.1007/s00265-010-1029-6.

Cafiero, C., Viviani, S. and Nord, M. (2017). Food security measurement in a global context: The food insecurity experience scale. *Measurement,* **116**, pp.146–152. [Online]. Available at: doi:10.1016/j.measurement.2017.10.065.

Caillon, S. et al. (2017). Moving beyond the human–nature dichotomy through biocultural approaches: Including ecological well-being in resilience indicators. *Ecology and Society*, **22** (4). [Online]. Available at: doi:10.5751/ES-09746-220427.

Calvet-Mir, L. et al. (2015). Payments for ecosystem services in the tropics: a closer look at effectiveness and equity. *Current Opinion in Environmental Sustainability*, **14**, pp.150–162. [Online]. Available at: doi:10.1016/j.cosust.2015.06.001.

Camfield, L., Crivello, G. and Woodhead, M. (2009). Wellbeing research in developing countries: Reviewing the role of qualitative methods. *Social Indicators Research*, **90** (1), pp.5–31. [Online]. Available at: doi:10.1007/s11205-008-9310-z.

Campbell, B. M., Sayer, J. A. and Walker, B. (2018). Navigating Trade-Offs: Working for Conservation and Development. *Ecology & Society*, **15** (2).

Capitani, C. et al. (2019). Scenarios of Land Use and Land Cover Change and Their Multiple Impacts on Natural Capital in Tanzania. *Environmental Conservation*, **46** (1), pp.17–24. [Online]. Available at: doi:10.1017/S0376892918000255.

Carr, J. A. et al. (2021). Anticipated impacts of achieving SDG targets on forests - a review. *Forest Policy and Economics*, **126**, p.102423. [Online]. Available at: doi:10.1016/j.forpol.2021.102423.

CBD (Convention on Biological Diversity). 2016. CoP 13 Decision XIII/6

Cerutti, P. O. (2014). Social impacts of the Forest Stewardship Council certification An assessment in the Congo basin. Volume 103. CIFOR.

Coad, L. et al. (2008). The Costs and Benefits of Forest Protected Areas for Local Livelihoods: a review of the current literature. p.45.

Corrigan, C. et al. (2017). Global Review of Social Indicators used in Protected Area Management Evaluation. *Conservation letters*, **11**, e12397. [Online]. Available at: doi:10.1111/conl.12397.

Costanza, R., Kubiszewski, I. and Giovannini, E. (2014). Time to leave GDP behind. *Nature*, **505**, pp.2–7.

Crawley, M. (2007). *The R book*. Chichester : John Wiley & Sons, Ltd. [Online]. Available at: doi:10.16242/j.cnki.umst.2014.04.005.

Cusack, J. J. et al. (2021). Measuring the intensity of conflicts in conservation. *Conservation Letters*, p.e12783. [Online]. Available at: doi:10.1111/conl.12783.

Dasgupta, P. (2001). *Human wellbeing and the natural environment*. Oxford University Press.

Davies, R. (2007). *The 2006 Basic Necessities Survey (BNS) in Can Loc District, Ha Tinh Province, Vietnam*. Pro Poor Centre. [Online]. Available at: doi:10.13140/RG.2.2.17664.28165.

Daw, T. et al. (2011). Applying the ecosystem services concept to poverty alleviation: the need to disaggregate human well-being. *Environmental Conservation*, **38** (04), pp.370–379. [Online]. Available at: doi:10.1017/S0376892911000506.

Dawson, N. et al. (2018). Assessing Equity in Protected Area Governance: Approaches to Promote Just and Effective Conservation. Conservation Letters, 11(2), p.e12388. [Online]. Available at: doi:10.1111/conl.12388.

Dawson, N. M. et al. (2017). Environmental justice research shows the importance of social feedbacks in ecosystem service trade-offs. *Ecology and Society*, **22**(3). [Online]. Available at: doi:10.5751/ES-09481-220312.

Dawson, N., Martin, A. and Danielsen, F. (2018). Assessing Equity in Protected Area Governance: Approaches to Promote Just and Effective Conservation. *Conservation Letters*, **11** (2), pp.1–8. [Online]. Available at: doi:10.1111/conl.12388.

de Lange, E., Woodhouse, E., Milner-Gulland, E.J., 2016. Approaches Used to Evaluate the Social Impacts of Protected Areas. *Conservation Letters*, **9**, 327–333. https://doi.org/10.1111/conl.12223

DeFries, R., Karanth, K. K. and Pareeth, S. (2010). Interactions between protected areas and their surroundings in human-dominated tropical landscapes. *Biological Conservation*, **143** (12), pp.2870–2880. [Online]. Available at: doi:10.1016/j.biocon.2010.02.010.

DeFries, R. uth et al. (2007). Land Use Change Around Protected Areas : Management To Balance Human Needs and Ecological Function. *Ecological Applications*, **17**, pp.1031–1038. [Online]. Available at: doi:10.1890/05-1111.

DFID, (2000). Sustainable Livelihoods Guidance Sheets. Department for International Development. Available: http://www.livelihoods.org/

Diao, X., Hazell, P. and Thurlow, J. (2010). The Role of Agriculture in African Development. *World Development,* **38** (10), pp.1375–1383. [Online]. Available at: doi:10.1016/j.worlddev.2009.06.011.

Dietz, T., Ostrom, E. and Stern, P. C. (2003). The struggle to govern the commons. *Science*, **302** (5652), pp.1907–1912. [Online]. Available at: doi:10.1007/978-0-387-73412-5_40.

Dormann, C. F. et al. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography,* **36** (1), pp.027–046. [Online]. Available at: doi:10.1111/j.1600-0587.2012.07348.x.

Dressler, W. et al. (2016). Land sharing not sparing in the "green economy": The role of livelihood bricolage in conservation and development in the Philippines. *Geoforum*, **76**, pp.75–89. [Online]. Available at: doi:10.1016/j.geoforum.2016.09.003.

Evans, K., Murphy, L. and Jong, W. De. (2014). Global versus local narratives of REDD : A case study from Peru' s Amazon. *Environmental Science and Policy*, **35**, pp.98–108. [Online]. Available at: doi:10.1016/j.envsci.2012.12.013.

Ewers, R. M. and Rodrigues, A. S. L. (2008). Estimates of reserve effectiveness are confounded by leakage. *Trends in Ecology and Evolution,* **23** (3), pp.113–116. [Online]. Available at: doi:10.1016/j.tree.2007.11.008.

FAO. (2020). *Global Forest Resources Assessment 2020: Main report*. Rome. [Online]. Available at: doi:https://doi.org/10.4060/ca9825en.

Fazey, I., Schäpke, N., Caniglia, G., Patterson, J., Hultman, J., Van Mierlo, B., Säwe, F., Wiek, A., Wittmayer, J., Aldunce, P. and Al Waer, H., 2018. Ten essentials for action-oriented and second order energy transitions, transformations and climate change research. *Energy Research & Social Science*, **40**, pp.54-70.

Fazey, I., Schäpke, N., Caniglia, G., Hodgson, A., Kendrick, I., Lyon, C., Page, G., Patterson, J., Riedy, C., Strasser, T. and Verveen, S. (2020). Transforming knowledge systems for life on Earth: Visions of future systems and how to get there. *Energy research & social science***, 70**, p.101724.

Ferraro, P. J. and Hanauer, M. M. (2014b). Quantifying causal mechanisms to determine how protected areas affect poverty through changes in ecosystem services and infrastructure. *Proceedings of the National Academy of Sciences*, **111** (11), pp.4332–4337. [Online]. Available at: doi:10.1073/pnas.1307712111.

Ferraro, P. J. and Hanauer, M. M. (2015). Through what mechanisms do protected areas affect environmental and social outcomes? *Philosophical Transactions of the Royal Society B: Biological Sciences*, **370** (1681), p.20140267. [Online]. Available at: doi:10.1098/rstb.2014.0267.

Ferraro, P. J., Hanauer, M. M. and Sims, K. R. E. (2011). Conditions associated with protected area success in conservation and poverty reduction. *Proceedings of the National Academy of Sciences*, **108**(34), pp.13913-13918. [Online]. Available at: doi:10.1073/pnas.1011529108.

Ferraro, P. J. and Miranda, J. J. (2014). The performance of non-experimental designs in the evaluation of environmental programs: A design-replication study using a large-scale randomized experiment as a benchmark. *Journal of Economic Behavior and Organization*, **107**, pp.344–365. [Online]. Available at: doi:10.1016/j.jebo.2014.03.008.

Finer, M. et al. (2008). Oil and gas projects in the Western Amazon: Threats to wilderness, biodiversity, and indigenous peoples. *PLoS ONE*, **3** (8). [Online]. Available at: doi:10.1371/journal.pone.0002932.

Fletcher, R. et al. (2016). Questioning REDD+ and the future of market-based conservation. *Conservation Biology*, **30** (3), pp.673–675. [Online]. Available at: doi:10.1111/cobi.12680.

Folke, C. et al. (2002). Resilience and sustainable development Sustainable Development: Building Adaptive Capacity in a World of Transformations. *Ambio,* **31** (5), pp.437–440. [Online]. Available at: doi:10.1579/0044-7447-31.5.437.

Franks, P. and Booker, F. (2018). *Governance Assessment for Protected and Conserved Areas (GAPA). Early experience of a multi- stakeholder methodology for enhancing equity and effectiveness*. London: IIED.

Franks, P., Booker, F. and Roe, D. (2018). *Understanding and assessing equity in protected area conservation*. London: IIED.

Franks, P. and Small, R. (2016). *Social Assessment for Protected Areas (SAPA) Methodology Manual for SAPA Facilitators*. London: IIED.

Fraser, N. (1995). Recognition or Redistribution? A Critical Reading of Iris Young's Justice and the Politics of Difference. *Journal of Political Philosophy*, **3** (2), pp.166–180. [Online]. Available at: doi:10.1111/j.1467-9760.1995.tb00033.x.

Frey, G., Charnley, S. and Makala, J. (2020). *The Costs, Benefits, and Sustainability of Community-based Forest Management and Certification for Community Forest Enterprises in Southeastern Tanzania*. Project report to U.S. Forest Service International Programs.

Fry, P. et al. (2015). Monitoring local wellbeing in environmental interventions: A consideration of practical trade-offs. *Oryx*, **44** (0), pp.1–9. [Online]. Available at: doi:10.1017/S003060531500112X.

Galvin, K. A., Beeton, T. A. and Luizza, M. W. (2018). African community-based conservation: A systematic review of social and ecological outcomes. *Ecology and Society*, **23** (3). [Online]. Available at: doi:10.5751/ES-10217-230339.

FSC (The Forest Stewardship Council) (2018). *The FSC National Forest Stewardship Standard of Tanzania Mainland.* July 2018. Available: https://fsc.org/en/document-centre/documents/resource/289

Gavin, M. C. et al. (2018). Protected land: Many factors shape success. *Science*, **361** (6402), p.561. [Online]. Available at: doi:10.1126/science.aau5168.

Gelman, A. and Imbens, G. (2013). *Why ask why? Forward causal inference and reverse causal questions*. National Bureau of Economic Research Working Paper 19614.

Gentle, P. and Narayan, T. (2012). Climate change, poverty and livelihoods: Adaptation practices by rural mountain communities in Nepal. *Environmental Science and Policy*, **21**, pp.24–34. [Online]. Available at: doi:10.1016/j.envsci.2012.03.007.

Gertler, P. J. et al. (2016). *Impact evaluation in practice*. 2nd edition. Washington, DC : World Bank. [Online]. Available at: doi:10.1109/WI-IATW.2006.145.

Gibbs, H. K. et al. (2010). Tropical forests were the primary sources of new agricultural land in the 1980s and 1990s. *Proceedings of the National Academy of Sciences of the United States of America*, **107** (38), pp.16732–16737. [Online]. Available at: doi:10.1073/pnas.0910275107.

Gibson, L. et al. (2011). Primary forests are irreplaceable for sustaining tropical biodiversity. *Nature*, **478** (7369), pp.378–381. [Online]. Available at: doi:10.1038/nature10425.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, **202**, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Goodwin, N. R. (2003). *Five Kinds of Capital: Useful Concepts for Sustainable Development*. G-DAE Working Paper, 03–07 (September).

Gross-camp, N. (2017). Tanzania' s community forests : their impact on human well-being and persistence in spite of the lack of benefit. *Ecology and Society*, **22** (1).

Gross-camp, N. et al. (2019). The Type of Land We Want: Exploring the Limits of Community Forestry in Tanzania and Bolivia. *Sustainability*, **11**(6) pp.1–16. [Online]. Available at: doi:10.3390/su11061643.

Guisan, A. and Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological modelling,* **135**, pp.147–186.

Gunderson, L. H. and Holling, C. S. (2002). *Panarchy: Understanding transformations in human and natural systems*. Washington : Island Press.

Gurney, G. G. et al. (2015). Integrated conservation and development: evaluating a community-based marine protected area project for equality of socioeconomic impacts. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **370** (1681), p.20140277. [Online]. Available at: doi:10.1098/rstb.2014.0277.

Hajjar, R. et al. (2020). A global analysis of the social and environmental outcomes of community forests. *Nature Sustainability*, **4**(3), pp.216-224. . [Online]. Available at: doi:10.1038/s41893-020-00633-y.

Hajjar, R., Engbring, G. and Kornhauser, K. (2021). The impacts of REDD+ on the social-ecological resilience of community forests. *Environmental Research Letters*, **16** (2), p.024001. [Online]. Available at: doi:10.1088/1748-9326/abd7ac.

Hajjar, R. and Oldekop, J. A. (2018). Research frontiers in community forest management. *Current Opinion in Environmental Sustainability*, **32**, pp.119–125. [Online]. Available at: doi:10.1016/j.cosust.2018.06.003.

Hansen, M. C. C. et al. (2013). High-Resolution Global Maps of. *Science*, **342**, pp.850–854. [Online]. Available at: doi:10.1126/science.1244693.

Haq, M. (1996). *Reflections on Human Development*. Oxford University Press: New York.

Hicks, B. C. C. et al. (2016). Engage key social concepts for sustainability. *Science,* **352** (6281).

Ho D.E, Imai K, King G, Stuart EA. (2007). MatchIt: nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software*. **42**(8), pp. 1-28. 32.

Holmes, G. (2007). Protection, Politics and Protest Understanding Resistance to Conservation. *Conservation & Society*, 5 (2), pp.184–201.

Hoque, S. F., Quinn, C. H. and Sallu, S. M. (2017). Resilience, political ecology, and well-being: An interdisciplinary approach to understanding social-ecological change in coastal Bangladesh. *Ecology and Society*, 22 (2). [Online]. Available at: doi:10.5751/ES-09422-220245.

Howard, R. J. et al. (2016). Which 'fairness', for whom, and why? An empirical analysis of plural notions of fairness in Fairtrade Carbon Projects, using Q methodology. *Environmental Science and Policy*, **56**, pp.100–109. [Online]. Available at: doi:10.1016/j.envsci.2015.11.009.

Howe, C. et al. (2014). Creating win-wins from trade-offs? Ecosystem services for human well-being: A meta-analysis of ecosystem service trade-offs and synergies in the real world. *Global Environmental Change*, **28** (1), pp.263–275. [Online]. Available at: doi:10.1016/j.gloenvcha.2014.07.005.

Humphries, S. et al. (2012). Are community-based forest enterprises in the tropics financially viable? Case studies from the Brazilian Amazon. *Ecological Economics*, **77**, pp.62–73. [Online]. Available at: doi:10.1016/j.ecolecon.2011.10.018.

Ingram, V., Ros-Tonen, M. A. F. and Dietz, T. (2015). A fine mess: Bricolaged forest governance in Cameroon. *International Journal of the Commons*, **9** (1), pp.41–64. [Online]. Available at: doi:10.18352/ijc.516.

ISRIC (International Soil Reference and Information Centre). 2019. GIS layers Soil pH product code PHIHOH_T__M, Soil cation exchange product code CEC_T__M and Soil clay content product code CLYPPT_T__M. available from https://data.isric.org/

James, C. A., Kershner, J. and Levin, P. S. (2012). A Methodology for Evaluating and Ranking Water Quantity Indicators in Support of Ecosystem-Based Management. *Environmental Management*, **49**(3), pp.703-719

Jayachandran, S. et al. (2017). Cash for carbon: A randomized trial of payments for ecosystem services to reduce deforestation. *Science*, **357**(6348), 267-273.

Jensen, O. and Wu, H. (2018). Urban water security indicators: Development and pilot. *Environmental Science and Policy*, **83**, pp.33–45. [Online]. Available at: doi:10.1016/j.envsci.2018.02.003.

Joppa, L. N. and Pfaff, A. (2011). Global protected area impacts. **Proceedings of the Royal Society B: Biological Sciences**, *278* (1712), pp.1633–1638. [Online]. Available at: doi:10.1098/rspb.2010.1713.

Joppa, L. and Pfaff, A. (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**, pp.135–149. [Online]. Available at: doi:10.1111/j.1749-6632.2009.05162.x.

Kaimowitz, D. (2003). Forest law enforcement and rural livelihoods. *International Forestry Review*, **5** (3), pp.199–210. [Online]. Available at: doi:10.4324/9781849771672.

Kalonga, S.K., Kulindwa, K.A., Mshale, B.I., and Håkon, L. (2014). *The influence of certified forest management on community socioeconomics: Implication for REDD+ safeguard standards implementation in Tanzania*. Proceedings of the International Conference on Reducing Climate Change Challenges through Forestry and Other Land Use Practices, 1.

Kalonga, S., Midtgaard, F. and Eid, T. (2015). Does Forest Certification Enhance Forest Structure? Empirical Evidence from Certified Community-Based Forest Management in Kilwa District , Tanzania. *International Forestry Review*, **17**(2), pp.182-194. [Online]. Available at: doi:10.1505/146554815815500570.

Kapos, V. et al. (2009). Outcomes, not implementation, predict conservation success. *Oryx*, **43** (3), pp.336–342. [Online]. Available at: doi:10.1017/S0030605309990275.

Karjalainen, E., Sarjala, T. and Raitio, H. (2010). Promoting human health through forests: Overview and major challenges. *Environmental Health and Preventive Medicine*, **15** (1), pp.1–8. [Online]. Available at: doi:10.1007/s12199-008-0069-2.

King, M. F., Renó, V. F. and Novo, E. M. L. M. (2014). The Concept, Dimensions and Methods of Assessment of Human Well-Being within a Socioecological Context: A Literature Review. *Social Indicators Research*, **116** (3), pp.681–698. [Online]. Available at: doi:10.1007/s11205-013-0320-0.

Klugman, J., Rodríguez, F. and Choi, H. J. (2011). The HDI 2010: New controversies, old critiques. *Journal of Economic Inequality*, **9** (2), pp.249–288. [Online]. Available at: doi:10.1007/s10888-011-9178-z.

Komives, K. et al. (2018). Conservation impacts of voluntary sustainability standards: How has our understanding changed since the 2012 publication of 'Towards sustainability: the roles and limitations of certification?' The Meridian Institute, (September). [Online]. Available at: merid.org/content/projects/supply_chain_sustainability_ research_fund.%0ADisclaimer:

Koning, J. de. (2011). Reshaping institutions: bricolage processes in smallholder forestry in the Amazon. Reshaping institutions: bricolage processes in smallholder forestry in the Amazon, Wageningen University. [Online]. Available at: doi:10.1063/1.1476300.

Kumar, S. (2002). Does 'participation' in common pool resource management help the poor? A social cost-benefit analysis of joint forest management in Jharkhand, India. *World Development*, **30** (5), pp.763–782. [Online]. Available at: doi:10.1016/S0305-750X(02)00004-9.

Kuznetsova, A., Brockhoff, P. B. and Christensen, R. H. B. (2017). Package 'lmerTest': Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, **82** (13), pp.1–26. [Online]. Available at: doi:10.18637/jss.v082.i13.

de Lange, E., Woodhouse, E. and Milner-Gulland, E. J. (2016). Approaches Used to Evaluate the Social Impacts of Protected Areas. *Conservation Letters*, **9** (5), pp.327–333. [Online]. Available at: doi:10.1111/conl.12223.

Latham, J. E. (2013). Evaluating Failures In Tropical Forest Management: Incorporating Local Perspectives Into Global Conservation Strategies. PhD Thesis, University of York.

Latham, J. E. et al. (2017). Examining the impact of forest protection status on firewood sufficiency in rural Africa. *Environmental Conservation*, *44* (3). [Online]. Available at: doi:10.1017/S0376892917000066.

Laurence. (2012). Averting biodiversity collapse in tropical forest protected areas. *Nature*, **489** (7415), Nature Publishing Group., pp.290–294. [Online]. Available at: doi:10.1038/nature11318.

Lefcheck, J. S. (2016). PIECEWISE SEM: Piecewise structural equation modelling in R for ecology, evolution, and systematics. *Methods in Ecology and Evolution*, pp.573–579. [Online]. Available at: doi:10.1111/2041-210X.12512.

Leggett, M. and Lovell, H. (2012). Community perceptions of REDD+: A case study from Papua New Guinea. *Climate Policy*, **12** (1), pp.115–134. [Online]. Available at: doi:10.1080/14693062.2011.579317.

Lele, S. et al. (2010). Beyond exclusion: Alternative approaches to biodiversity conservation in the developing tropics. *Current Opinion in Environmental Sustainability*, **2** (1–2), Elsevier B.V., pp.94–100. [Online]. Available at: doi:10.1016/j.cosust.2010.03.006.

Leroy, J. L. et al. (2015). Measuring the Food Access Dimension of Food Security: A Critical Review and Mapping of Indicators. **36** (2), pp.167–195. [Online]. Available at: doi:10.1177/0379572115587274.

Loveridge, R. et al. (2020). Measuring human wellbeing : A protocol for selecting local indicators. *Environmental Science and Policy*, **114**, pp.461–469. [Online]. Available at: doi:10.1016/j.envsci.2020.09.002.

Lund, J. F. et al. (2017). Promising Change, Delivering Continuity: REDD+ as Conservation Fad. *World Development*, 89 (13), pp.124–139. [Online]. Available at: doi:10.1016/j.worlddev.2016.08.005.

Luttrell, C. et al. (2013). Who should benefit from REDD+? Rationales and realities. *Ecology and Society*, 18 (4). [Online]. Available at: doi:10.5751/ES-05834-180452.

Magessa, K., Wynne-jones, S. and Hockley, N. (2020). Forest Policy and Economics Does Tanzanian participatory forest management policy achieve its governance objectives? *Forest Policy and Economics*, **111** (November 2019), p.102077. [Online]. Available at: doi:10.1016/j.forpol.2019.102077.

Mahanty, S. (2018). Insights from a Cultural Landscape : Lessons from Landscape History for the Management of Rajiv Gandhi (Nagarahole) National Park. Insights from a Cultural Landscape (1), pp.23–47.

Mariki, S. B., Svarstad, H. and Benjaminsen, T. A. (2015). Elephants over the Cliff: Explaining wildlife killings in Tanzania. *Land Use Policy*, **44**, pp.19–30. [Online]. Available at: doi:10.1016/j.landusepol.2014.10.018.

Maron, M. et al. (2016). Interactions Between Biodiversity Offsets and Protected Area Commitments: Avoiding Perverse Outcomes. *Conservation Letters*, **9** (5), pp.384–389. [Online]. Available at: doi:10.1111/conl.12222.

Martin, A. et al. (2014). Geoforum Whose environmental justice? Exploring local and global perspectives in a payments for ecosystem services scheme in Rwanda. *Geoforum*, **54**, pp.167–177. [Online]. Available at: doi:10.1016/j.geoforum.2013.02.006.

Martin, A. et al. (2019). Fair ways to share benefits from community forests? How commodification is associated with reduced preference for equality and poverty alleviation. *Environmental Research Letters*, **14**, p.064002.

Martin, T. G. et al. (2011). Eliciting Expert Knowledge in Conservation Science. *Conservation Biology*, **26** (1), pp.29–38. [Online]. Available at: doi:10.1111/j.1523-1739.2011.01806.x.

Massarella, K. et al. (2018). REDD + , hype , hope and disappointment: The dynamics of expectations in conservation and development pilot projects. *World Development*, **109**, pp.375–385. [Online]. Available at: doi:10.1016/j.worlddev.2018.05.006.

Massarella, K., Sallu, S. M. and Ensor, J. E. (2020). Reproducing injustice: Why recognition matters in conservation project evaluation. *Global Environmental Change*, **65** (February), p.102181. [Online]. Available at: doi:10.1016/j.gloenvcha.2020.102181.

Max-Neef, M. et al. (1989). *Human scale development: an option for the future.* Development Dialogue, Uppsala, Sweden

McCaslin, M. L. (2012). *The SAGE Encyclopedia of Qualitative Research Methods*. In: Given, L. (Ed). The SAGE Encyclopedia of Qualitative Research Methods. Thousand Oaks . pp.672–675. [Online]. Available at: doi:10.4135/9781412963909.

Mcgillivray, M. (1991). The Human Development Index: Yet Another. *World Development*, **19** (10), pp.1461–1468. [Online]. Available at: doi:10.1016/0305-750X(91)90088-Y.

Mcgregor, A. and Sumner, A. (2010). Beyond Business as Usual: What Might 3-D Wellbeing Contribute to MDG Momentum? *IDS Bulletin*, **41** (1).

Mcgregor, J. A. (2018). *Reconciling Universal Frameworks and Local Realities in Understanding and Measuring Wellbeing in The Politics of Wellbeing*. Palgrave Macmillan. [Online]. Available at: doi:10.1007/978-3-319-58394-5.

McGregor, J. A., McKay, A. and Velazco, J. (2007). Needs and resources in the investigation of well-being in developing countries: illustrative evidence from Bangladesh and Peru. *Journal of Economic Methodology*, **14** (1), pp.107–131. [Online]. Available at: doi:10.1080/13501780601170115.

Mckinnon, M. C. et al. (2016). What are the effects of nature conservation on human well-being ? A systematic map of empirical evidence from developing countries. *Environmental Evidence*, pp.1–25. [Online]. Available at: doi:10.1186/s13750-016-0058-7.

McKinnon, M. C. et al. (2016). What are the effects of nature conservation on human well-being? A systematic map of empirical evidence from developing countries. *Environmental Evidence*, **5** (1), BioMed Central., p.8. [Online]. Available at: doi:10.1186/s13750-016-0058-7.

McNeely, J. A. (2015). A political future for protected areas. *Oryx*, **49** (2), pp.189–190. [Online]. Available at: doi:10.1017/S0030605315000150.

McShane, T. O. et al. (2011). Hard choices: Making trade-offs between biodiversity conservation and human well-being. *Biological Conservation*, **144** (3), pp.966–972. [Online]. Available at: doi:10.1016/j.biocon.2010.04.038.

MFP. (2016). NATIONAL FIVE YEAR DEVELOPMENT PLAN. The United Republic of Tanzania, (June).

Millenium Ecosystem Assessment. (2005). Ecosystems and Human Well-Being: Biodiversity Synthesis. *Ecosystems*, p.285. [Online]. Available at: doi:10.1057/9780230625600.

Miller, D. C. and Hajjar, R. (2019). Forests as pathways to prosperity : Empirical insights and conceptual advances. *World Development*, **125,** p.104647. [Online]. Available at: doi:10.1016/j.worlddev.2019.104647.

Mitchell, S. et al. (2018). The Millennium Villages Project: a retrospective, observational, endline evaluation. *The Lancet Global Health*, **6** (5), pp.500–513. [Online]. Available at: doi:10.1016/S2214-109X(18)30065-2.

Miteva, D. A., Loucks, C. J. and Pattanayak, S. K. (2015). Social and environmental impacts of forest management certification in Indonesia. *PLoS ONE*, **10** (7), pp.1–18. [Online]. Available at: doi:10.1371/journal.pone.0129675.

Muradian, R. et al. (2013). Payments for ecosystem services and the fatal attraction of win-win solutions. *Conservation Letters*, **6** (4), pp.274–279. [Online]. Available at: doi:10.1111/j.1755-263X.2012.00309.x.

Murtaugh, P. A. (2009). Performance of several variable-selection methods applied to real ecological data. *Ecology Letters*, **12** (10), pp.1061–1068. [Online]. Available at: doi:10.1111/j.1461-0248.2009.01361.x.

Naidoo, R. et al. (2019). Evaluating the impacts of protected areas on human well-being across the developing world. *Science Advances*, **5** (4), pp.1–7. [Online]. Available at: doi:10.1126/sciadv.aav3006.

Narayan, D. et al. (2000). *Voices of the poor: Can Anyone Hear Us?* Oxford University Press.

Naughton-Treves, L. (1997). Farming the Forest Edge: Vulnerable Places and People around Kibale National Park , Uganda. *Geographical Review*, **87**(1), pp.27-46.

Naughton-Treves, L., Holland, M. B. and Brandon, K. (2005). the Role of Protected Areas in Conserving Biodiversity and Sustaining Local Livelihoods. *Annual Review of Environment and Resources*, **30** (1), pp.219–252. [Online]. Available at: doi:10.1146/annurev.energy.30.050504.164507.

Nelson, F. (2011). *Community rights, conservation and contested land: the politics of natural resource governance in Africa*. In: Nelson, F. (Ed). Community rights, conservation and contested land: the politics of natural resource governance in Africa. Oxford: Earthscan. pp.3–31. [Online]. Available at: doi:10.1016/j.cosust.2016.02.003.

Newton, P. et al. (2016). Carbon, biodiversity, and livelihoods in forest commons: synergies, trade-offs, and implications for REDD+. *Environmental Research Letters*, **11** (4), IOP Publishing. [Online]. Available at: doi:10.1088/1748-9326/11/4/044017.

OECD (Organisation for Economic Cooperation and Development). 2013. Measuring Wellbeing for Development, Discussion paper 3.1. 2013 OECD GLOBAL FORUM ON DEVELOPMENT 4-5 APRIL. Available: www.oecd.org/site/oecdgfd/

Oindo, B. O. and Skidmore, A. K. (2010). Interannual variability of NDVI and species richness in Kenya. *International Journal of Remote Sensing*, **1161** (23:2), pp.285–298. [Online]. Available at: doi:10.1080/01431160010014819.

Oldekop, J. A. et al. (2016). A global assessment of the social and conservation outcomes of protected areas. *Conservation Biology*, **30** (1), pp.133–141. [Online]. Available at: doi:10.1111/cobi.12568.

Oldekop, J. A. et al. (2019). decentralized forest management in Nepal. *Nature Sustainability*, **2** pp.421–428. [Online]. Available at: doi:10.1038/s41893-019-0277-3.

Oldekop, J. A. et al. (2020). Forest-linked livelihoods in a globalized world. *Nature Plants*, **6** (12), Springer US., pp.1400–1407. [Online]. Available at: doi:10.1038/s41477-020-00814-9.

Ostrom, E. (2000). Reformulating the Commons. *Swiss Political Science Review*, **6**, pp.29–52.

Ostrom, E. (2007). A diagnostic approach for going beyond panaceas. *Proceedings of the National Academy of Sciences*, **104** (39), pp.15181–15187. [Online]. Available at: doi:10.1073/pnas.0702288104.

Oyanedel, R. (2020). Motivations for (non-) compliance with conservation rules by small-scale resource users. *Conservation Letters*, **13**(5) , pp.1–9. [Online]. Available at: doi:10.1111/conl.12725.

Pailler, S. et al. (2015). Impacts of community-based natural resource management on wealth, food security and child health in Tanzania. *PLoS ONE*, **10** (7). [Online]. Available at: doi:10.1371/journal.pone.0133252.

Parkes, D., Newell, G. and Cheal, D. (2003). Assessing the quality of native vegetation: The 'habitat hectares' approach. *Ecological Management and Restoration*, 4, pp.29–38. [Online]. Available at: doi:10.1046/j.1442-8903.4.s.4.x.

Pascual, U. et al. Biodiversity and the challenge of pluralism. *Nature Sustainability*. [Online]. Available at: doi:10.1038/s41893-021-00694-7.

Pascual, U. et al. (2014). Social Equity Matters in Payments for Ecosystem Services. *Bioscience*, **64** (11), pp.1027–1036. [Online]. Available at: doi:10.1093/biosci/biu146.

Pereira, L. (2021). Imagining better futures using the Seeds approach By Laura M. Pereira, Stockholm Resilience Centre, Stockholm University, Sweden. *Social Innovations Journal*, 5.

Persha, Lauren, M. C. (2015). Is Tanzania's joint forest management programme a triple win? Understanding causal pathways for livelihoods, governance and forest condition impacts. International Initiative for Impact Evaluation. London.

Persha, L., Agrawal, A. and Chhatre, A. (2011a). Social and Ecological Synergy. *Science*, **331** (6024), p.1606. [Online]. Available at: doi:10.1126/science.1199343.

Persha, L., Agrawal, A. and Chhatre, A. (2011b). Social and Ecological Synergy: Local Rulemaking, Forest Livelihoods, and Biodiversity Conservation. *Science*, **331** (6024), pp.1606–1608. [Online]. Available at: doi:10.1126/science.1199343.

Pettorelli, N. et al. (2005). Using the satellite-derived NDVI to assess ecological responses to environmental change. *Trends in Ecology and Evolution*, **20** (9). [Online]. Available at: doi:10.1016/j.tree.2005.05.011.

Pfaff, A. et al. (2014). Governance, Location and Avoided Deforestation from Protected Areas: Greater Restrictions Can Have Lower Impact, Due to Differences in Location. *World Development*, **55**, Elsevier Ltd., pp.7–20. [Online]. Available at: doi:10.1016/j.worlddev.2013.01.011.

Pfaff, A. and Robalino, J. (2017). Spillovers from conservation programs. *Annual Review of Resource Economics*, **9** (1), pp.299–315. [Online]. Available at: doi:10.1146/annurev-resource-100516-053543.

Pfeifer, M. et al. (2012). Protected Areas: Mixed Success in Conserving East Africa's Evergreen Forests. *Plos One*, **7** (6), pp.1–10. [Online]. Available at: doi:10.1371/journal.pone.0039337.

Pfeifer, M. et al. (2016). Mapping the structure of Borneo's tropical forests across a degradation gradient. *Remote Sensing of Environment*, 176, pp.84–97. [Online]. Available at: doi:10.1016/j.rse.2016.01.014.

Piketty, M. G. et al. (2015). Annual cash income from community forest management in the Brazilian Amazon: Challenges for the future. *Forests*, **6** (11), pp.4228–4244. [Online]. Available at: doi:10.3390/f6114228.

Porter-bolland, L. et al. (2012). Forest Ecology and Management Community managed forests and forest protected areas: An assessment of their conservation effectiveness across the tropics. *Forest Ecology and Management*, **268**, pp.6–17. [Online]. Available at: doi:10.1016/j.foreco.2011.05.034.

Poudyal, M. et al. (2017). Who bears the cost of forest conservation? *PeerJ,* **6** pp.1–35. [Online]. Available at: doi:10.7717/peerj.5106.

Pullin, A. S. et al. (2013). Human well-being impacts of terrestrial protected areas. *Environmental Evidence*, **2** (1). [Online]. Available at: doi:10.1186/2047-2382-2-19.

Rasmussen, L. V. et al. (2017). Bridging the practitioner-researcher divide: Indicators to track environmental, economic, and sociocultural sustainability of agricultural commodity production. *Global Environmental Change*, **42**, pp.33–46. [Online]. Available at: doi:10.1016/j.gloenvcha.2016.12.001.

Rasolofoson, R. A. et al. (2015). Effectiveness of Community Forest Management at reducing deforestation in Madagascar. *Biological Conservation*, **184**, pp.271–277. [Online]. Available at: doi:10.1016/j.biocon.2015.01.027.

Rasolofoson, R. A. et al. (2017). Impacts of Community Forest Management on Human Economic Well-Being across Madagascar. *Conservation Letters*, **10** (3), pp.346–353. [Online]. Available at: doi:10.1111/conl.12272.

Rasmussen, L.V., Bierbaum, R., Oldekop, J.A., Agrawal, A., 2017. Bridging the practitioner-researcher divide: Indicators to track environmental, economic, and sociocultural sustainability of agricultural commodity production. *Global Enviromental Change,* **42**, 33–46.

Redford, K. H., Padoch, C. and Sunderland, T. (2013). Fads, Funding, and Forgetting in Three Decades of Conservation. *Conservation Biology*, **27** (3), pp.437–438. [Online]. Available at: doi:10.1111/cobi.12071.

Reed, J., Ros-Tonen, M.A.F. and Sunderland, T.C.H. (2020). Operationalizing integrated landscape approaches in the tropics. CIFOR. [Online]. Available at: doi:10.17528/cifor/007800.

Reed, J. et al. (2020). Integrated landscape approaches in the tropics: A brief stock-take. *Land Use Policy*, 99 (June), p.104822. [Online]. Available at: doi:10.1016/j.landusepol.2020.104822.

Ribot, J. C., Agrawal, A. and Larson, A. M. (2006). Recentralizing While Decentralizing: How National Governments Reappropriate Forest Resources. *World Development*, **34** (11), pp.1864–1886. [Online]. Available at: doi:10.1016/j.worlddev.2005.11.020.

Robinson, E. J. Z., Albers, H. J. and Williams, J. C. (2011). Sizing reserves within a landscape: The roles of villagers' reactions and the ecological-socioeconomic setting. *Land Economics*, **87** (2), pp.233–249. [Online]. Available at: doi:10.3368/le.87.2.233.

Rode, J., Gómez-baggethun, E. and Krause, T. (2014). Motivation crowding by economic incentives in conservation policy : A review of the empirical evidence. *Ecological Economics*, **109**, Elsevier B.V. [Online]. Available at: doi:10.1016/j.ecolecon.2014.11.019.

Romero, C. et al. (2017). Evaluation of the impacts of Forest Stewardship Council (FSC) certification of natural forest management in the tropics: a rigorous approach to assessment of a complex conservation intervention. *International Forestry Review*, **19**, pp.36–49.

Romero, C. and Putz, F. E. (2018). Theory-of-Change Development for the Evaluation of Forest Stewardship Council Certification of Sustained Timber Yields from Natural Forests in Indonesia. [Online]. Available at: doi:10.3390/f9090547.

SAGCOT (2011). *Southern Agricultural Growth Corridor of Tanzania Investment Blueprint*. Available: http://www.sagcot.com/uploads/media/Invest-Blueprint-SAGCOT_High_res.pdf

Salerno, J. et al. (2020). More robust local governance suggests positive effects of long-term community conservation. *Conservation Science and Practice*, (February 2020), pp.1–13. [Online]. Available at: doi:10.1111/csp2.297.

Sandbrook, C. et al. (2013). Social Research and Biodiversity Conservation. *Conservation Biology*, **27** (6), pp.1487–1490. [Online]. Available at: doi:10.1111/cobi.12141.

Sayer, J. et al. (2013). Ten principles for a landscape approach to reconciling agriculture, conservation, and other competing land uses. *Proceedings of the national academy of sciences*, **110** (21), pp.8349-8356. [Online]. Available at: doi:10.1073/pnas.1210595110.

Sayer, J. A. et al. (2017). Measuring the effectiveness of landscape approaches to conservation and development. *Sustainability Science*, **12** (3), pp.465–476. [Online]. Available at: doi:10.1007/s11625-016-0415-z.

Sayer, J., Campbell, B. and Petheram, L. (2007). Assessing environment and development outcomes in conservation landscapes. *Biodiversity Coonservation*, **16**, pp.2677–2694. [Online]. Available at: doi:10.1007/s10531-006-9079-9.

Scheba, A. and Mustalahti, I. (2015). Rethinking 'expert' knowledge in community forest management in Tanzania. *Forest Policy and Economics*, **60**, pp.7–18. [Online]. Available at: doi:10.1016/j.forpol.2014.12.007.

Scheba, A. and Rakotonarivo, O. S. (2016). Territorialising REDD+: Conflicts over market-based forest conservation in Lindi, Tanzania. *Land Use Policy*, **57**, pp.625–637. [Online]. Available at: doi:10.1016/j.landusepol.2016.06.028.

Schleicher, J. et al. (2017a). Conservation performance of different conservation governance regimes in the Peruvian Amazon. *Scientific reports,* **7**, pp.1–10. [Online]. Available at: doi:10.1038/s41598-017-10736-w.

Schleicher, J. et al. (2017b). Poorer without It? The Neglected Role of the Natural Environment in Poverty and Wellbeing. *Sustainable Development*, **17**. [Online]. Available at: doi:10.1002/sd.1692.

Schleicher, J. et al. (2020). Statistical matching for conservation science. *Conservation Biology*, **34** (3), pp.538–549. [Online]. Available at: doi:10.1111/cobi.13448.

Schreckenberg, K. et al. (2010). Social assessment of conservation initiatives.

[Online]. Available at: doi:10.1017/S0030605310001183.

Schreckenberg, K. et al. (2016). Unpacking equity for protected area conservation. *Parks*, **22** (2), pp.11–26. [Online]. Available at: doi:10.2305/IUCN.CH.2016.PARKS-22-2KS.en.

Scoones, I. (2016). The Politics of Sustainability and Development. *Annual Review of Environment and Resources*, **41** (1), pp.293–319. [Online]. Available at: doi:10.1146/annurev-environ-110615-090039.

Sheraz, U. (2021). Used Futures as Stumbling Blocks to Sustainable Development. *Social Innovations Journal*, **5**, pp.1–8.

Sims, K. R. E. (2010). Conservation and development: Evidence from Thai protected areas. *Journal of Environmental Economics and Management*, **60** (2), pp.94–114. [Online]. Available at: doi:10.1016/j.jeem.2010.05.003.

Sims, K. R. E. and Alix-garcia, J. M. (2017). Parks versus PES : Evaluating direct and incentive-based land conservation in Mexico. *Journal of Environmental Economics and Management*, **86**, pp.8–28. [Online]. Available at: doi:10.1016/j.jeem.2016.11.010.

Stephanson, S. L. and Mascia, M. B. (2014). Putting People on the Map through an Approach That Integrates Social Data in Conservation Planning. *Conservation Biology*, **28** (5), pp.1236–1248. [Online]. Available at: doi:10.1111/cobi.12357.

Stephenson, P. J. et al. (2015). Overcoming the challenges to conservation monitoring: integrating data from in-situ reporting and global data sets to measure impact and performance. *Biodiversity*, **16** (2–3), pp.68–85. [Online]. Available at: doi:10.1080/14888386.2015.1070373.

Sterling, E. J. et al. (2017). Sustainability indicators across scales. *Nature Ecology & Evolution*, **17**, pp.99-117. [Online]. Available at: doi:10.1038/s41559-017-0349-6.

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science*, **25** (1), pp.1–21. [Online]. Available at: doi:10.1214/09-STS313.

Sunderland, T. and Campbell, B. (2007). Conservation and Development in Tropical Forest Landscapes: A Time to Face the Trade- offs? *Environmental conservation*, **34** (4), pp.276-279. [Online]. Available at: doi:10.1017/S0376892908004438.

Sunderlin, W. D. et al. (2005). Livelihoods, forests, and conservation in developing countries: An overview. *World Development*, **33** (9), pp.1383–1402. [Online]. Available at: doi:10.1016/j.worlddev.2004.10.004.

Sunderlin, W. D. (2006). Poverty alleviation through community forestry in Cambodia, Laos, and Vietnam: An assessment of the potential. *Forest Policy and Economics*, **8** (4), pp.386–396. [Online]. Available at: doi:10.1016/j.forpol.2005.08.008.

Thornton, D., Branch, L. and Murray, D. (2020). Distribution and connectivity of protected areas in the Americas facilitates transboundary conservation. *Ecological Applications*, **30** (2), pp.1–10. [Online]. Available at: doi:10.1002/eap.2027.

TNBS (Tanzania National Bureau of Statistics) 2012. Population Distribution of Pwana and Lindi Regions by District, Ward and Village/Mtaa; in 2012 Population and Housing Census. Available: https://www.nbs.go.tz/index.php/en/census-surveys/population-and-housing-census/

Tu, Y. K. and Gilthorpe, M. S. (2007). Revisiting the relation between change and initial value: A review and evaluation. *Statistics in Medicine*, **26** (2), pp.443–457. [Online]. Available at: doi:10.1002/sim.2538.Tanzania National Bureau of Statistics (2012). *2012 Census data*. Available: http://www.nbs.go.tz/

Trochim, W.M.K. 2006. Types of reliability. Research methods knowledge base. *Web Center for Social Research Methods*. http://www. socialresearchmethods. net/kb/reltypes. php.

Turnhout, E. (2014). 'Measurementality' in biodiversity governance: knowledge , transparency , and the Intergovernmental Science – Policy Platform on Biodiversity and Ecosystem Services (IPBES). *Environment and Planning A*, (January). [Online]. Available at: doi:10.1068/a4629.

UNDP (United Nations Development Programme). 2015. United Nations Sustainable Development Goals – 17 goals to transform our world. http://www.undp.org/content/undp/en/home/sustainable-development-goals/goal-3-good-health-and-wellbeing.html (accessed 10/09/2019)

UNDP (United Nations Development Programme). 2017. Human Development Report 2016: Human Development for Everyone. United Nations Development Programme

UNEP-WCMC and IUCN (2019), Protected Planet: The World Database on Protected Areas (WDPA) [Online], [2019], Cambridge, UK: UNEP-WCMC and IUCN. Available at: www.protectedplanet.net**.**

United Nations. (2020). *The UN Decade on Ecosystem Restoration Strategy*. [Online]. Available at: https://www.decadeonrestoration.org/strategy.

van der Ven, H. and Cashore, B. (2018). Forest certification: the challenge of measuring impacts. *Current Opinion in Environmental Sustainability*, **32**, Elsevier B.V., pp.104–111. [Online]. Available at: doi:10.1016/j.cosust.2018.06.001.

Venables, W. N. & Ripley, B. D. (2002). *Modern Applied Statistics with S*. Fourth Edition. Springer, New York. ISBN 0-387-95457-0.

Walker, B. et al. (2006). A Handful of Heuristics and Some Propositions for Understanding Resilience in Social-Ecological Systems. *Ecology & Society*, **11** (1).

Walpole, M., & Leader-Williams, N. (2001). Masai Mara tourism reveals partnership benefits. *Nature*, **413**(6858), 771-771.

Watson, J. E. M. et al. (2014). The performance and potential of protected areas. *Nature*, **515** (7525), pp.67–73. [Online]. Available at: doi:10.1038/nature13947.

Wells, M. and Brandon, K. (1992). *People and Parks: Linking Protected Area Management With Local Communities*. World Bank.

Wells, M. P. and McShane, T. O. (2004). Integrating protected area management with local needs and aspirations. *Ambio*, **33** (8), pp.513–519. [Online]. Available at: doi:10.1579/0044-7447-33.8.513.

West, P., Igoe, J. and Brockington, D. (2006). Parks and Peoples: The Social Impact of Protected Areas. *Annual Review of Anthropology*, **35** (1), pp.251–277. [Online]. Available at: doi:10.1146/annurev.anthro.35.081705.123308.

Whittingham, M. J. et al. (2006). Why do we still use stepwise modelling in ecology and behaviour? *Journal of Animal Ecology*, **75** (5), pp.1182–1189. [Online]. Available at: doi:10.1111/j.1365-2656.2006.01141.x.

Wiersum, K. F., Humphries, S. and van Bommel, S. (2013). Certification of community forestry enterprises: Experiences with incorporating community forestry in a global system for forest governance. *Small-scale Forestry*, **12** (1), pp.15–31. [Online]. Available at: doi:10.1007/s11842-011-9190-y.

Wilkinson, H., Hills, D. and Barbrook-johnson, P. (2021). Building a system-based Theory of Change using Participatory Systems Mapping. *Evaluation*, **27** (1), pp.80-101. [Online]. Available at: doi:10.1177/1356389020980493.

Willcock, S. et al. (2016). Land cover change and carbon emissions over 100 years in an African biodiversity hotspot. *Global Change Biology*, pp.2787–2800. [Online]. Available at: doi:10.1111/gcb.13218.

Wood, B. T. et al. (2018). Socially just triple-wins? A framework for evaluating the social justice implications of climate compatible development. *Sustainability*, **10** (1). [Online]. Available at: doi:10.3390/su10010211.

Woodhouse, E. et al. (2015). Guiding principles for evaluating the impacts of conservation interventions on human well-being. *Philisophical Transactions B*, **370** (20150103). [Online]. Available at: doi:doi.org/10.1098/rstb.2015.0103.

Woodhouse, E., De Lange, E. and Milner-Gulland, E. (2016). *Evaluating the impacts of conservation interventions on human well-being: guidance for practitioners*. IIED, International Institute for Environment and Development, London. [Online]. Available at: doi:10.1017/S0030605316001423.

Woodhouse, E. and McCabe, J. T. (2018). Well-being and conservation: Diversity and change in visions of a good life among the maasai of northern Tanzania. *Ecology and Society*, **23** (1). [Online]. Available at: doi:10.5751/ES-09986-230143.

World Bank (2018). World Bank Datazone. Available: https://data.worldbank.org/indicator/ER.LND.PTLD.ZS?locations=TZ

Wunder, S. (2007). The efficiency of payments for environmental services in tropical conservation: Essays. *Conservation Biology*, **21** (1), pp.48–58. [Online]. Available at: doi:10.1111/j.1523-1739.2006.00559.x.

Yue, A. (2012). *Validity*. Mills, Albert, J. et al. (Eds). SAGE Publications, Inc. [Online]. Available at: doi:https://dx.doi.org/10.4135/9781412957397.

Zafra-calvo, N. (2018). Exploring local people' s views on the livelihood impacts of privately versus community managed conservation strategies in the Ruvuma landscape of North Mozambique-South Tanzania. *Journal of Environmental Management* **206**, pp.853–862. [Online]. Available at: doi:10.1016/j.jenvman.2017.11.065.

Zuur, A. F., Ieno, E. N. and Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, **1** (1), pp.3–14. [Online]. Available at: doi:10.1111/j.2041-210X.2009.00001.x.