September 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'?

Authors[1]

Kristin Komives, Ashleigh Arton, Ellen Baker, Elizabeth Kennedy, Catherine Longo, Deanna Newsom, Alexander Pfaff, and Claudia Romero

Acknowledgements

This report is the product of an intense collaborative research process that brought together academic researchers and practitioners from the voluntary sustainability standards community. We are grateful to the Gordon and Betty Moore Foundation and The Meridian Institute for the opportunity to work together in this way. Many others made important contributions to this effort, including technical writing by Sofia Perez; report screening and results coding by research assistants Rita Mendez, Rebecca Banbury Morgan, Sophie Pierce, and Adam Romo; helpful advice on our approach by Jeff Milder, Samantha Cheng, Gillian Petrokofsky, Michael Mascia, and Zuzana Burivalova; feedback on plans and drafts by Robyn Paulekas and Diana Portner; and creative contributions to our debates about how best to visualize research results by Samantha Cheng, Beth Askham, Johanna Rijkenberg, Tom Allnut, Dan Rademacher, Sofia Perez, and Chris Fiorello, under the guidance of Lara Koritzke. We also acknowledge the indirect contribution of the ISEAL-led Value and Impacts Analysis (VIA) Initiative and of the researchers and corporate partners involved in that effort: our research approach builds on the evidence typology and visualization approach originally developed through the VIA Initiative.

This research is funded by the Gordon and Betty Moore Foundation through grant GBMF5710 to Meridian Institute to support the work of the Supply Chain Sustainability Research Fund. The purpose of the Research Fund is to increase the collective understanding of outcomes associated with market-based approaches to conservation, including corporate sustainability commitments, certification and standards, and finance sector strategies.

About the Gordon and Betty Moore Foundation: The Gordon and Betty Moore Foundation fosters path-breaking scientific discovery, environmental conservation, patient care improvements and preservation of the special character of the Bay Area. Visit Moore.org and follow @MooreFound.

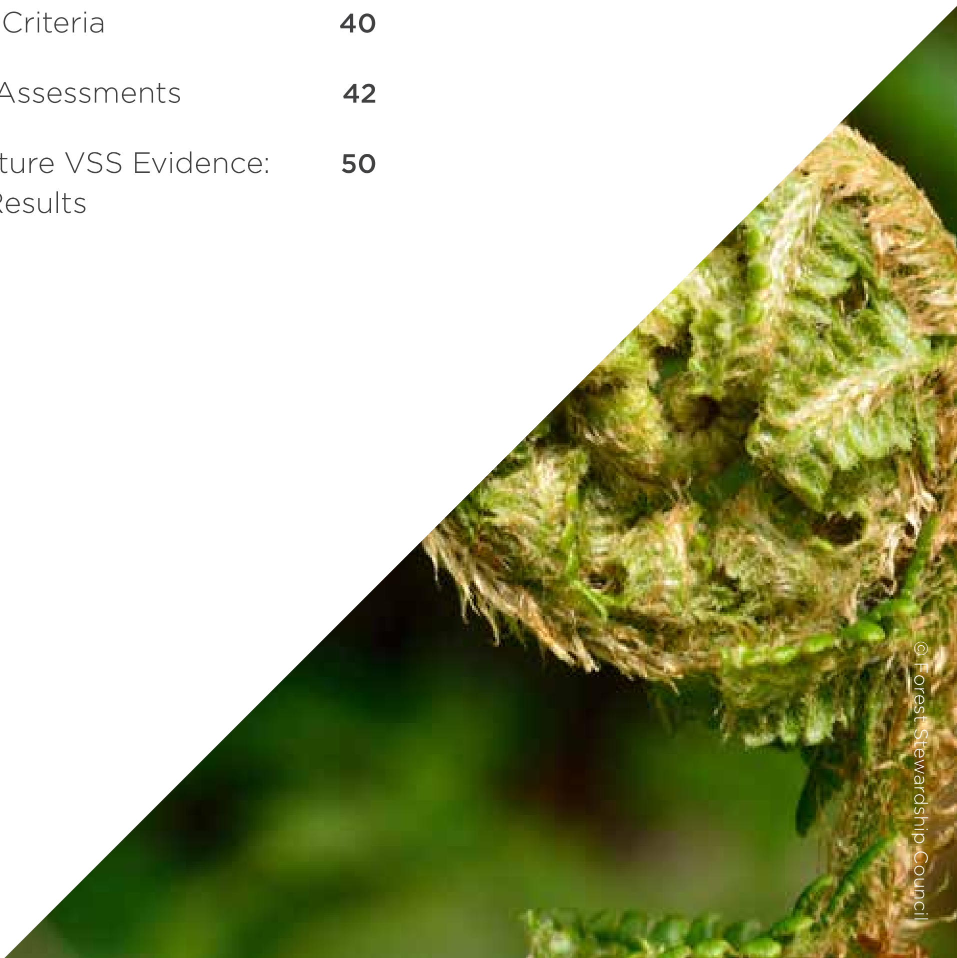
About Meridian Institute: Meridian Institute helps people solve complex and often controversial problems, make informed decisions, and implement solutions that improve lives, the economy, and the environment. Visit merid.org and follow @MeridOrg.

Suggested citation: Komives, K., Arton, A., Baker, E., Kennedy, E., Longo, C., Newsom, D., Pfaff, A., Romero, C. 2018. How has our understanding of the conservation impacts of voluntary sustainability standards changed since the 2012 publication of "Toward sustainability: The roles and limitations of certification?" Meridian Institute. Washington, DC. Available at merid.org/content/projects/supply_chain_sustainability_research_fund.

Date of publication: September 2018

1 ⁄ The authors are members of a working group brought together to identify, and then help fill, critical research gaps about voluntary sustainability standards. This was one of three working groups examining the impacts of market-based approaches to conservation as part of the work of the Supply Chain Sustainability Research Fund. This report was produced on behalf of the authors by the ISEAL Alliance.

Contents



Executive Summary

In 2012, a committee of international experts from academia, business, and civil society published *Toward Sustainability: The Roles and Limitations of Certification.* In addition to describing the history, key features and actors in voluntary standard systems (VSS), the report summarised the state of knowledge regarding VSS use and their potential to achieve conservation and other goals. It also enumerated existing evidence about VSS impacts, finding few studies and weak study designs. Since then, considerable effort has been made to fill research gaps. In this report, we review new VSS studies in the agricultural, forestry, marine fisheries and aquaculture sectors to revisit the issue of the state of knowledge about their conservation impacts and, going forward, consider how best to advance VSS impacts research.

Approach

Using a systematic search of literature published since 2011, we identified 32 studies with designs that make it possible to credit observed differences in conservation outcomes to VSS adoption. We recorded the findings of each study then used a quality screen to assess the level of confidence that could be assigned to reported results. Our final report summarizes the findings by sector and by conservation outcome, and in doing so, puts greater weight on the strongest studies with robust study designs.

Agriculture

Sixteen studies about agricultural VSS met our criteria. All examined coffee or palm oil certification in developing and emerging economies, with some studies covering multiple standards. The Rainforest Alliance/ Sustainable Agriculture Network standard (RA), is by far the most frequently studied VSS (10 studies), followed by Roundtable for Sustainable Palm Oil (RSPO – 4 studies). Smithsonian's Bird Friendly, Fairtrade International, UTZ, and C.A.F.E Practices were each addressed in one study.

Since 2011, there has been a substantial improvement in evidence about agricultural VSS adoption and two important issues of concern: deforestation, which leads to soil erosion, flooding, desertification, biodiversity loss, and climate impacts; and loss of biodiversity, which can impact the provision of ecosystem services and undermine the critical role that biodiversity plays in both human health and protection of our food supply.

Studies conclude that RA and RSPO reduced deforestation rates on certified farms and plantations in some settings, specifically Ethiopia and Colombia (RA) and primary forests in Indonesia (RSPO). They find no significant difference between certified and uncertified areas in other settings (RA in Brazil and RSPO in peatlands in Indonesia). Multiple studies find greater plant biodiversity on certified farms (RA certified farms in Ethiopia and various Latin American settings; Bird Friendly certified farms in Mexico). For mammal, insect, and bird biodiversity, half of the reported results suggest positive impacts of certification, while the other half found no significant differences.

Forestry

All 10 qualifying studies concerning the impacts of forestry VSS are about FSC certification. One also assessed Chile's Joint Solutions Project and CERTFOR, while 2 studies considered both FSC and PEFC.

Geographically, the 10 new studies cover 9 countries (with multiple studies in Gabon and Indonesia). All of these studies were set in tropical countries, except for one study on Chile and one on Sweden.

The bulk of the forestry VSS papers rely exclusively on satellite imagery and remote sensing data on forests in order to analyse the impacts of certification. Six of these papers evaluated whether certified forests had less forest cover loss than uncertified forests over the period of study, finding mixed results overall. Forest cover loss is expected in forestry activities as a function of timber harvesting practices, but is likely to be temporary, unlike deforestation from agricultural expansion. Of the four particularly strong forest-outcomes studies (relying on large sample sizes and data from both before and after certification), a study of Chile finds a meaningful decrease in measured reductions in forest cover (lower rates of forest conversion to plantations) due to FSC certification. The other studies (Mexico, Cameroon, and Peru) provide reliable evidence that (for those settings and time periods) the adoption of forestry VSS did not meaningfully reduce forest cover loss.

We emphasize that the available remote-sensing data may not always capture the outcomes one may most expect from VSS concerning timber harvesting. The studies using remote sensing data to examine reductions in tree cover have not addressed, for instance, the extent to which set-asides (e.g., high conservation value forests) are maintained or whether logging activities occurred in areas according to the specifications of management and annual operation plans. Moreover, these data do not consider other forest outcomes of interest such as changes in forest quality and associated degradation. Both can affect species habitat, carbon storage and other ecosystem services and both may be more likely than large-scale reductions in forest cover to be affected by VSS adoption.

We found only three new papers that examine forest degradation and forest-quality-related outcomes. Each covers different locations and uses a different research approach, making it impossible to see patterns or draw general conclusions. The one strong study about FSC and PEFC and forestry in Sweden found no effect from either of these VSS on forest degradation.

Marine fisheries and aquaculture

For these sectors, we identified four qualifying papers on the Marine Stewardship Council (MSC), two for the Aquaculture Stewardship Council (ASC) and none for other VSS in the area.

In marine systems, overharvesting has similar relevance as deforestation has for terrestrial systems, being one of the most concerning threats to species' and ecosystems' health. Overharvesting affects biodiversity, ecosystem structure and function and has consequences on long term sustainability of fisheries themselves as well as on other species that are captured accidentally (called 'bycatch').

Three studies on MSC focus on stock status -- whether harvested populations are maintained at healthy levels or are successfully rebuilding towards those targets. All three are high-quality studies that rely on publically available global data sets, with extensive data on fisheries in Europe, the United States, Canada, Australia and New Zealand, and less in other regions where there is little available public information. These studies show improvement over time in stock status of most MSC certified fisheries around the world, and mixed, though predominantly favourable, results in certified versus non-certified fisheries comparisons.

A further study compared the bycatch and discard rates between MSC certified and noncertified fisheries in the United States. It found no difference in marine mammal mortality, in a context in which U.S. federal legislation requires reducing marine mammal bycatch to nearly zero. Discard rates of MSC fisheries were found to be lower than those for their noncertified counterparts.

For aquaculture, major sustainability concerns include the effects of organic matter from fish farming, inputs such as antibiotics on water quality and surrounding ecosystems, and contributions to global warming. The two qualifying studies examined these issues as well as resource use of aquaculture operations, finding lower impacts from ASC-certified operations. Both examine Vietnam, so the ASC evidence base remains narrow.

The way forward

While evidence has increased substantially since 2011, it is clear that research about VSS impacts faces significant challenges, such as finding quality comparison groups (matches), establishing a picture of the pre-certification situation, and including adequate sample sizes. These challenges will not go away, yet future studies can improve by learning from the problems other researchers have faced and the solutions they have found.

To continue to advance our understanding of the conservation impacts of VSS, more and better studies will be needed to extend substantive and methodological learnings to date. We strongly encourage continuity of the systematic search process we used to find studies for this report. Our database is a public resource to which new studies can be added, so that we always have an up to date picture of what research is telling us about conservation impacts. In addition, we feel particular investments in public data would enable stronger research in the future. These include:

- Investments in multi-year public data sets on outcomes of interest, such as remote sensing data on forest degradation and large-scale efforts to measure water quality or biodiversity; and

- Investments in public access to regularly updated information on the location of certified entities, to be able to match the increasing array of georeferenced data with information on VSS.

Our report highlights the specific research gaps that future research should seek to fill. At the same time, we recognize that the type of research we focused on in this review is not the only valuable type of information concerning VSS and conservation outcomes. Other complementary research and evidence types can help to illustrate and explain VSS impacts. For example, evaluations combined with performance monitoring (tracking of specific indicators in certified entities over time) can help to us to understand how certified operations have changed and where things may have gone wrong. Other complementary approaches include:

- modelling approaches that conceptually estimate what may occur in the absence of VSS (which is potentially useful when good comparators and pre-certification information are not available);
- studies to reveal barriers or enabling conditions for VSS adoption and solid implementation; and
- studies of VSS impacts on the adoption of good practices that are expected to lead to conservation.

Putting it all in perspective

Research on the impacts of complex interventions like VSS can produce complicated and nuanced results. We need to be realistic about what to expect from research and humble about what we can claim to know. In drawing conclusions about the effectiveness of standards systems, we should take to heart findings from strong studies. At the same time, there are three important points to bear in mind:

- Conservation impacts are just one dimension of sustainability that VSS try to advance, and by looking only at conservation impacts we ignore possible trade-offs between objectives;
- Research questions put forward by research teams may not always correspond to a standard's own theory or promise of change;
- VSS and other supply-chain interventions are limited in their ability to address landscape- or seascape-level conservation outcomes such as deforestation and habitat or water quality.

1. Introduction

In 2012, a multi-stakeholder committee of international experts, drawn from academia, business, and civil society organizations, published *Toward Sustainability: The Roles and Limitations of Certification.* This independent research review provided an early assessment of the past performance of voluntary standard systems (VSS) and their future potential to achieve sustainability goals. The report brought together existing knowledge about who used VSS and why, the direct and indirect impacts of VSS, and the conditions that enhanced or impeded their effectiveness.

In one of the chapters of the *Towards Sustainability* report, the authors evaluated existing evidence about the conservation, ecological, social and economic impacts of VSS, noting differences in the literatures across sectors while also describing identified trends. They found that most of the studies were qualitative and failed to examine changes over time. The methodologies that were employed in studies available at that time made it very difficult to attribute any recorded changes to the adoption of a VSS or to extrapolate beyond the specific cases studied. As the report stated: "In the literature, unambiguous, clear attribution of on-the-ground impacts is rare." Yet it also noted there was reasonable suggestive evidence that standards had positive ecological impacts in some cases.

Why revisit the state of impacts evidence?

Reviewing evidence about impacts of standards was just one small piece of the *Towards Sustainability* report. Still, at the time of its publication, it was one of the most intensive efforts to review existing research on the impacts of sustainability standards and certification and provide conclusions to a broad audience. Since then, there has been a considerable body of new research on the sustainability impacts of VSS, a development that can be attributed to growing interest in the subject among the members of the research and donor communities, as well as by the VSS themselves.

The aim of this document is to revisit the state of evidence about conservation impacts of standards. We advance on the efforts made in the *Towards Sustainability* report by collecting and examining evidence published since 2011, then determining what this new evidence reveals about the conservation impacts of sustainability standards and certification within the agriculture, forestry, fisheries, and aquaculture sectors. Our scope is much narrower than the original report and our methods differ. First, whereas we focus just on conservation impacts (which were called 'ecological' impacts in the *Towards Sustainability* report), *Towards Sustainability* also covered social and economic impacts and broader considerations about the contribution and effectiveness of VSS. Second, our approach to literature review was more systematic than that used for the previous report, more closely following internationally accepted best practices for evidence gathering and synthesis, to the extent feasible in the available time frame.

We focus on conservation impacts because this review is part of an effort funded by the Gordon and Betty Moore Foundation and convened by Meridian Institute to build a knowledge base on how market-based strategies, such as VSS, contribute to conservation outcomes. In line with the objectives laid out by Meridian Institute, we searched for and have reviewed new evidence about the on-the-ground or in-the-water impacts of VSS, covering a range of highly relevant conservation-related issues.

Our work was guided by three overarching questions:

- What trends do we observe in the literature on conservation impacts produced since 2011 (e.g. outcome, commodity, geography, and VSS studied; methodologies used)?
- What conclusions do these studies draw about the impacts of sustainability standards on conservation at any scale?
- What are the remaining challenges and gaps in our knowledge of sustainability standards and conservation, and how can these gaps be addressed in the future?

Following a discussion of our approach and methods, the remaining sections of this report address each research question in turn.

2. Approach & Methods

This report was researched and produced in a collaborative manner by a working group whose members included academic researchers and VSS practitioners. As a group, we feel that a collaboration of this type increases the value of the resulting report, which benefits in equal measure from the rigor and external perspective provided by the academic experts, and the intimate understanding of VSS that its practitioners can supply.

Search strategy

Using the *Towards Sustainability* report as the point of departure, we drew on systematic searches of recent academic and grey literature to identify the relevant studies that had been conducted in the intervening years (2011-2018). Given the constraints under which we were operating, our search strategy built in first instance on two recent and ongoing systematic mapping[2] exercises to identify literature that looked at the conservation outcomes of VSS:

- A systematic map that was carried out by an independent research team in late 2017, as part of an ISEAL-commissioned report (Petrokofsky and Jennings, 2018): This mapping examined evidence of changes in the adoption of sustainability practices resulting from VSS, and it covered all sectors with a global scope and a starting date of 1990.

- A systematic map that was carried out in 2018 under the supervision of several members of our working group and as part of this same Moore-Meridian initiative. The mapping exercise examined terrestrial conservation outcomes of VSS in the agriculture sector, with similar but somewhat less stringent inclusion criteria than what we have used for our current effort. Our working group produced an experimental visualization of the results of this systematic mapping exercise, which is included in Annex 4.

To identify papers to include in our review, we started with the full list of studies that had qualified for full-text review in these two systematic mapping exercises (thereby benefiting from the larger search included in those efforts). To this research pool, we added grey literature and very recent academic articles that were provided by ISEAL, VSS representatives, and researchers in response to an open call for papers. We then submitted all of these papers to a new full-text review and coding process to identify empirical studies that met our specific criteria for this report based on an agreed-upon protocol.

The initial review was done by a small group of coders—graduate students and/or junior research professionals—who examined the papers; noted the conservation outcomes, study type, and other content; and determined which papers would advance to the next stage. Once the coders completed the initial mapping, members of our group were assigned to review and do a quality assessment of all papers on one or more conservation outcomes within one or more sectors, according to each member's expertise.

Study type and other inclusion criteria

To be considered in our final report, each paper had to meet our inclusion criteria: be of a qualifying study type (see Annex 1), meet the requirements for PICOs (population, intervention, comparator, and outcome – see Annex 2 for the inclusion and exclusion criteria related to each), and meet other inclusion criteria (such as a publication date of 2011 or later). Study populations were drawn from the agriculture, forestry, fisheries, and aquaculture sectors, and the intervention we examined was being or becoming certified or verified by recognized global, national, or regional VSS. We considered only VSS that are owned by non-state actors and were developed in partnership with a range of non-governmental actors (including civil society organizations and businesses). Given the limited timeframe, we excluded organic and ISO certification, as well as other interventions and populations that did not fit in the sectors included in the report. In the agricultural sector, for example, livestock was included but wild harvesting was not.

2 ⁄ Systematic maps conduct a systematic and replicable search of literature with the goal of producing a searchable database of studies on a defined topic as well as a descriptive information of this body of evidence.

We searched for papers on the three broad categories of conservation outcomes examined in the Towards Sustainability report: ecological integrity, biodiversity, and pollution and waste. We also specifically looked for papers about natural population health, ecosystem services, ecosystem function, and habitat quality.

Whereas *Towards Sustainability* examined both outcomes and specific practices and approaches that were expected to lead to desired outcomes (e.g., better water management, etc.), we focused only on research that reported on outcomes. We excluded papers on conservation-friendly practices except when the implementation of a practice provided information on the achievement of a desired outcome.

For the sake of efficiency and conceptual clarity, we limited ourselves to examining the strongest empirical evidence generated since 2011 about the conservation impacts of VSS. This meant that we included only study types that facilitate the attribution of impacts to sustainability standards, enabling us to conclude with reasonable confidence that VSS produced the results. We searched for the following study types: systematic reviews and meta analyses (which combine the results of multiple rigorous studies) and individual empirical studies with experimental or quasi-experimental research methodologies, which provide a clear comparison between certified and noncertified entities, or between the same entity before and after certification, or both (see Annex 1).

As expected, our approach generated a relatively small number of papers (though still many more than were available to the *Towards Sustainability* authors). For the marine sector, we ultimately widened the inclusion parameters because there were fewer studies that matched our guidelines than was the case for agriculture and forestry. The next section discusses additional information about the inclusion parameters for this sector.

Study quality assessment and ranking

After the working group readers conducted a full-text review of all papers that addressed their assigned outcomes within each sector, they assessed each study and rated its design and quality, considering how the evidence was collected, what questions it sought to answer, and how well it answered them—taking care to avoid conflating studies that were designed to address different questions, or had inconsistencies. In that sense, we also looked at how each study addressed common research challenges (see box below) and whether the studies' design and implementation addressed the questions they aimed to ask, remaining attentive to unexpected findings.

As a group, and once the assessment was done, we discussed and agreed on the rankings that were assigned to each study—from "strong" for those that had controlled for factors other than standards that could explain the results and that employed appropriate methods to answer the study questions, to "weak" for studies that examined a very small number of units, used questionable techniques to identify comparison groups or did not have appropriate units of analysis. Studies that applied an inappropriate method to answer the study question, had no matching, or failed to report how the matching was done were excluded from the quality assessment and from our analysis for this report. The ratings were used to inform our discussions and the resulting evidence summaries presented later in this report. (See Annex 3 for a full list of ratings and paper descriptions).

Collaborative review of the findings

In the end, 32 papers fully met our criteria for inclusion in this report: 16 papers regarding VSS in agriculture, 10 in forestry, 4 in marine fisheries, and 2 in aquaculture. Most papers were reviewed by at least two, and as many as four, readers.

The entire working group came together, virtually and in person, to conduct a collaborative review of the resulting evidence. This exercise was done by outcome, considering what each study had explicitly measured, its findings, and its quality. We also considered the level of agreement in results across studies, for the few cases where we had a handful of studies all looking at the same outcome.[3] While we were clearly working with more evidence than was available in 2011, it is still important to realize that the conclusions we were able draw in this report are based on a fairly small number of studies.

An external writer participated in some working group meetings and drafted the initial text of most of this report. The content was reviewed, edited, and approved by the entire working group.

3 / In our discussions about what conclusions could be drawn from the evidence, we considered the study quality ratings and were inspired by the approach to evidence synthesis used by the International Panel on Climate Change (IPCC), which considers how much and what type of evidence exists on a particular question as well as the level of agreement of the evidence. This framework presented an appealing approach for pulling out summary statements. The application of this approach in practice however, proved difficult because of the limited number of studies looking at the same outcome for the same sector.

CHALLENGES FACING RESEARCH ON VSS

IDENTIFYING CONTROL GROUPS: Experimental research designs (see Study Typology in Annex 1) require random assignment to treatment or control groups (that do not receive the intervention). This concept of assignment goes against the nature of a voluntary, supply chain driven intervention. Treatment and control groups shrink and get 'contaminated' when buyers, capacity builders, workers, or certified operations change their plans or strategies.

Quasi-experimental designs require that operations in the treatment group be "matched" to similar control operations, but adequate controls can be difficult to identify. In some cases, there may simply not be any or many similar operations to serve as controls. High quality matching requires rich information about the operations and the context in which they operate. More often than not, much of the information researchers would like to have to ensure good matches is simply not available. But strong comparison-group construction also requires an understanding of the motivations of companies behind the standards adoption decisions, or their voluntary selection into certification. For instance, the roles of direct incentives (e.g. subsidies to cover partial certification costs or tax reduction) and indirect incentives (e.g. hard to assess but related to reputational advantages) to adopt certification would need to be factored into the analyses.

SAMPLE SIZE: The explanatory power of experimental and quasi-experimental designs is greatly enhanced with large sample sizes. A higher sample size often implies higher financial cost of the evaluation activities. Remote sensing data and other global public information can help make large scale, less expensive data collection possible. However, these approaches are not currently applicable to all of the conservation outcomes of interest. Even in the event that costs were not an issue, there are many cases in which there simply are not that many certified operations or similar operations with which to compare. Marine fisheries has many examples of this (see box in report section 3)

ISOLATING THE EFFECTS OF VSS: For a study to isolate the contribution and impact of VSS, it is important to rule out the possibility that other factors may have contributed to the outcomes observed (e.g. economic crisis or boom, other development interventions, etc). The matching process discussed above helps address this issue. Adjustments can also be made during the analysis to address any systematic differences between treatment and control groups. VSS interventions, and the natural systems they seek to influence, are particularly challenging to study because they operate at multiple scales (e.g. water flows and pollution are not isolated from certified operations), are located within spaces that are also affected by several other programs, and as is the case for fisheries, are not easily confined to specific locations.

There can also be spill over of knowledge and capacity – for example farmers neighboring certified operations may learn from observing their neighbors, and employees who had worked in certified management operations may take knowledge gained to other places. Spill over to 'control' areas will reduce the estimate of the impact of VSS because control and treatment groups will looks similar.

© Rainforest Alliance

3. New Evidence: Research Trends from 2011 to Today

The authors of the *Towards Sustainability* report had a limited base of evidence about impacts to work with – few studies and fairly weak study designs. They concluded that much more work needed to be done to assess the ability of VSS to drive local and large-scale change.

They provided several explanations for the small number of rigorous studies available at the time. One was the methodological challenges associated with impact evaluations of VSS, challenges that are introduced in the text box in the previous chapter and discussed again later in our current report. They also noted that implementing rigorous studies takes time, effort, skills, and resources, and that many standards are relatively new. Moreover, demand for hard evidence about the impacts of standards had been quite low until relatively recently. When VSS first emerged, there was a widespread assumption that the implementation of standards would lead to positive impacts.

As the early standards—such as the Forest Stewardship Council (FSC), Marine Stewardship Council (MSC), and Fairtrade—matured, and with a substantial increase in new standards and certification programs, demand for evidence of impacts increased on the part of, among others, the companies that entered VSS programs. Likewise, donors and other organizations that have strongly committed resources to support standards are more and more motivated to understand the impacts of their investments.

In the six years since the publication of the *Towards Sustainability* report, considerable effort has been made to fill research gaps. Figures 1-3 show the number of papers identified and screened for inclusion in our review, and the number that finally met our inclusion criteria. As Figure 4 demonstrates, 2016 and 2017 were particularly prolific in the production of studies that met our inclusion criteria. 2018 is also off to a good start.[4]

In the remainder of this section of the report, we review in more detail the types and focus of new studies that have emerged. In the following section, we will examine the results of our evidence synthesis for each sector.

Figure 1. Number of forestry impact studies examined for this report, by year (total: 47). Blue dots indicate the total number of studies that were screened; orange dots indicate the subset that qualified and were included in the final analysis (10 papers).

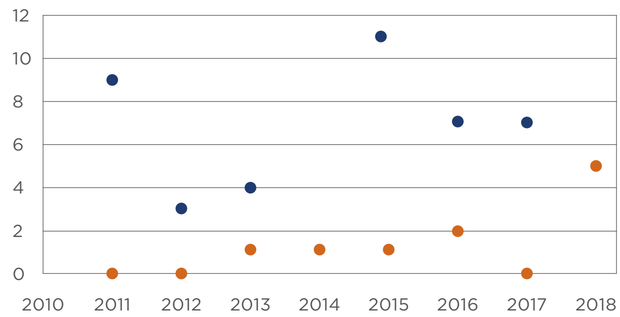


Figure 2. Number of marine/aquaculture impact studies examined for this report, by year (total: 36). Blue dots indicate the total number of studies that were screened; orange dots indicate the subset that qualified and were included in the final analysis (6 papers). .

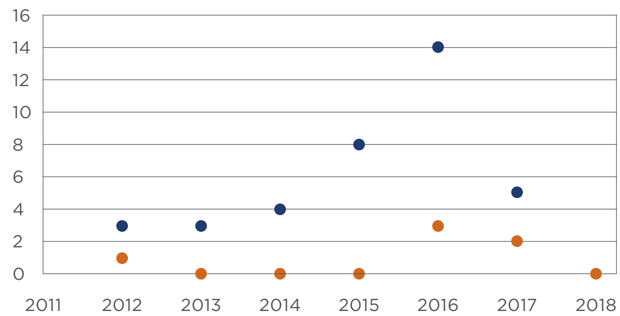
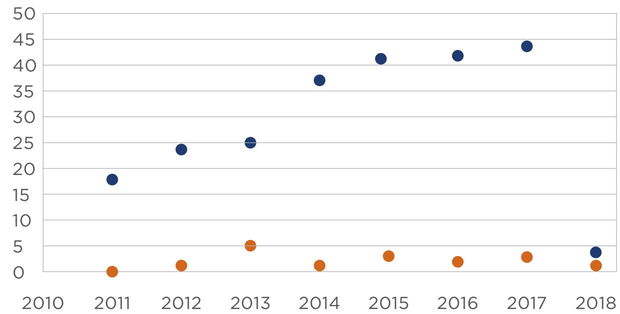
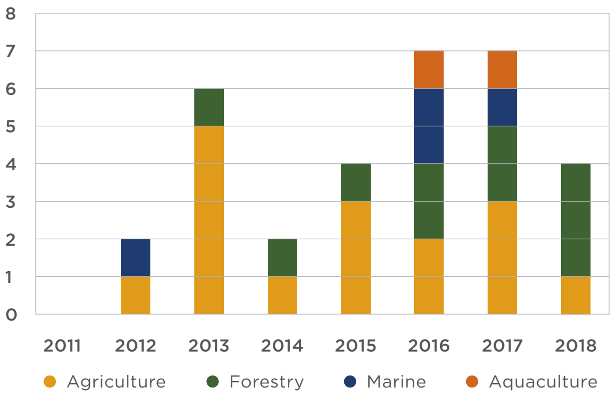


Figure 3. Number of agriculture impact studies examined for this report, by year (total: 235). Blue dots indicate the total number of studies that were screened; orange dots indicate the subset that qualified and were included in the final analysis (16 papers).



4 ⁄ 2018 was not part of the initial search (which was initiated at the end of 2017), but our call for papers did identify a number of 2018 papers. The figures presented do not represent the full research output for 2018.

Figure 4. Number of new qualifying studies on conservation impacts of VSS in the agriculture, forestry, marine, and aquaculture sectors, by year of publication - 2011-2018 (partial).



New studies about VSS in agriculture

The bulk of the literature on agricultural VSS impacts that was featured in the Towards Sustainability report was about organic standards, which was outside the scope of our current effort. There was almost no evidence available on the conservation impacts of non-organic agricultural standards; only three agriculture studies cited in that report (Komar, 2010; Melo and Wolf, 2005, 2007) looked at non-organic systems.

In contrast, sixteen new papers on non-organic agricultural standards met our criteria for inclusion in this report (See Table 1). All examined coffee and palm oil, which means that they were all localized in developing and emerging economies. The strongest studies looked at a very small number of locations (e.g. for deforestation studies, a focus on coffee certification in Colombia and Ethiopia and on palm oil certification in Indonesia and Malaysia). Rainforest Alliance (RA) was by far the most frequently studied standard. One study looked at Roundtable for Sustainable Palm Oil (RSPO), one at Smithsonian's Bird Friendly, and one at Fairtrade, UTZ, and C.A.F.E Practices.

Table 1. Geographic and commodity focus of new agriculture sector papers.

Country	Number of studies about coffee	Number of studies about palm oil
Mexico	1	0
Brazil	1	0
Colombia	5	0
El Salvador	1	0
Nicaragua	1	0
Ethiopia	3	0
Indonesia	0	3
Malaysia	0	1
Total	12	4

The qualifying papers measured a range of specific outcomes related to forest landscapes (including deforestation rates, inside and outside farm or concession areas; incidence of fire; and forest connectivity, habitat protection, and landscape heterogeneity); shade trees and agroforestry practices; preservation of riparian buffers and forest fragments; and changes in biodiversity (specifically with regard to plants, insects, birds, and mammals). Many papers addressed several topics, a few covered multiple different study locations, and many included more than one measure of our outcomes of interest. Table 2 reports on the number of papers that addressed each of the major outcomes.

Table 2. Geographic and commodity focus of new agriculture sector papers.

Outcome	Number of papers about coffee	Number of papers about palm oil
Deforestation	5	1
Incidence of fire (related to deforestation)	0	2
Forest connectivity, habitat availability, and landscape heterogeneity	2	1
Shade trees and agroforestry	2	0
Preservation of riparian buffers and forest fragments within farm areas	2	0
Biodiversity	7	0
Water quality	1	0
Total	19	4

Clearly the biggest advance since 2012 has been on literature about deforestation and literature about biodiversity. Remote sensing data has made the new work on deforestation possible. Studies on biodiversity employed a variety of methodologies (i.e., interviews/surveys, and field work/direct observation) and covered a range of taxonomic groups (four plant biodiversity, one each for mammal, insect, and bird studies since 2011). There are also new studies on other on-farm conservation outcomes, but the evidence base there is weaker, with only one or two studies on any one outcome.

We did not find papers for all of the conservation outcomes we were interested in, meaning that there is still some uncovered territory in the literature on agricultural VSS and conservation impacts. For example, no papers that qualified for our study examined conservation outcomes related to HCV

(High Conservation Value) areas or ecosystem services. Only one study looked at the effects of agricultural certification on the conservation of non-forested ecosystems (i.e., cerrado or savannah).

New studies about forestry VSS

When *Towards Sustainability* was published, some of the best sources of indirect evidence about the ecological changes promoted by forestry certification were corrective action reports, or CARs, which are issued by certifiers to their clients to highlight areas where improvements are necessary. Since 2012, various VSS in the agriculture and marine sectors have begun to do their own repeated analysis of CARs to monitor how certified entities change over time (e.g. Newsom and Milder 2018; MSC 2017). CARs studies, however, fall outside the scope of our current effort because they look only at change within certified operations, without a comparison group.

Beyond CARs studies, Towards Sustainability report included four studies that looked at forestry certification: Johansson and Lidestav, 2011 on forest biodiversity in Sweden; Hughell and Butterfield, 2008 on forest-cover outcomes in Guatemala; and Kreveld and Roerhorst, 2010 and Lagan et al., 2007, both on animal habitats. These studies would likely not meet the requirements that we applied to our current literature search: as stressed in Towards Sustainability, none of these studies provided clear attribution of impacts to VSS.

In our current review, we found 10 papers on forestry VSS published since 2011 that met our criteria for inclusion in this report. All 10 compared certified and noncertified areas, using different approaches (e.g. panel regression, propensity score matching, etc.) to limit the effect of other factors that co-vary with certification and thus could distort the results. Study quality varied based on these methods as well as data. The most common data source was remote sensing, sometimes combined with secondary data. Several papers were based upon data from field work, including ecological/ecosystem measures; surveys; formal state inspections; and combinations of qualitative and quantitative data.

All ten studies looked at FSC certification, and among these, one also assessed JSP (Joint Solutions Project[5]) and Chile's CERTFOR, while two also considered the Programme for Endorsement of Forest Certification (PEFC). Geographically, the evidence base touched on a total of nine countries (with multiple studies in Gabon and Indonesia). All but two studies addressed tropical countries, with several papers addressing more than one country.

As with the agricultural studies, many of the forestry papers examined more than one conservation outcome. The bulk of the papers (6) covered outcomes about forest-cover changes (which are expected in forestry and more likely to be temporary than forest-cover changes resulting from agricultural expansion) and about forest quality and conservation of important areas (3 papers). There is one new paper on each air quality, carbon emissions, and biodiversity. We looked for, but did not find any, qualifying papers in the forestry sector about HCV set-asides.

Table 3. Number of papers regarding forestry VSS that examined different outcomes, by region. One study examined three regions and was included in the table three times.

Outcome	Studies set in...				
	Mexico	South America	Northern Europe	Africa	South-east Asia
Forest-cover changes	1	3	0	3	2
Forest quality and conservation areas	0	0	1	1	1
Carbon emissions	0	0	0	0	1
Biodiversity	0	0	0	1	0
Air quality	0	0	0	0	1
Total	1	3	1	1	5

New studies about VSS for marine fisheries and aquaculture

Towards Sustainability included no assessment of research on aquaculture impacts, which is not surprising given that the most widely-known standard in the aquaculture space, the Aquaculture Stewardship Council (ASC) was only created in 2010. The report did discuss two papers on marine fisheries, both studies commissioned by MSC. These reports examined improvements to management practices in MSC-certified fisheries as well as indicators of fishery

5 / JSP: A voluntary commitment made by Chilean companies not to clear natural forests on their properties.

sustainability, such as stock status. It is likely that neither would meet our criteria for inclusion in the current review.

Today the number of qualifying papers remains low: we found only six new papers about VSS for marine fisheries and aquaculture that met our inclusion criteria. Four of the papers analyzed the impacts of MSC certification, while the other two focused on the ASC.

In general, these papers assess whether certified operations avoid the potential negative environmental outcomes that can stem from marine fishing (e.g. bycatch) and fish farming (e.g. pollution) (See Table 4). For marine fisheries, three out of four studies focus on 'stock status' - whether harvested populations are maintained at healthy levels (i.e., the fishing intensity and population abundance is such that they are able to reproduce and be harvested in the future, or are successfully rebuilding towards those targets).

Table 4. Number of papers regarding marine fisheries and aquaculture VSS that examined different outcomes. Three studies examined multiple outcomes and were included in the table more than once.

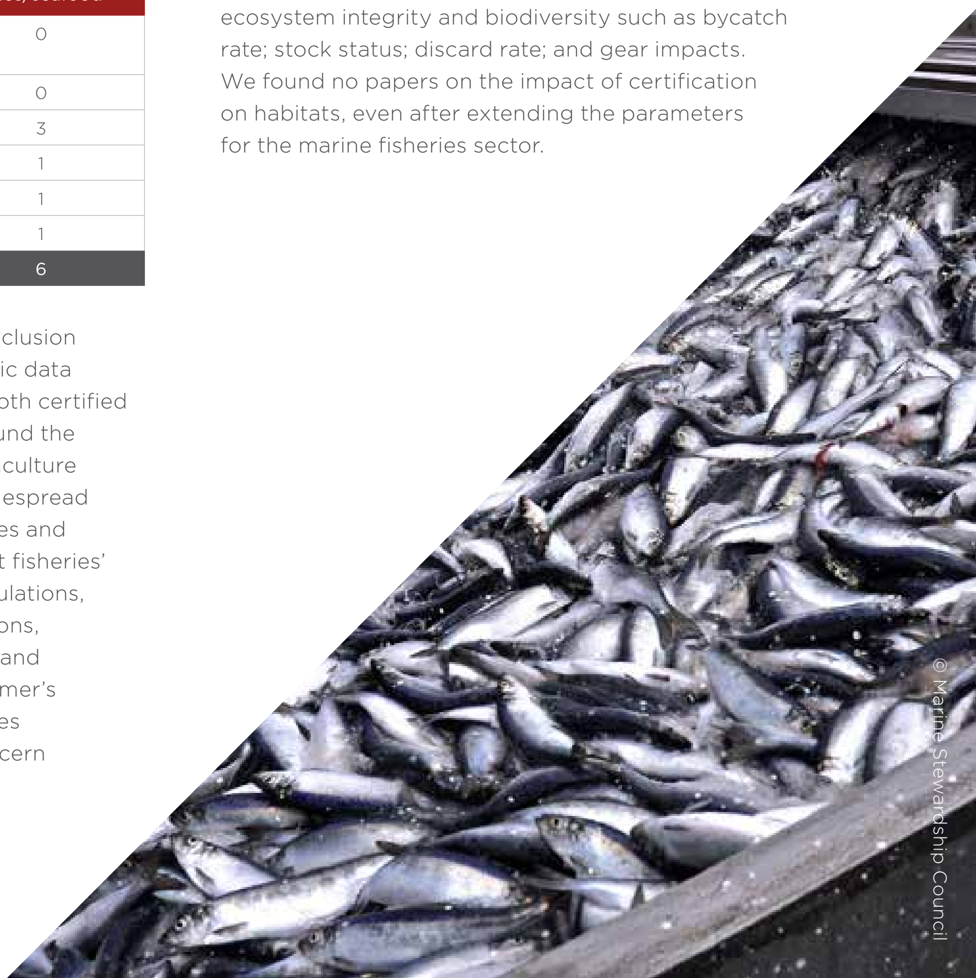
Outcome	Number of papers covering farmed seafood	Number of papers covering wild fisheries/seafood
Pollution and greenhouse gases	2	0
Resource use	2	0
Stock status	0	3
Bycatch rate	0	1
Discard rate	0	1
Gear impacts	0	1
Total	4	6

The marine fisheries papers that met our inclusion criteria make extensive use of existing public data sources to examine changes over time in both certified and noncertified wild capture fisheries around the world. By contrast, the two qualifying aquaculture studies are both based in Vietnam. The widespread availability of public data for marine fisheries and not aquaculture is not surprising, given that fisheries' managers regulate harvest of the wild populations, sometimes crossing transnational jurisdictions, while, in aquaculture, the location, amount and type of harvest on a farm are under the farmer's control, on comparably smaller spatial scales with the main causes of environmental concern being (generally) localized.

As we discuss in more detail in the final chapter of this report, all researchers who study VSS and certification in the agriculture, forestry, marine, and aquaculture sectors face thorny challenges when it comes to designing and implementing the type of research that qualified for our review. The field of wild-capture fisheries, however, poses particularly challenging barriers for researchers (See Box: Challenges in assessing impacts of wild capture fisheries).

Recognizing these challenges, we broadened our inclusion parameters for this sector, and found an additional five papers, and supplementary outcomes in one of the studies already included, that met our criteria for population, intervention, and outcome, but that did not have as strong a qualifying comparator or counterfactual. These papers -- all about the MSC -- included case studies of bycatch reduction, three papers on stock status, an in depth analysis of the impacts of certification in Australia and Mexico, and an analysis on the improvements made by Canadian certified fisheries to reduce their environmental impacts.

Taken together, the papers about marine fisheries and aquaculture covered measures of pollution (e.g., eutrophication, acidification, etc.) and impacts on ecosystem integrity and biodiversity such as bycatch rate; stock status; discard rate; and gear impacts. We found no papers on the impact of certification on habitats, even after extending the parameters for the marine fisheries sector.



CHALLENGES OF ASSESSING IMPACTS OF WILD-CAPTURE FISHERIES

By their very definition, randomized control trials require controls, as well as an adequate consideration of covariates (variables or characteristics that confound or conflate observed changes after standards adoption). For land-based commodities, a control group might take the form of neighboring noncertified farms in similar types of locations, but wild-capture fisheries operate in environments that are particularly difficult to monitor and control, for the reasons outlined below. This makes proper counterfactuals—cases that represent what would have happened if VSS had not been adopted—much harder to find, if they even exist at all.

ALWAYS ON THE MOVE: While a farm or forest has a fixed location with clearly delimited coordinates, fishing fleets and the marine populations they harvest may shift location on a yearly, seasonal, and even daily basis. Given the large area over which some marine species travel, the depths at which some vulnerable ecosystems are found, and the scale of oceanographic and climatological drivers of productivity, migration patterns, etc., it can be prohibitively costly or logistically impractical (or both) to collect information on such dynamic covariates in order to control for confounding effects.

HIGH VARIABILITY: Marked differences among fisheries can make it challenging to find appropriate comparators. In South Africa, for example, there is only one purse-seine fishery, and researchers would need to look at neighboring countries for a control that featured a similar scale of operations, target species biology, gear type, etc. Even if they could identify such a control, the comparator might operate under different contexts requiring consideration of additional covariates.

LEAKY BOUNDARIES: The steps taken by fisheries to become or remain certified can affect other fisheries that operate in the same area or target the same population. For example, a certified fishery may lobby managers to implement new sea-bottom monitoring and, in so doing, reveal vulnerable ecosystems that all fleets are subsequently asked to avoid. Although this particular improvement might be precipitated by the certified fishery, other local fisheries would follow the same practice, thereby attribution of the effects to certification is not possible through comparisons with noncertified fisheries operating in the same area.

TAILORED STUDY DESIGN: as described above, when certified fleets operate in areas that overlap those of noncertified harvesters, it can be more informative to monitor indirect changes among certified fleets (such as the implementation of measures to mitigate seabird bycatch) than to calculate direct changes (such as shifts in seabird populations). The gold standard for study design on land may not have an equivalent at sea, or may fail to detect certain impacts of improved fisheries management.

For these reasons, it is not surprising that the majority of impact studies on marine certification has focused on a single intervention with no controls, featuring descriptions of outcomes in that single case, and have used time (before and after certification) as the comparator. While data analyzed in these types of studies may be quantitative, the comparisons are often qualitative.

© Marine Stewardship Council

4. New Evidence: Results by Sector and Outcome

In this section of the report, we dive into the new papers produced since 2011 to examine what we learned about the emerging picture of VSS impacts thanks to this new research. The results are summarized by sector and grouped by outcome. More detail on individual studies, quality assessments and rankings are included in Annex 3.

4a. Research results regarding agricultural VSS

OUTCOME: Forest cover change and deforestation

Thanks in large part to technologies for remote sensing and to the adoption of more robustly-designed methodological approaches, solid evidence has begun to emerge on the forest-cover impacts of agricultural VSS. With multiple years of data on tree cover change, a number of studies were able to compare certified and noncertified areas, as well as changes in pre- versus post-certification conditions. The studies confronted problems with 'selection effects' -- important initial differences between certified and uncertified areas, which can distort conclusions if not appropriately accounted for in the study – but the authors did a fairly good job of addressing this issue in their research designs and analysis.

There is emerging evidence from robust studies that both RA coffee certification and RSPO palm oil certification reduce rates of tree-cover change in certified farms and plantations in some specific settings. There is very little evidence about whether these agricultural VSS have wider effects on deforestation, outside the boundaries of certified areas.

Among the seven studies that looked at rates of change in tree or forest cover, five focused on RA coffee certification, specifically in Colombia (Rueda et al., 2015; Rueda and Lambin, 2013), Ethiopia (Takahashi and Todo, 2013, 2014, 2017), and Brazil (Hardt et al., 2015), while the remaining study (Carlson et al., 2018) examined rates of change in forest cover and RSPO palm oil certification in Kalimantan and Sumatra, Indonesia.

DEFORESTATION, DEGRADATION AND TEMPORARY FOREST-COVER CHANGE

Of all the conservation issues facing terrestrial production systems, deforestation is arguably the one that currently receives the most attention, and understandably so. The loss of trees and other natural vegetation has wide impacts. It contributes to soil erosion, flooding, desertification and increased greenhouse gas emissions with climate change consequences, while reducing biodiversity and, often, crop productivity and quality, and water quality. When forest loss occurs across a landscape, it also disrupts the connectivity of vital habitat for flora and fauna—which particularly imperils species that are already endangered, threatened, and protected. This can be the case either for permanent reductions in forest cover, as from agricultural expansion, or temporary reductions as are common in forestry. Millions of people depend on dense forests for their subsistence and income.

Many factors fuel both permanent and temporary reductions in forest cover—including poor logging, farming, mining, road-building, and other development activities. Unsustainable and illegal logging activities contribute to forest reduction, forest loss and degradation in significant ways. While forestry activities by definition result in at least the temporary reduction of forest cover within a forestry concession, the scale and distribution of forest-cover change, the degradation of forest quality, and the rate at which forests regenerate can vary widely among managed forestry concessions.

Agriculture is considered the largest driver of deforestation and forest degradation, through the conversion of forestland to farms and pastures. In addition to outright conversion, deforestation can occur on an existing farm, when remnants of intact forests are degraded or destroyed.

Rueda and Lambin (2013), in a strong study with matched pairs, interviewed farmers on RA certified and noncertified coffee farms in Colombia and found that certified farmers planted significantly more trees outside of coffee plots than noncertified farmers. This finding was confirmed in a 2015 study that used remote sensing data to look at forest-cover change in the same area: Rueda et al., 2015, compared trends in tree cover and landscape connectivity in the 2003-2009 period. While tree cover rose in all coffee-growing areas during this period, certified farms increased their tree cover more than non-certified farms.

Takahashi and Todo published three papers about the differences between certified and noncertified coffee production in a forested coffee-growing areas of Ethiopia. In their 2013 paper, which relied on remote-sensing data that was collected in 2005 and again in 2010, the authors found lower rates of forest-cover loss in forests that had certified coffee production than in forests without coffee. By contrast, forest-cover loss in forests with non-certified coffee production were not significantly different than those without coffee. They concluded that certification significantly reduced the probability of forest-cover loss in this context. In their 2014 paper, they found that economically poor producers tended to clear forest, but that RA certification motivated producers to preserve forest. Their 2017 paper observed that certification reduced forest degradation up to 100 meters outside of the certified operation's boundary.

Hardt et al. 2015 looked at rates of forest-cover change in coffee-growing areas of Brazil, with a small sample size comparison of certified and non-certified farms and with measures taken both before and nine years after certification. They found no significant difference in rates of forest-cover change between certified and noncertified farms, though rates in certified farms were lower than in surrounding areas.

The evidence of forest-cover change and palm oil certification is limited to one strong study -- Carlson et al. 2018. This research team evaluated rates of tree cover loss in Kalimantan and Sumatra from 2001 to 2015 using annual satellite data, controlling for differences between certified and noncertified plantations. This study's overall conclusion was that RSPO reduced rates of forest-cover change in plantation areas, with two important caveats: 1) most of the impact occurred in Kalimantan, and 2) the areas that later became RSPO certified had already loss substantially more in forest cover (due to higher pre-certification rates of forest-cover reduction). Importantly, the authors found a significant and positive effect of RSPO certification on avoiding the loss of primary forest, though no significant effect on reducing peatland clearing.

OUTCOME: Incidence of fire

In the context of palm oil, deforestation associated with fire is a major concern. Three studies (Cattau et al., 2016; Noojipady et al., 2017; Carlson et al., 2018) examined the impacts of RSPO palm oil certification on fire incidence in Indonesia. The results are mixed and inconclusive. Noojipady et al. (examining change from 2002 to 2014) found lower fire incidence on certified plantations during El Niño events, while Cattau et al. (focused on 2012-2015 outcomes) found no difference in fire activity in peatland areas between certified and noncertified concessions, but a significantly lower level of fire activity in RSPO certified concessions in non-peatland areas during wet years.

Noojipady and Carlson both recognized that significant differences between certified and uncertified areas prior to certification could have affected their results. Carlson et al. chose not to report results for this reason. Noojipady et al. speculated that their finding of lower fire-related deforestation in certified areas could be because less remaining forest cover in these areas leads to smaller, harder-to-detect losses.

OUTCOME: Forest connectivity, habitat availability, and landscape heterogeneity

Studies that use remote sensing data to examine deforestation rates miss potentially important changes in forest degradation and quality. There is still little evidence that looks at this more subtle dimension of forest change, which can have important implications for species habitats and biodiversity. The three studies we reviewed on this topic use a variety of methodologies and measures to look at a range of different, though related outcomes, making it impossible to draw any overall conclusions.

Two papers are about RA certification and coffee: the Hardt et al., 2015, and Rueda and Lambin, 2013, discussed above. Both suggest that RA certified farms were already better at conservation measures than other farms before they become certified, and that they maintain or increase this difference once certified.

In the case of Colombia, Rueda and Lambin (2013) found that certified farms had larger and better connected forest areas than non-coffee regions.

Hardt et al.'s study of coffee-growing areas of Brazil, looked at habitat availability for two mammals, the giant armadillo and grey slender mouse opossum, which have different habitat requirements. Their analysis revealed that certified farms had done better job of preserving native vegetation prior to becoming certified, and maintained this difference with non-certified plantations over time.

Azhar et al. 2015 considered RSPO and landscape diversity (heterogeneity) in peninsular Malaysia, under the assumption that higher landscape heterogeneity would likely benefit biodiversity. The authors compared mean size and number of forest patches between certified plantations, noncertified large-scale plantations, and smallholder palm oil planting areas (without providing detail on any matching procedures). The authors concluded that smallholder palm oil planting areas had significantly more forest patches than either large-scale noncertified plantations or certified plantations; they found no difference between large-scale plantations and certified plantations.

OUTCOME: Biodiversity

Among those studies that met our inclusion criteria, seven addressed biodiversity outcomes—five looked at plant species (Haggar et al., 2017; Takahashi and Todo, 2017; Rueda and Lambin, 2013; Everage and Ingersoll, 2013; Caudill and Rice, 2016), one evaluated mammals (Caudill and Rice, 2016), and two assessed other types of fauna (Hughell and Newsom, 2013; Komar, 2012). All seven included comparisons of certified versus noncertified farms, with two looking at a broader range of comparisons (Komar 2012, and Caudill and Rice, 2016).

The biodiversity studies employed a variety of methodologies (i.e., interviews/surveys, and field work/direct observation) and represent a mix of quality, with four of the seven considered being particularly strong. All but one were cross-sectional and all used matching techniques to compare farms that were similar in most respects other than certification. Each study focused on different types of outcome metrics. Some used direct measures of biodiversity (e.g. species counts) while others like Rueda and Lambin and Everage and Ingersoll used indirect measures (e.g. visual assessment of tree canopy).

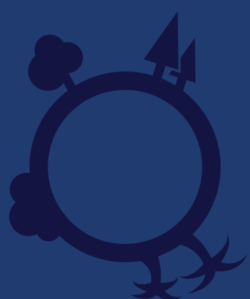
The many differences across studies makes it difficult to draw general conclusions beyond the individual study results. Two-thirds of the biodiversity metrics measured in these studies showed a positive statistically significant impact; for one-third, the authors found no statistically significant difference between certified and noncertified farms; no studies reported negative effects. Details are provided below.

Of the five studies that looked at tree diversity, the strongest based on our quality review were Haggar et al., 2017,and Takahasi and Todo, 2017, both of which find positive impacts of RA certification. Haggar et al. used propensity scores to match farms and found that RA, Fairtrade, and UTZ-certified farms in Nicaragua demonstrated higher tree diversity when compared to noncertified and C.A.F.E. Practices certified peers. Takahasi and Todo, using remote sensing technologies, found that RA forest coffee certification had a positive impact on forest quality in forested coffee growing areas of Ethiopia.

TERRESTRIAL BIODIVERSITY AND HABITATS

The preservation of terrestrial biodiversity and of habitats to support rich land-based ecosystems are conservation outcomes at the top of international agendas. Human health relies on biodiversity in myriad ways, including the provision of medicine and the protection of our food supply. A wide range of species—such as bats, bees, frogs, salamanders, spiders, and worms, to name just a few—makes agriculture possible through pollination, pest management, and soil construction.

Biodiverse systems also protect ecosystem services, including clean air, clean water, and healthy soil; as we lose biodiversity, these services become jeopardized. Finally, biodiversity is a key component of system resilience. Unlike monocultures, biodiverse agricultural systems create buffers against disease and a variety of disturbances, such as fires and floods, allowing our ecosystems to rebound or persist in an adjusted equilibrium after a shock.



Two additional studies focused on coffee farms in Mexico. Caudill and Rice, 2016, assessed biodiversity across 23 sites in Mexico comprised of forest, Bird Friendly shade-coffee farms, conventional shade-coffee farms, and sun-coffee habitats. The study controlled for elevation, and proximity to roadways, water and residential areas. Tree species richness was not significantly different, but tree density was statistically higher for Bird Friendly farms relative to conventional shade and sun coffee habitat types, but still statistically lower than forest remnants. Everage and Ingersoll, 2013, evaluated biodiversity through a visual observation of canopy diversity and a survey of certified and noncertified farmers. Certified farms in their study had significantly less degraded and denser canopies than the control group. They did not specify which VSS were assessed.

Rueda and Lambin, 2013, in their study on Colombia, found that RA-certified farmers not only planted more trees (as mentioned above) but also had greater variety of tree species on their farms than noncertified farmers.

Three new studies address fauna biodiversity. The three had mixed results for the specific taxon and metrics assessed. Caudill and Rice's study in Mexico found that some, but not all, mammal species density measures were significantly higher on Bird Friendly farms than in forests, conventional shade or sun coffee farms. Hughell and Newsom looked at arthropods (such as insects) on Colombian coffee farms. Arthropod richness was found to be significantly higher on RA certified than on noncertified farms, but there were no significant differences in arthropod abundance or diversity.

Komar examined the survival rate and site fidelity of resident and migratory forest birds across the several habitats in El Salvador (natural forest, forest fragments, RA-certified coffee farms, non-certified coffee farms with high agrochemical inputs ('technified farms'), and open pastureland. The study controlled for elevation, distance between study sites, and distance to roads. On measures of resident birds, the authors found no significant differences. For migratory birds, various measures were significantly greater on certified farms than on 'technified farms', but not always greater than on forest fragments, natural forests, and open pastureland.

OUTCOME: Shade trees and agroforestry

Retaining trees on agricultural land can have important benefits for ecosystems and biodiversity. Agroforestry (see box) is one way to do this. There is little evidence available to date on VSS and agroforestry.

The two studies that looked at agroforestry outcomes evaluated certified coffee -- RA in Colombia (Hughell and Newsom, 2013) and UTZ, Fairtrade, and C.A.F.E. Practices in Nicaragua (Haggar et al., 2017). Both studies had samples of around 35 farms of each certification and matched these to control farms that are similar on some important variables. Neither paper found a significant difference in shade cover between certified and noncertified coffee farms—a result that the authors suggest may be due to the prevalence of similar agroforestry practices in both areas that were studied. Nor did Haggar et al. observe any significant difference in tree density between certified and non-certified farms in Nicaragua.

AGROFORESTRY AND PRESERVATION OF IMPORTANT ON-FARM AREAS

Preserving important forest areas and trees on agricultural land can help retain the ecosystem value of forests. Agroforestry systems for example integrate trees, crops, and/or livestock on the same plot of land, as in the case of shade-grown coffee or cocoa farms. Agroforestry can exist both in native forests and those established by landholders, and can diversify and sustain agricultural production while increasing environmental, social, and economic benefits to communities and other land managers. Benefits of agroforestry systems include improved soil health and fertility; decreased runoff and soil erosion; improved efficiency of water and solar energy use; greater levels of biodiversity, decreased pest and disease outbreaks, and increased absorption of nitrogen, all of which lessen the need for pesticides and fertilizers.

Similarly, the preservation of streamside management zones, forest fragments on farms, or high conservation value areas are important measures that can be taken in farms or forestry concessions to protect water quality, decrease soil erosion, and promote species abundance and richness.

OUTCOME: Preservation of riparian buffers and forest fragments within farm areas

Two studies about preservation of riparian buffers met our criteria for inclusion in this report (Rueda and Lambin, 2013; and Lentijo and Hostetler, 2013). Both are about coffee farms in Colombia. Both found that certified coffee farms were significantly more likely to take action to protect streams through the establishment or preservation of buffer zones, and that these farmers made further improvements while certified. These studies' conclusions about riparian buffer protection are also supported by Hughell and Newson (2013), which found that certified coffee farms in Colombia had significantly higher streamside vegetation cover than noncertified farms. The Lentijo and Hostetler study also shows that certified farms are more likely to preserve forest fragments within their boundaries.

OUTCOME: Water quality

Hughell and Newsom (2013) is the only study we found to look at water quality outcomes, present results of water quality comparisons between RA certified and noncertified coffee farms in two regions of Colombia (Cundinamarca and Santander). In both regions the study compared various water quality measures, including measures of macroinvertebrate abundance and richness, and streambed and riparian area integrity. The authors conclude that there is significant evidence that streams flowing through certified coffee farms in Cundinamarca have higher water quality than those flowing through noncertified farms. However, for Santander, only some variables were significantly higher for certified farms. The authors highlight that some of the effects of certification may have been masked due to drought conditions during the study period in the Santander region.

4b. Research results regarding forestry VSS

OUTCOME: Forest cover change

Six of the ten papers on forestry VSS that qualified for inclusion in this report attempted to evaluate whether or not forestry certification (principally FSC) affected rates of forest-cover change. Four studies looked at the rate at which tree cover was reduced over time within forestry concessions. Three assessed changes outside the concession area as well, including one paper that looked at forest conversion to plantations.

This body of research about FSC and forest-cover change includes a number of strong studies with large sample sizes and remote sensing data from both before and after certification, and that considered a range of confounding factors in the study design and analysis. One of these strong papers finds no significant impacts of FSC on rates of forest-cover change in Mexico between 2000 and 2012 (Blackman et al. 2018). Two others (Panlasigui et al. 2018; Rico et al. 2018) estimate statistically significant but small (< 0.1%) impacts on rates of forest-cover change for a subset of certified forests included in the studies in Cameroon and Peru, and no significant effect for certified forests from other regions within the same countries. Panlasigui et al. 2018 found that FSC significantly but very slightly (0.02%) decreased rates of forest-cover reduction over the 2000-2013 period in only one out of four regions. Rico et al. 2018 found a similarly small decrease in Peru (<0.1%) in one out of three regions over the same period of time.

By contrast, Heilmayr and Lambin 2016 - also a study with a large sample size -- found a consistent gain from FSC (-13% reduction in the baseline forest conversion to plantation rate). The authors examined the impact of adoption of FSC, CERTFOR, and JSP (alone and in combination) on the rates of avoided forest substitution in company-owned land in Chile between 1986-2001 (historic rates) and 2001-2011 (period since VSS were adopted). Overall, joint adoption of the three interventions achieved higher reduction of forest conversion to plantations that any other combination or individual interventions. FSC certified companies had 33% lower historic rates of forest-cover reduction when compared to companies adopting the other two interventions, either jointly or individually, and FSC certified companies had lower forest substitution rates than any of the other two interventions. This study also tested for, and found no significant effects on company-owned areas neighboring certified areas (i.e. it found no 'leakage').

With the exception of this one study (Heilmayr), the emerging body of strong remote sensing-based studies provides increasing evidence of non-impact of FSC on measured rates of forest-cover change. A finding of non-significance in these stronger studies is more meaningful, suggesting the result is 'effectively zero'.

Less strong studies that used remote-sensing data were Rana and Sills (2018) and Miteva et al. (2015). Rana and Sills 2018 is a low-powered study examined one concession in each of three countries over 2001-2012. Controlling for covariates and using a novel method for counterfactual construction (synthetic control method), they found that differences between certified and non-certified units varied through time in one country, with a higher tree-cover reduction in Brazil, reduced tree-cover reduction in Indonesia, and no significant difference in Gabon. Miteva et al. 2015 looked at rates of forest-cover change between 2000 and 2008 at the village level in Kalimantan, Indonesia. Certification is granted at the forest-management-unit level within the village (so any village would generally be 'partially certified' or not certified at all). The study found that FSC certification decreased rates of forest-cover change at the village level by 5%. There is some noise in this study because certification history was not able to be taken into account (e.g., these units may have lost certification at the time of study, affecting also the villages that are included as controls).

OUTCOME: Forest quality and conservation of important areas

Three of the forestry papers looked at indicators related to forest quality, forest degradation, and the conservation of ecologically important areas. As the papers cover different locations and different measures, it is not possible to draw general conclusions.

The most rigorous of these studies is Villalobos et al. 2018, which examines whether either the FSC or the PECF certification avoid forest degradation in Sweden (given requirements in the Swedish Forestry Act which appear to be similar to the requirements for each certification scheme). Proxies for degradation were the post-felling preservation of environmentally important areas and establishment of additional set-asides (EIAs, established by the Swedish Forestry regulations), number of trees and of high stumps across groups of certified and comparable non-certified forest owners as reported by state agencies. The article also demonstrated that the majority of those managers in treated and control groups do not comply with environmental regulations. Authors found no effect of FSC, as well as no effect of PECF on avoiding degradation or upon any of other possible deleterious conservation effects of forest management.

Miteva et al. 2015 looked at effects of FSC certification on forest fragmentation and found increased fragmentation rates (4%) in villages with certified units in Indonesia. Medjibe et al. 2013 compared outcomes of forest management in one each certified and noncertified unit through intensive sampling in Gabon. While an attempt was made to choose similar sites, variables that may affect the outcomes of interest were different (concession area, # workers, logging intensity), which could obscure magnitude of changes found. Differences before and after logging were determined for a range of variables (gap area and # trees damaged as a function of harvested tree DBH; # trees damaged/ unit skidtrail length expressed on a per area and on a per volume basis; and changes in post-logging species compositions), but the study design was not able to statistically attribute effects to certification adoption.

OUTCOME: Air quality

The Miteva et al. study in Indonesia (2015) is the only study to examine air quality impacts. The authors find some positive impacts of certification on self-reported air pollution at the village level, but offer no explanatory mechanism.

OUTCOME: Carbon emissions

Griscom et al. 2014 assessed CO2emissions associated with several logging activities in Indonesia (felling, skidding and hauling), as a result of biomass damage and loss. They compared emissions performance in 3 certified and 6 noncertified concessions, with some attempt to control for confounding factors such as logging intensity. Due to the sample size, and inability to control for all covariates, results are more suggestive than conclusive. They found skidding related emissions to be lower in FSC certified units, but found no significant differences on other measures, and could not conclude on this difference being due to FSC certification.

OUTCOME: Biodiversity

Kalonga et al. 2016 studied the impacts of different forest management regimes -- community-based FSC certification, open access in village forests, and state forests -- on tree and seedling richness, density, and diversity in Tanzania. This study included 2 units in each certified and control group and attempted to create a counterfactual using information on potential confounding factors. They found higher tree diversity, density, and richness in FSC areas when compared to the other management regimes. They conclude they cannot fully attribute differences observed to FSC exclusively because of contrasting logging histories across sites and because they had no noncertified community-managed forest to which to compare the FSC community-managed forests.

4c. Research results regarding marine fisheries and aquaculture VSS

OUTCOME: Stock status (marine fisheries)

Stock status refers to the health of a particular stock—a fish population (or subset of a population) of interest to fisheries managers. Well-managed stocks are those whose status ensures the long-term sustainability of that particular stock—i.e., the stock is harvested at a rate that will not jeopardize the future productivity of the population or of the species that feed on said stock. It also means that the food security and livelihoods of people who rely on that stock may be safeguarded.

Three studies that met our criteria for inclusion in this report looked at stock status -- Gutierrez et al. (2012) and the MSC's 2016 and 2017 Global Impacts Reports. All papers compared MSC certified and noncertified stocks.

Using publicly available data from the International Council for the Exploration of the Sea (ICES), the MSC 2016 report analyzed the performance of Northern European stocks by examining data that captured fishing mortality and spawning stock biomass (SSB) relative to the management target in 2000 (before certification) and 2014 (after certification) (see Box: Stock status: what is it and how is it measured?). This study showed that median biomass had gone up in certified stocks since 2000, and was higher than noncertified stocks. The study also showed that certified stocks were under less intense fishing pressure.

The MSC's 2017 Global Impacts undertook a similar analysis. The report examined stocks around the world, looking at the following regions, which all had MSC certified fisheries and data for at least two fisheries (sample size in parentheses): Alaska (7 fisheries); Australia (9); western Canada (5); eastern Canada(3); Indian Ocean (4); Europe, Northern non-EU European countries (12), Europe, EU countries (37); New Zealand (18); South Africa (3). In some cases, the sample size was small, that is less than 10, due to the low number of certified fisheries in the area and/or the number of stock assessments available for the time period and reference points of interest. In others, it was considerably larger.

MARINE BIODIVERSITY AND HABITATS

Preservation of marine biodiversity and habitats is an internationally-recognized conservation goal, ensuring a range of ecosystem services, as articulated in the UN Sustainable Development Goal 14 'Conserve and sustainably use the oceans, seas and marine resources for sustainable development'. In marine systems, overharvesting has similar prominence to deforestation as one of the most concerning threats to species and ecosystem health. Direct effects of harvesting, combined with indirect effects, mediated through trophic cascades or mechanical impacts of fishing gear, can seriously impact biodiversity, ecosystem structure and function with feedback consequences on long term sustainability of fisheries themselves. Harvest impacts are evaluated through the combined outcomes on the species that are directly targeted (stock status), and those that are captured accidentally (bycatch), while mechanical impacts are captured through habitats outcomes. In the fisheries sector, there is concern both with the health of fished species and with other marine populations.

Even when different stocks share the same ecosystem, they often respond to fishing pressures and environmental drivers independently of each other, and fishing pressure on one stock might have little effect on a neighboring stock.

Fisheries can affect the integrity of habitats, especially on the sea bottom. Some types of gear can be more impactful than others—line-fishing, for example, has little to no effect on habitats, while bottom-trawling can harm seabeds and the creatures that live there.

STOCK STATUS: WHAT IS IT AND HOW IS IT MEASURED?

Stock status is generally evaluated through the combination of a stock's population abundance and the harvesting pressure it faces, often measured as the mortality rate caused by fishing. There are three commonly-used categories for a stock's population abundance, all of which are relative to a reference point for optimal harvest rates, known as the Maximum Sustainable Yield (MSY). These are over-exploited, when harvesting has reduced biomass below biologically sustainable levels; fully exploited, when the population can be harvested sustainably at MSY is; or under-exploited, when the population is more abundant than what can deliver MSY. Appreciating that these are highly dynamic systems, managers increasingly refer to MSY as a range rather than a target. The MSC definition of a healthy stock is one that 'fluctuates around' the biomass that supports MSY.

Fishing mortality is defined relative to the harvest that delivers MSY when a stock is healthy, and is considered sustainable when it is at or below this level. Other reference points are also used by different agencies that manage fisheries around the world, depending on the data available, harvest strategies, etc., but most are based on the concept of MSY. Stock assessments are performed by using data on the species' biology (e.g., longevity, fecundity, natural mortality, often by size or age class), fishing effort (e.g., days at sea, catchability), and some measure of abundance (e.g., catch per unit of effort, egg count), collected from fisheries and /or scientific surveys and inputted in stock assessment models. These models use a set of equations to estimate the stock size and exploitation rate experiences by the stock, generally incorporate several sources of uncertainty, can also include environmental forcing, and sometimes even impacts of predators and prey availability.

Often management advisory bodies will produce management advice based on an established process delivered through expert working groups, using official datasets. These are published as formal advice that management authorities then need to consider to set targets and quotas. Such formal assessments constitute the most commonly available public data on stock status.

In this report, outcomes were calculated by dividing the stock biomass (B) by a reference point (Bref) that is used by local fishery managers to determine whether or not a stock is within safe biological limits. Stock status was compared before and after certification, as well as between certified and noncertified fisheries. The objectives of the study were two-fold: 1) showing that certified stocks were managed sustainably (either with biomass above safe limits, or in the process of adjusting effort to rebuild biomass), and 2) that they had, in general, healthier stocks than noncertified fisheries and had generally improved more than noncertified stocks through time (i.e., higher median biomass and very few stocks below safe biological limits). The authors found that this was true in seven out of nine regions. In the Indian Ocean and South Africa median biomass had declined between 2000 and 2014, although it was above safe biological limits.

In the European Union, the median biomass of certified fisheries was lower than non-certified fisheries (with two stocks below biological levels that were progressively rebuilding through effort reduction). However, these European stocks have improved over time.

Though the sample sizes in some regions are quite small, these studies did rely on all publicly available data in each region at the time. Both show improvement over time in certified fisheries in most cases and mixed, though largely favourable comparisons, of certified over noncertified fisheries.

Similarly, Gutierrez et al., (2012) analyzed 45 certified and 179 noncertified stocks using data from a global stock assessment database. The authors found that the ratio (Bcurrent/BMSY) was significantly different between certified and noncertified fisheries ($p < 0.005$). Furthermore 74% of certified stocks were above sustainable target biomass levels, compared with 44% of noncertified stocks. Additionally, 82% of certified stocks had current exploitation rates that were expected to maintain the stocks at BMSY or allow for rebuilding to BMSY compared with 65% of noncertified stocks.

All of these studies ranked as high quality because they were allowing for comparisons between certified and non-certified, as well as through time, and utilizing all data available, thus compensating for the challenges of finding matched comparators in marine fisheries.

We also examined three other studies that looked at stock status using less rigorous approaches that did not fully meet our selection criteria. The first (Agnew et al., 2013) examined the stocks exploited by 45 MSC-certified fisheries and determined that none of them could be defined as overfished. This study did not meet our selection criteria because it lacked comparators. The second study (Opitz et al., 2016), which compared 31 Northern European stocks in their first year of certification with their stock status in 2015, when fisheries had been certified between 3 and 10 years. The authors found that the percentage of stocks subject to overfishing had decreased (from 52% to 44%); the percentage of stocks outside of safe biological limits had increased (from 16% to 21%); and the percentage of stocks with less than two times the biomass at the precautionary reference point, which is considered a proxy for the biomass at which MSY is achieved, had increased slightly from 64% to 67%. This study compared certified stocks through time, without an outside comparator. Importantly, this study did not use the same definition of sustainable harvest as the MSC Fishery Standard, thus it is measuring a different outcome.[6] The third study by Bellchambers et al., 2016 looked at two MSC-certified lobster fisheries, in Mexico and Australia. It found that practices adopted since certification had resulted in better environmental outcomes in both fisheries, particularly with regards to the health of their stocks.

OUTCOME: Bycatch rate and discard rate (marine fisheries)

Bycatch are marine populations that are unintentionally captured while fishing for a target species. This can include individuals of the same target species that are of a different size or sex than those sought (e.g., juveniles that are too small for consumption or regulation etc.), as well as other non-targeted species, such as sharks, sea turtles, or dolphins. "Discards," which are a subset of bycatch, refers specifically to species that are considered of low or absent market value and are mostly discarded at sea.

Bycatch contributes to the total mortality of species and can have significantly negative impacts on populations, leading either to the over-exploitation of certain stocks, or to the decline of endangered, threatened, or protected species.

6 / The authors considered any fishery harvested with fishing mortality above Fmsy to be unsustainable, while the MSC standard defines sustainable management as one that adapts fishing effort to stock abundance (i.e., higher harvests are allowed when stock biomass is shown to be very high and models suggest it can temporarily withstand higher levels of fishing, but must be monitored closely, while effort needs to be reduced and stock biomass monitored for signs of rebuilding if biomass falls below established thresholds).

To reduce their impacts on the environment, fisheries should ensure that their bycatch rates are within biologically sustainable limits, which are defined as those that avoid serious harm to a species' productivity.

Only one study that examined marine mammal bycatch (Selden et al., 2016) fully passed our quality screening. This study examined a large sample size of US fisheries—49 MSC certified and 56 noncertified—over two time periods, from 2002 to 2005 and again from 2006 to 2010. It found no difference between MSC certified and noncertified fisheries with regard to marine mammal mortality.

These results may be attributable to the fact that the Marine Mammals Protection Act requires that US fisheries reduce incidental bycatch of marine mammals to an insignificant number approaching zero; a target reference point known as the Zero Mortality Rate Goals (ZMRG). All fisheries that exceed ZMRG are required to establish a "Take Reduction Plan." Therefore, all US fisheries, certified and noncertified alike, already operate under national guidelines that mandate that manageable bycatch rates.

The unique dataset for this study came from the U.S. National Bycatch Report, the 2013 update to the National Bycatch Report, the 2012 NMFS List of Fisheries, and NOAA Marine Mammal Stock Assessment Reports.

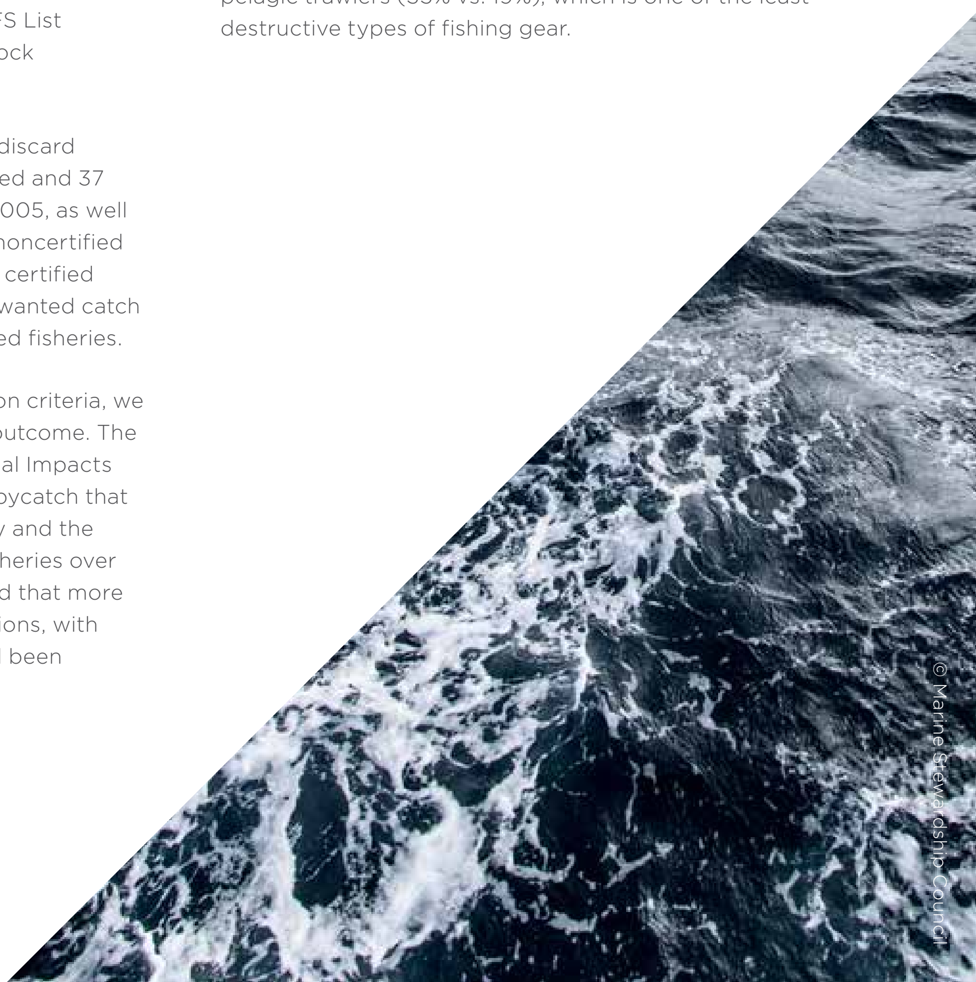
The same study also looked specifically at discard rates. Using a sample size of 28 MSC certified and 37 noncertified fisheries located in the US in 2005, as well as a second sample of 33 certified and 43 noncertified fisheries in 2010, the study found that MSC certified fisheries had significantly lower rates of unwanted catch being discarded overboard than noncertified fisheries.

Beyond this one study that met our inclusion criteria, we also reviewed several case studies on this outcome. The above-mentioned 2016 and 2017 MSC Global Impacts reports included three case studies about bycatch that documented decreases in seabird mortality and the bycatch of juvenile fish in three certified fisheries over time. A separate report by SeaChoice found that more than half of the habitat and bycatch conditions, with which Canadian fisheries must comply, had been met by MSC certified fisheries (which must comply with existing national regulations to become certified). 15% of these conditions drove practice changes and 85% resulted in assessment/research/monitoring.

OUTCOME: Gear impacts (marine fisheries)

In the fisheries sector, some types of gear can be more impactful than others. This outcome refers to the negative environmental impacts that particular types of gear may produce on marine habitats and ecosystems.

The Selden et al., 2016 paper cited above also compared the severity of gear impacts among MSC certified and noncertified fisheries. The authors found that, on average, MSC certified fisheries used gear that had lower impacts on the environment, though specific information about the kinds of impacts produced or their severity was not included in the report. The results appear to have been categorized according to the impact on seabeds, with bottom trawling considered the most destructive and pelagic trawling (midwater trawling conducted higher in the water column and above the ocean bed) the least. Among certified fisheries, a higher percentage of landings were from pelagic trawlers (35% vs. 19%), which is one of the least destructive types of fishing gear.



© Marine Stewardship Council

OUTCOME: Pollution and greenhouse gases (aquaculture)

Pollution and greenhouse gas emissions were the main focus of the two aquaculture papers included in our review. These papers also looked at resource use, such as energy use, which also shapes the sector's effect on global warming.

The first study, conducted by WWF Austria, WWF Vietnam and the ASC (2016), examining Pangasius and shrimp farms in Vietnam, found that those certified by the Aquaculture Stewardship Council (ASC) demonstrated lower usage of chemicals, drugs, and fuels than noncertified farms. However, the data were collected via interviews at a small number of farms (nine certified and five uncertified Pangasius farms, and six certified and five uncertified shrimp farms), making it difficult to draw statistically significant conclusions.

The other paper that evaluated this outcome (Nhu et al., 2016) also focused on Vietnam and involved a life-cycle analysis of ASC certified and noncertified pangasius farms (ten certified and two hundred and twelve uncertified farms). The authors analyzed marine and freshwater eutrophication, total resource use, impacts on global warming, acidification, and land resource use. For all examined categories except freshwater eutrophication (for which the analysis was inconclusive), ASC certified farms demonstrated lower environmental impacts than noncertified farms, and these results were statistically significant.

POLLUTION

In the aquaculture sector, main threats to conservation are represented by the impact of environmental inputs and alteration of local physical-chemical water column properties. Pollution is a major source of concern. One major type of pollution is eutrophication, which is the process by which an aquatic ecosystem experiences an increase in the supply of organic matter. These nutrients—nitrogen or phosphorus, most commonly—stimulate the growth of algae and other microorganisms, and can also produce toxic phytoplankton blooms that kill macroalgae, invertebrates, and fish. Increased growth in phytoplankton can also result in water column anoxia (the absence of oxygen) and other changes to ecosystems. While some of this organic matter enters aquatic ecosystems as fertilizer runoff from sources such as farms and lawns, aquaculture can also release large amounts of extra nutrients into the water, from faecal matter and excess food, which can contribute to eutrophication.

In aquaculture, the application of antibiotic drugs (and other medications) is widespread. While this practice is done to ensure the health of the farmed stock, it can also lead to increases in resistant bacteria, both in farmed and wild environments, and in fact, traces of antibiotics and antibiotic-resistant bacteria have also been found in wild-caught fish. The use of these drugs can cause increases in resistant bacteria, which may harm fish health.

Another concern is the potential contribution of aquaculture to global warming; greenhouse gas (GHG) emissions can result from aquaculture production through energy use, and contribute to climate change and its associated impacts.

5. Looking Ahead

Advances to date

Quantity

There are considerably more studies today than at the time of the *Towards Sustainability* report. The last six years have brought a substantial increase in research on impacts of sustainability standards. Forest cover changes and biodiversity are the most commonly studied outcomes within this new body of evidence.

Data and methods

There are also better studies now than existed at the time of the *Towards Sustainability* report. The questions we wanted to address for this report required us to focus on the most robust studies only, and we took the task to further differentiate within this group by study quality. New sources of data and new methods made many of the strongest of these new studies possible: publically available satellite imagery and remote sensing data facilitated forestry and some agriculture studies, and global data sets of formal assessments made by scientific advisory bodies underlie several strong marine fisheries studies.

Results

Results from high-quality quasi-experimental studies with large sample sizes are much more likely to be true effects, and not simply artefacts of a small sample size, poor data, or weak analysis. More numerous and methodologically stronger studies thus allow us to have more confidence in study results, whether these results indicate positive, negative or simply no significant impacts of VSS.

With more research now available, we were able to examine the results by outcome, rather than conflate all studies about conservation impacts to reach a single summary judgement about "the" impact of VSS. This is important for two reasons. First, given the complex and multi-faceted objectives of VSS, it is plausible that standards could have positive impacts on some outcomes of interest and no impact or a negative impact on others, particularly where important trade-offs between objectives exist. These patterns may also evolve over time. Second, different actors can have very different expectations for sustainability standards. These expectations affect which outcomes are examined in studies (e.g. deforestation rates within concessions versus outside of concession areas) and can also lead to different conclusions about determinations of "failure" or "success" of VSS, even based on the same results for each outcome. Describing results by outcome helps distinguish objective evidence concerning the changes or differences in particular outcomes from variations in perspectives on which outcomes best indicate "success".

Gaps remaining

Missing outcomes

In section 3 of this report, we compared the outcomes addressed by new studies with the range of outcomes that we had hoped to find covered within the literature. We noted some important gaps: for agricultural standards, we found no literature on conservation of HCV (High conservation value) areas or much on ecosystem services; for forestry standards, there was little on biodiversity and habitat and on HCVs; for marine standards, we were missing research on marine ecosystems and habitats; and for aquaculture, most conservation outcomes other than pollution had no qualifying research.

Beyond these coverage gaps, we also note that important dimensions of some key outcomes were missed in the studies we assessed. A good example is forest degradation, which the remotely-sensed data used in the most robust papers on deforestation do not capture well. The data also require further processing to reflect issues relevant for habitat, such as fragmentation versus intact forest landscapes. [See Box: Deforestation and Degradation: Current Challenges with Remote Sensing]. Yet these are both important outcomes of VSS adoptions. Likewise, large global data sets now being used in analyses of fisheries stock status do not reflect important details about particularly vulnerable habitats and species.

In the future, we would like to see more robust studies on agricultural VSS impacts on HCVs and ecosystem services, on forestry VSS impacts on biodiversity and forestry degradation/forest quality, and on marine fisheries VSS impacts and ecosystem and habitat health. All new studies on aquaculture VSS would be very welcome, as the evidence base for this sector is still very weak.

DEFORESTATION AND DEGRADATION: CURRENT CHALLENGES WITH REMOTE SENSING

The analyses in most of the papers on forest-cover change related to forestry certification have used the Hansen et al. (2013) global pixel data. As Hansen and colleagues themselves have emphasized, these data on tree-cover change cannot directly provide good measures for all forest outcomes of interest, such as habitat quality or degradation.

Outcomes concerning habitat, extent of fragmentation and cover by 'intact forest landscapes' (IFLs) certainly could be calculated, yet often they are not. IFLs provide a way to examine habitat when large areas are the focus of the analysis, and they can be generated from the Hansen data. Some have urged use of this data for evaluation of interventions, including standards. Yet attributing changes in such spatial metrics to VSS can currently be challenging. For instance, small concessions that comply with a commitment not to clear trees can still lose 100% of their IFL over time because of reductions in forest area that occurred in neighboring concessions.

Focusing on variation in forest degradation – say for any given level of temporary reduction in forest cover – is of interest because degradation affects species habitat, availability of hunting species, carbon storage and other ecosystem services. Yet further processing of globally available forest data would be required to measure degradation, and that effort is costly.

Analogously, for other methods, a number of papers we considered examined forest degradation or forest fragmentation through intense fieldwork, which also involves considerable and costly investments. This field information could also provide insights regarding on-the-ground aspects of VSS adoption, new knowledge that will be missed when using remote-sensing data.
The relationship between the information gained through these two complementary methodological approaches is worth exploring.

Missing standards & commodities

The growing literature on the impacts of standards is still focused on very few VSS and a small number of commodities. For example, all of the new marine fisheries papers are about the MSC, even after expanding the acceptable pool of evidence to include studies that did not meet our selection criteria. This is not altogether surprising, as MSC is the most established of the small number of international VSS in the fisheries sector; others are Friends of the Sea and Fairtrade-USA.

The few new aquaculture papers are about ASC. The deforestation-related studies examined primarily RA standards for coffee or FSC standards for forest management. We found one paper on RSPO and deforestation. For biodiversity, there are multiple papers about RA, but only one each for Bird Friendly, Fairtrade, and UTZ.

Going forward, we would like to see researchers consider examining multiple VSS in strongly-designed research studies and programmes. We also encourage the less studied VSS to commission research about their conservation impacts and to build relationships with researchers who can conduct independent assessments about their systems.

Missing geographies

VSS are applied in many different contexts around the world. With relevant differences in institutions, infrastructure, information, laws, economies across settings, there are good reasons to expect heterogeneity of impacts. Even for sectors and standards that have been studied to some extent, the number of strong papers is too small, and the geographic gaps in our evidence base too large, to help us understand with confidence how impacts vary by context. For instance, the forest cover change studies we reviewed for this report were all focused on emerging or developing economies; there are no qualifying publications on standards and forest cover loss in northern, heavily forested countries. Similarly, all new papers on agricultural standards and biodiversity are for Latin America; we have no such evidence about biodiversity impacts in other areas. Very few papers set out to explicitly examine the same outcome with comparable measures in different contexts in order to contrast results.

For the future, it would be valuable to round out our understanding of the most commonly studied standards (FSC, MSC, and RA) and outcomes (forest cover change and biodiversity) by investing in robust research in under-studied locations. Investments in new empirical research and new research synthesis activities that explicitly compare how the results of one VSS vary across different settings and times, and that examine the causes of these variations would also be extremely useful.

Missing replicates

There were not many studies by different authors looking at the same outcome and using similar measures for one single context. Only for FSC's impacts on forest cover in Indonesia did we find several distinct studies that reinforced each other's conclusions about impacts in that context, even if their units differed (e.g. certified concessions versus certified villages). In the case of biodiversity, we found multiple RA studies in Latin America but they had very different foci (plant, mammal, birds, insects) and different metrics, making it impossible to draw overarching conclusions.

Replication of studies might seem like a lot to ask for, in light of the fact that whole standards and settings are missing from the evidence base. However, for consequential decisions about resource allocations by both private and public actors, replication can raise confidence in the conclusions – especially when their results seem to be consistent with each other. That may well be seen as a worthwhile endeavor for high-cost interventions, or if natural and human (not examined in this review) impacts are significant. Such efforts could also reveal changes in impacts over time, given the dynamic and evolving nature of institutions and other factors affecting resource management.

Missing sharp conclusions about performance

Many of the studies we reviewed for this report did not find any significant difference between certified and noncertified operations, or in pre- and post-certification outcomes. As noted earlier, this can be due at least in part to small sample sizes and other study design issues, which make it harder to detect impact. With studies on some outcomes becoming more robust, we are getting a clear picture of the impacts of VSS, including when findings of “no significant impact” really mean that the adoption and use of standards had no impacts on the studied outcomes.

Yet while the quality of research designs and studies is rising, our review and several recent papers (Milder et al. 2014; Baylis et al., 2015; LeVelly and Dutilly 2016; Romero et al. 2017) highlight the challenges of research on impacts of standards. If not addressed in the research process, these problems can mask the impacts of standards or their lack thereof.

The highest quality studies we reviewed did a good job of addressing most of these issues, yet in some cases there are intrinsic problems that cannot be solved by a good research design. The following paragraphs look at these issues in more depth.

- ***It is challenging to control for all relevant factors that might affect observed changes or differences in order to accurately estimate the impacts of VSS.*** A particular challenge arises from the fact that firms' choices to certify are far from random. Differences in context and in the firms themselves can explain why some actors pursue certification yet others do not. For instance, large firms who export to rich-country consumers might be especially eager to certify, while others may be required to be certified to retain an important business-to-business client. Should the influences of such differences be strong enough – and various cases in papers suggest this could be the case – then all certified firms could be of a particular type (e.g. large exporters) while all uncertified firms could be of another (e.g. small non-exporters). When this is the case, it may be literally impossible to separate the effect of standards from other factors, as in this case, firm size and exports.

- ***It can be hard to perceive the contributions of standards when those impacts are small relative to the influences of other factors – even if small VSS impacts are meaningful.*** This problem in theory is addressed by controlling for these other factors in the research design and analysis and by increasing sample size. However, in practice, it is not always possible to find large numbers of certified operations or controls, or to have the necessary information to effectively control for all the important influences.

- ***Choosing neighboring areas or operations as controls is not always the best choice.*** A number of studies use proximity as a proxy for similarity of controls. In these cases, researchers choose areas adjacent to certified areas as controls under the assumption that areas nearby share many characteristics and are subject to similar important influences. If this assumption is true, adjacent land areas or operations could, indeed, be useful controls. But what if the benefits or disadvantages of standards and certification adoption also accrue to neighboring properties? This could be the case if having a well-protected forest next door allows species to survive or if aquaculture operations can learn from one another – even when they are not all certified. If such positive spillovers occurred, using areas near certified units as controls biases impact estimates since the gains in the certified units are being compared not to units unaffected by VSS, as should be the case, but rather to units that also enjoyed some gains from the spillover. Alternatively, negative spillovers could hide disadvantages of standards and certification.

- ***Spillovers or changes over time can also bias, downwards or upwards, the estimates of impacts from certification***. One obvious example is when the process of getting certified itself involves significant improvements that occur before a certificate is awarded. Fisheries studies have found that the more dramatic improvements occurred in the period leading up to certification (Martin et al). The same is suggested by looking at a time-series of assessments of fish population health through time in the period preceding certification (e.g, Gutierrez et al., 2012). It is also possible that the situation could deteriorate before certification: one of the deforestation papers we reviewed for this report observed more historical land clearing in areas that later became certified than in the areas that did not (Heilmayr and Lambin 2016). Without a pre-certification baseline or data from pre-certification years, these effects are lost in impact evaluations. Of the studies we reviewed, only those using remote sensing data were able to establish the situation before certification and only one addressed the possibility of anticipatory changes, that is, those occurring before certification is adopted (Blackman et al. 2018).

- ***Assessing impacts soon after certification or simultaneously studying operations that have been certified for different periods of time can be problematic.*** Certifications within a given landscape start at different moments and improvements may occur over several years post-certification – as one would expect from standards with continual improvement requirements, for example, or if there is a lag between practice adoption and changes in related outcomes. Differences in time since certification creates 'noise' in the treatment – some units have in fact had more exposure to certification than others. One problem for some of the studies we reviewed was having data only for a few post-certification years, which limited the statistical power and also detection of the change that could potentially have been realized. Some studies attempted to compare annual rates of change in outcome as a function of the years since certification yet, overall, identification by averaging the impacts across several firms certified at different points in time remains difficult.

These sorts of challenges will not go away – they are inherent to studies of VSS. Encouraging dialogue and learning between researchers about how best to address these problems will continue to be important going forward.

Missing determinants of performance

Measuring the conservation impacts of standards is only part of the story. If the ultimate goal is to design interventions to achieve conservation, then we also need to understand why an intervention works or does not work in a particular context. Filling in research gaps for specific settings helps to solve this problem, but understanding of why and how an intervention works also requires different types of information. For example, Romero et al. 2017 propose adding a 'process' evaluation to studies of FSC impact, to ask whether FSC was implemented according to the system's design specifications. Poor implementation, rather than poor requirements, could be responsible for the poor outcomes.

Some of the studies that we examined analyzed factors that contributed to observed impacts, or the lack thereof. Carefully reviewing this aspect of the authors' analysis was outside of the scope of our current effort, unless it was linked to constraints on methods. Such a review should be done in the future, however, to help ensure that the wealth of new studies can most contribute to our collective understanding about how to improve the effectiveness of conservation interventions. The information generated through this type of work could greatly contribute to detecting and addressing implementation issues and potentially result in enhanced performance of individual VSS.

The way forward

Research about the conservation impacts of sustainability standards and certification has come a long way since the *Towards Sustainability* report from 2012. Still, as we have seen, there are still many gaps in our understanding of the impacts of standards, and research to fill those gaps faces important challenges. Where do we go from here?

Towards more and better studies

For individual researchers and those funding research, we hope that this report provides guidance on what still needs to be studied and what to consider in designing a rigorous study of standards' impacts. Individual studies can be improved by learning from the problems other researchers faced and the solutions they found: working to increase sample sizes, obtain good quality matches, and capture a pre-certification picture will all be key. Finding ways to incentivize noncertified units, farms or fisheries to participate in studies would also be helpful.

In general, it is clear that significant investment in both more and better studies, and in communicating the substantive and methodological learnings from those studies, will be necessary to continue to advance our understanding of the impact of standards on conservation.

Investing to strengthen the counterfactual research base

Our review suggests that a number of key investments that could go a long way towards enabling further strong research, going forward:

Multi-year public data sets on outcomes of interest

As seen above, multi-year public data sets with satellite imagery are incredibly powerful and have opened up new possibilities for research. The studies that we examined used remote sensing to look at tree cover loss, for example, but this technology has the potential to do more. With more processing, it could be used to examine forest degradation rather than just forest loss.

In the marine sector, remote sensing can be used to track the movement of fishing fleets. Beyond remote sensing, large-scale systematized efforts at water quality or biodiversity monitoring (or compilations of existing monitoring data) hold equal promise for enabling better evaluations of certification and other conservation-oriented interventions.

Certification atlas

To link these types of large scale data sets to certification, it is important at a minimum to know which fisheries or which land areas are certified. Ideally, researchers also are able to obtain some basic information about certified areas or fisheries, such has year of and history with certification. ISEAL is currently working with its members to build capacity to collect georeferenced data to be able to map certified areas. The ultimate goal is to create a 'certification atlas' data layer that can be combined with other spatial data for analysis. This will require sharing location data across standards and with researchers, and ideally even making this information public, creating systems and expertise for storing, quality-control, handling, maintenance, and update of this information base. Connecting these data with key information about the characteristics of certified entities and with external spatial data have potential to make it a powerful analysis and visualization tool of standards' reach and ability to produce impacts.

Systematic evidence synthesis and reviews

Systematic reviews and meta-analyses use statistical techniques to combine the quantitative findings of multiple studies in order to draw further or even overarching conclusions about impacts. To date there have not been any systematic reviews or meta-analyses conducted about the conservation impacts of sustainability standards. With the growing evidence base on conservation impacts, however, this could soon be possible. To enable this type of data aggregation, it is critical that all of the studies in question carefully report methodologies, sample sizes and strategies, as well as detailed results, including confidence intervals and caveats. This is not the case with all of the studies we reviewed, so there is work to do in encouraging researchers to ensure that their work is able to contribute to such wider learning efforts. While we get to the point of abundant robust evidence on VSS impacts, there are other useful evidence synthesis methods: this report is an example of a variation on systematic mapping approach, which could well inform on how to strategically target research efforts and formalize evidence statements.

Investing in complementary ways of knowing

Given the challenges of implementing rigorous impact evaluations of complex interventions, and the many questions we would like to answer about impacts of certification, it is unlikely that we will exhaust the demand for research on this topic any time soon, even with investments in more and better counterfactual studies, and large publicly-available data sets. We limited the scope of our systematic mapping exercise to rigorous research designs aimed at reaching conclusions about the impacts of standards and certification. But we also acknowledge that these studies are not the only valuable type of study or evidence. In general, the choice of research approach and the study design should be tailored to the specific research question of interest. For example, for those interested in knowing whether certified operations in their supply chain are improving over time, either through stopping bad practices or implementing conservation-friendly practices, it could be enough to track progress within certified operations, as long as their operators do not require strong evidence that this performance change was brought about by standards. The key in exploring and using different types of evidence is in staying true to what each type of study can and cannot tell us.

While there are many promising alternative and complementary modes of relevant research, below are three that we feel could help address the specific challenges that we encountered in our review.

Modelling

Modelling offers a potential alternative to counterfactual analysis when it is just not possible to find a reasonable control for the intervention of interest. It is commonly used in marine fisheries but is not common in the other sectors we examined here. RA has recently commissioned a study of the Upper Tana watershed in Kenya, in which roughly 95% of the tea farmers are RA certified and potential members of a non-certified control group are scarce. To address this challenge, and using established ecosystem services modelling techniques and existing datasets, the researchers have attempted to estimate the effect of farm-level sustainable farming practices, such as erosion control measures, upon watershed-level delivery of selected ecosystem services that are related to water quality. More experimentation with this type of research could be useful.

Implementation process

Understanding the implementation process can provide understanding of how certification adoption happens on the ground, and what effects differences in implementation process have on outcomes (Romero et al., 2017). VSS implementation frequently involves many actors, only one of which is the standard owner. Others can include project donors, capacity building organisations, certification bodies, government actors, and, importantly, those making and implementing resource management decisions. Examining the implementation process, and degree of implementation, can reveal any factors that are acting as barriers or enabling conditions to certification adoption and solid implementation.

Practices and outcomes

One of the background papers that was prepared to feed into the *Towards Sustainability* report (Newsom et al. 2012) looked into the possibility of unbundling 'certification' into the different practices that are required to become certified. The idea behind that strategy is that if it were possible to establish with confidence, through other studies (not exclusive to certification), that adoption of a particular practice would in fact deliver a desired outcome, then we could more safely assume that certification successfully brings about that outcome through adoption of the practice. This is appealing when it is easier to measure an observed practice than the outcome of interest.

Examining practice adoption in research studies can also help understand the link between VSS and observed outcomes. Many more factors affect the achievement of outcomes, making it harder to isolate the role of VSS. As one example, comparing changes in gear deployment between certified and uncertified fishers may more readily indicate the impact of standards than trying to evaluate whether population abundance increased as a result of certification. In reality, the combination of these two approaches would render a stronger evidence on the impacts, but this triangulation on outcomes may not always be possible.

Identifying which practices are reliable indicators of outcomes would be a useful contribution to strengthen research going forward, alongside additional particular understanding about what factors, other than a specific practice, are likely to affect the outcome that the practice is intended to influence.

Combining monitoring and impact evaluation

Milder at al. 2015 recently proposed a multi-faceted approach to building the evidence base on the impacts of agricultural and forestry certification. Focused impact evaluations are just one of three components. The other two both involve performance monitoring of certified entities only: first, system-wide monitoring of basic characteristics and practice adoption of all certified entities, and then also in-depth field monitoring of important intermediate conservation outcomes. Field-based outcome monitoring is costly to do across a whole portfolio of certified operations, so sampled monitoring could be conducted on a stratified sample of sites.

These two types of performance monitoring taken together would provide a good picture of how certified operations change over time and where things may go wrong. This understanding would support the design and interpretation of impact evaluations that compare certified to noncertified entities, ideally at multiple periods of time. Setting up this kind of system would require investment and also coordination between standards systems and independent researchers. Fortunately, standards systems are investing in building their monitoring systems, while efforts like the current collaboration to prepare this report are bringing standards systems and researchers together. A particular challenge to date has been finding funding and an on-going business model for the in-depth field monitoring component of this system. Solving this problem would unlock very valuable insights.

Keeping on top of what we know

Our evidence-review-and-synthesis exercise has generated a public database of studies that consider the conservation impacts of standards. That includes a clear framework for categorizing studies, coding results, and analyzing study quality. It is also now a resource to be built upon, as new studies come out, which we see as one of the principal benefits of a systematic approach to evidence review and synthesis. We encourage continued investment in this area, with a commitment to maintain the highest quality processes.

More could also be done to make research results accessible. Presenting them visually to informed non-academic audiences, in accurate and transparent ways, clearly could help. Annex 4 presents an initial, experimental effort to present results across multiple studies in an accessible and yet still accurate way. It attempts to guide users in understanding the evidence base by differentiating these studies by the strengths of their research designs.

Within this visualization effort, we used research design as a proxy for the strength of the evidence arising from a study because it was the only measure that we possess for all of the studies we wanted to include in trying to accurately portray a diverse literature. However, it is important to realize that, even with the strongest design, a poorly implemented study could be less conclusive than a superbly implemented study with a weaker design. Thus, in the future it would be useful and informative to do a quality review of all studies in this type of visualization, just as we have done for the studies reviewed within this report (see above). We recommend further exploring how aligned quality ratings of studies can be institutionalized, producing ongoing reviews of additions to such literature, and then also how those quality ratings could inform visual presentations of results that improve learning and inform both VSS theory and practice.

Putting it all in perspective

Conducting any assessment of rigorous studies about the impacts of complex interventions like sustainability standards can be sobering. As Romero et al. (2017) note, 'a thorough evaluation of the complex intervention of tropical forest management certification by FSC can be expected to yield complicated and heavily nuanced answers'. Faced with this reality, we need to be both realistic about what we can expect from research and humble about what we can claim to know. At the same time, we cannot lose sight or ignore conclusions that are emerging from the strongest studies or from the predominance of evidence. It is important that standards systems, their users, and their critics all confront and acknowledge both promising and disappointing findings.

In drawing conclusions about the effectiveness of standards systems based on emerging results, there are three important points to bear in mind. The first is that conservation impacts are just one dimension of sustainability that multi-faceted standard systems are trying to achieve. There are inevitable trade-offs between objectives, and these trade-offs take different shapes over time. By looking only at conservation impacts, we are inevitably missing part of the bigger picture of how standards function, particularly in what pertains to socio-economic and policy impacts of standard adoption and their mutual relationships.

It is also important to recognize that there are trade-offs between objectives, and these trade-offs may be perceived and valued differently in different contexts by different people. For instance, in a marine setting in which fish are currently quite abundant while local actors are quite poor, short-run gains in local consumption may be of more concern than short-run negative impacts on fish stocks. But when fish stocks are low enough to potentially soon collapse, or when poverty is less widespread, local priorities could be very different.

A second, and related point is that the research question asked in a particular study may not correspond to the standard system's own theory, or promise, of change. An example of this is deforestation impacts outside of certified areas – while this is a very relevant policy question, the standard systems studied do not actually claim to reduce deforestation at this wider scale. Performance against a sustainability outcome that was actually promised or claimed by a VSS is very different from delivery of a broader policy goal. Still, studies that find effects that go beyond what VSS claim to deliver (such as limited deforestation effects outside certified forest boundaries in Chile, Colombia and Ethiopia) are useful because they suggest there are indirect benefits of certification about which the standards community should be aware.

Finally, while building strong evidence about the contribution of VSS to conservation impacts is a very important research objective, this may not be the most important question we should be asking for pursuing such outcomes of interest. Standards, even when they are widely successful at certifying operations that employ best-in-class sustainable practices, cannot alone address critical environmental challenges like deforestation and biodiversity loss. These issues are affected by a multitude of forces under the influence of a wide array of actors in many sectors and places.

Most standards currently work through supply chains and seek to drive change or highlight good performance in one sector. The inherent limitations of any strictly supply chain driven intervention to address landscape or seascape issues is one of the reasons that standards and many other actors are experimenting with collaborative models that seek to use the strengths of different organisations, including standards, to achieve impact at a broader scale. We would be wise to bring the knowledge gained through this review of recent evidence into these efforts – drawing on what we have learned about what works and what may not, and also about how to set up strong evaluations of these collaborative efforts.

Bibliography

- Agnew, D., Gutierrez, N. L., Stern-Pilot, A., Smith, A. D. M., Zimmerman, C., Sainsbury, K. (2013) Rebuttal to Froese and Proelss ''Evaluation and legal assessment of certified seafood'', Marine Policy, 38: 551–553
- Arnold, S., Roebuck, K. (2017) What's behind the label? Assessing the impact of MSC and ASC seafood Certification in Canada, SeaChoice, September 2017.
- Azhar, B., Saadun, N., Puan, C. L., Kamarudin, N., Aziz, N., Nurhidayu, S., & Fischer, J. (2015). Promoting landscape heterogeneity to improve the biodiversity benefits of certified palm oil production: Evidence from Peninsular Malaysia. Global Ecology and Conservation, 3, 553-561.
- Baylis et al., 2015, Honey-Rosés, J. , J., Börner, J., Corbera, E., Ezzine-de Blas, D., Ferraro, P.J., Lapeyre, R., Persoon, U.R., Pfaff, A., and Wunder, S. 2015. Mainstreaming impact evaluation in nature conservation. Conservation Letters. http://dx.doi.org/10.1111/conl.12180
- Bellchambers, L. M., Phillips, B. F., and Perez-Ramirez, M. (2016) From certification to recertification the benefits and challenges of the Marine Stewardship Council (MSC): A case study using lobsters, Fisheries Research, 182: September 2015
- Blackman, A., Goff, L., & Planter, M. R. (2018). Does eco-certification stem tropical deforestation? Forest Stewardship Council certification in Mexico. Journal of Environmental Economics and Management, 89, 306-333.
- Carlson, K.M., Heilmayr, R., Gibbs, H.K., Noojipady, P., Burns, D.N., Morton, D.C., Walker, N.F., Paoli, G.D. & Kremen, C., (2018). Effect of oil palm sustainability certification on deforestation and fire in Indonesia. Proceedings of the National Academy of Sciences, 115(1), 121-126.
- Cattau, M. E., Marlier, M. E., & DeFries, R. (2016). Effectiveness of Roundtable on Sustainable Palm Oil (RSPO) for reducing fires on oil palm concessions in Indonesia from 2012 to 2015. Environmental Research Letters, 11(10), 105007.
- Caudill, S. A., & Rice, R. A. (2016). Do Bird Friendly® Coffee Criteria Benefit Mammals? Assessment of Mammal Diversity in Chiapas, Mexico. PloS one, 11(11), e0165662.
- Everage, L., Ingersoll, C. (2013). The COSA Measuring Sustainability Report: Coffee and Cocoa in 12 Countries, COSA.
- Froese, R., Proelss, A. (2012) Evaluation and legal assessment of certified seafood, Marine Policy, 36(6):1284–1289
- Griscom, B., Ellis, P., & Putz, F. E. (2014). Carbon emissions performance of commercial logging in East Kalimantan, Indonesia. Global change biology, 20(3), 923-937.
- Gutiérrez, N., Valencia, S., Branch, T., Agnew, D., Baum, J., Bianchi, P., Cornejo-Donoso, J., Costello, C., Defeo, O., Essington, T., Hilborn, R., Hoggarth, D., Larsen, A., Ninnes, C., Sainsbury, K., Selden, R., Sistla, S., Smith, A., Stern-Pirlot, A., Teck, S., Thorson, J., and Williams, N. (2012) Eco-label conveys reliable information on fish stock health to seafood consumers, Plos One: 7(8), 1-8.
- Haggar, J., Jerez, R., Cuadra, L., Alvarado, U., & Soto, G. (2012). Environmental and economic costs and benefits from sustainable certification of coffee in Nicaragua. Food Chain, 2(1), 24-41.
- Haggar, J., Asigbaase, M., Bonilla, G., Pico, J., & Quilo, A. (2015). Tree diversity on sustainably certified and conventional coffee farms in Central America. Biodiversity and Conservation, 24(5), 1175-1194.
- Haggar, J., Soto, G., Casanoves, F., & de Melo Virginio, E. (2017). Environmental-economic benefits and trade-offs on sustainably certified coffee farms. Ecological indicators, 79, 330-337.
- Hardt, E., Borgomeo, E., dos Santos, R. F., Pinto, L. F. G., Metzger, J. P., & Sparovek, G. (2015). Does certification improve biodiversity conservation in Brazilian coffee farms?. Forest Ecology and Management, 357, 181-194.
- Heilmayr, R., & Lambin, E. F. (2016). Impacts of nonstate, market-driven governance on Chilean forests. Proceedings of the National Academy of Sciences, 113(11), 2910-2915.
- Hughell, D. & Butterfield, R. (2008). Impact of FSC Certification on Deforestation and the Incidence of Wildres in the Maya Biosphere Reserve, Rainforest Alliance.
- Hughell, D., & Newsom, D. (2013). Impacts of Rainforest Alliance certification on coffee farms in Colombia. New York: Rainforest Alliance.
- Johansson, J. and Lidestav, G. (2011). Can voluntary standards regulate forestry? -- Assessing the environmental impacts of forest certification in Sweden. Forest Policy and Economics 13(3):191-198
- Kalonga, S. K., Midtgaard, F., & Klanderud, K. (2016). Forest certification as a policy option in conserving biodiversity: An empirical study of forest management in Tanzania. Forest Ecology and Management, 361, 1-12.
- Komar, O. (2012). Are Rainforest Alliance certified coffee plantations bird-friendly. SalvaNatura Fundacion Ecologica, San Salvador. Online: http://www. rainforest-alliance. org. uk/publications/komar-bird-study.
- Lentijo, G. M., & Hostetler, M. E. (2013). Effects of a participatory bird census project on knowledge, attitudes and behaviors of coffee farmers in Colombia. Environment, development and sustainability, 15(1), 199-223.
- LeVelly. G and Dutilly, C. (2016) Evaluating Payments for Environmental Services: Methodological Challenges. PLOS One, 11(2): e0149374
- McKinnon, M. C., Cheng, S. H., Dupre, S., Edmond, J., Garside, R., Glew, L., Holland, M. Levine, E. Masuda, Y., Miller, D., Oliveira, I., Revenaz, J., Roe, D., Shammer, S., Wilkie, D., Wongbusarakum, S., and Woodhouse, E. (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), 8. https://doi.org/10.1186/s13750-016-0058-7
- Medjibe, V. P., Putz, F. E., & Romero, C. (2013). Certified and uncertified logging concessions compared in Gabon: changes in stand structure, tree species, and biomass. Environmental management, 51(3), 524-540.
- Melo, C. J., & Wolf, S. A. (2005). Empirical assessment of eco-certification: The case of Ecuadorian bananas. Organization & Environment, 18(3), 287-317.

- Milder JC, Arbuthnot M, Blackman A, Brooks SE, Giovannucci D, Gross L, Kennedy ET, Komives K, Lambin EF, Lee A, Meyer D, Newton P, Phalan B, Schroth G, Semroc B, Van Rikxoort H, & Zrust M. (2015) Conservation Biology, 29(2):309-20.
- Miteva, D. A., Loucks, C. J., & Pattanayak, S. K. (2015). Social and environmental impacts of forest management certification in Indonesia. PloS one, 10(7), e0129675.
- MSC (2016) Marine Stewardship Council: Global Impacts Report 2016. MSC, London, UK.
- MSC (2017) Marine Stewardship Council: Global Impacts Report 2017. MSC, London, UK.
- Mupepele, A.-C., Walsh, J.C., Sutherland, W.J. & Dormann, C.F., 2016. An evidence assessment tool for ecosystem services and conservation studies. Ecological Applications, 26(5), pp.1295–1301.
- Newsom, D., Kennedy, E., Miller, J., Bahn, V., and Adhikari, S. Testing a BMP-Based Approach for Assessing Gaps in Certification Impacts Research, Appendix to Toward sustainability: The roles and limitations of certification. Washington, DC: RESOLVE, Inc.
- Noojipady, P., Morton, D.C., Schroeder, W., Carlson, K.M., Huang, C., Gibbs, H.K., Burns, D., Walker, N.F. & Prince, S.D., (2017). Managing fire risk during drought: The influence of certification and El Niño on fire-driven forest conversion for oil palm in Southeast Asia. Earth System Dynamics, 8(3), 749.
- Nhu, T. T., Schaubroeck, T., Henriksson, P. J., Bosma, R., Sorgeloos, P., & Dewulf, J. (2016). Environmental impact of non-certified versus certified (ASC) intensive Pangasius aquaculture in Vietnam, a comparison based on a statistically supported LCA. Environmental pollution, 219, 156-165.
- Opitz, S., Hoffmann, J., Quaas, M., Matz-Lück, N., Binohlan, C., and Froese, R. (2016) Assessment of MSC-certified fish stocks in the Northeast Atlantic, Marine Policy 71: September 2016, 10-14
- Panlasigui, S., Rico, J., Pfaff, A., Swenson, J., & Loucks, C. (in press) Impacts of certification, uncertified concessions, hunting zones and protected areas on forest loss in Cameroon, 2000 to 2013. Biological Conservation.
- Petrokofsky, J. and Jennings, S. (2018). The Effectiveness of Standards in Driving Practice Adoption: A State of Knowledge Report. ISEAL Alliance/3 KEEL
- Rana, P., & Sills, E. O. (2018). Does certification change the trajectory of tree cover in working forests in the tropics? An application of the synthetic control method of impact evaluation. Forests, 9(3), 98.
- Rico, J., Panlasigui, S., Loucks, C. J., Swenson, J., & Pfaff, A. (2017). Logging concessions, certification and protected areas in the Peruvian Amazon: forest impacts from development rights and land-use restrictions (No. 2017.22).
- Romero, C., Sills, E.O., Guariguata, M.R., Cerutti, P.O., Lescuyer, G., Putz, F.E. (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, Vol 19 (S2)
- Rueda, X., & Lambin, E. F. (2013). Responding to globalization: impacts of certification on Colombian small-scale coffee growers. Ecology and Society, 18(3).
- Rueda, X., Thomas, N. E., & Lambin, E. F. (2015). Eco-certification and coffee cultivation enhance tree cover and forest connectivity in the Colombian coffee landscapes. Regional environmental change, 15(1), 25-33.
- Selden, R. L., Valencia, S. R., Larsen, A. E., Cornejo-Donoso, J., & Wasserman, A. A. (2016). Evaluating seafood eco-labeling as a mechanism to reduce collateral impacts of fisheries in an ecosystem-based fisheries management context. Marine Policy, 64, 102-115.
- Steering Committee of the State-of-Knowledge Assessment of Standards and Certification. (2012). Toward sustainability: The roles and limitations of certification. Washington, DC: RESOLVE, Inc.
- Takahashi, R., & Todo, Y. (2013). The impact of a shade coffee certification program on forest conservation: A case study from a wild coffee forest in Ethiopia. Journal of environmental management, 130, 48-54.
- Takahashi, R., & Todo, Y. (2014). The impact of a shade coffee certification program on forest conservation using remote sensing and household data. Environmental Impact Assessment Review, 44, 76-81.
- Takahashi, R., & Todo, Y. (2017). Coffee certification and forest quality: evidence from a wild coffee forest in Ethiopia. World Development, 92, 158-166.
- Tejeda-Cruz, C., Silva-Rivera, E., Barton, J. R., & Sutherland, W. J. (2010). Why shade coffee does not guarantee biodiversity conservation. Ecology and Society, 15(1).
- Villalobos, L., Coria, J., & Nordén, A. (2018). Has Forest Certification Reduced Forest Degradation in Sweden?. Land Economics, 94(2), 220-238.
- WWF Austria, WWF Vietnam, ASC (2017) Lessons learned from conducting a cost - benefit analysis for Aquaculture Stewardship Council certified farms in Vietnam: The business case to illustrate value of certification through case studies of ASC certified farm.

September 2018

ANNEXES

CONSERVATION IMPACTS OF VOLUNTARY SUSTAINABILITY STANDARDS

How Has Our Understanding Changed Since the 2012 Publication of 'Towards Sustainability: The Roles and Limitations of Certification'?

1. Typology of Study Types

The evidence typology used to classify research and determine which types of research would qualify for our study was developed by Jeff Milder and Deanna Newson as part of the ISEAL / Rainforest Alliance/ WWF Global Impacts Platform project. We, the authors of the current report, had an opportunity to feed into the development of this evidence typology and reviewed several versions.

The typology includes syntheses of multiple empirical works (BLOC 1), individual empirical studies (BLOC 2), empirical datasets and analysis thereof (BLOC 3), and secondary analysis, modelling and predictive works based on such empirical works (BLOC 4). Only BLOC 1 and 2 were relevant for our use in this systematic map of literature on conservation impacts.

This typology builds on a classification of evidence originally developed by a technical group of researchers as part of ISEAL's VIA project (Value in Impact Assessment). It also relates to a typology prepared by McKinnon and collaborators (2016) as part of a large initiative on systematizing conservation evidence. The table below shows the relationship between all of these frameworks.

For this report, we included papers that were in all cells in Bloc 1 and each of the 3 bolded cells in Bloc 2 below.

BLOC 1: SYNTHESIS OF MULTIPLE INDIVIDUAL STUDIES

· Uses an a priori systematic methodology, including explicit search processes and inclusion criteria, to characterize a discrete body of literature. Such methodology should be accepted by appropriate scientific and/or technical communities, and should be transparent and repeatable.

· Such studies may enable generalization at a broader level (i.e., across multiple commodities, geographies, or standards systems) than for individual studies, depending on the depth, breadth, and findings of the studies included in the synthesis.

Evidence type	Elaboration	Source	Notes
1a- Systematic review	A study that collates, critically appraises, and synthesizes all available studies relevant to a question. Reviewers use pre-defined methods to minimize bias.	Peer-reviewed literature, grey literature, or other scientific sources (inclusion criteria may be further specified)	VIA typology category #1
1b- Meta-analysis	A study that synthesizes findings of individual papers, reports, etc., to draw conclusions that hold for broader geographic regions or time periods, or test emerging hypotheses based on the syntheses. Quantitative focus.		VIA typology category #2

Note: "systematic map" is not included as a category because it is not a synthesis of information that seeks to generate new conclusions or insights; rather it is a way of organizing and communicating existing evidence in a systematic way.

BLOC 2: DISCRETE INDIVIDUAL STUDIES

Empirical studies that interrogate cause & effect of intervention(s) (i.e., studies have explicitly defined independent and dependent variable[s])

· Study conducts primary or secondary empirical data collection

· Generalizability of results may be possible to a degree, contingent on the study design, methodology, sampling frame, statistical approach and results interpretation

	Randomized Control Trial	Randomized Control Trial		
		Covariates considered through matching (quasi-experimental)	Covariates not considered	Studies without a control group
Time series data, collected before and after an intervention*	2 - Outcomes compared in treatment and control group, before and after an intervention, with random assignment to each group. (VIA #3, McK #1)	Peer-reviewed literature, g 3 - Outcomes compared in treatment and control groups, before and after an intervention with covariates considered through matching. (VIA #4 and #5, McK #1)	4 - Outcomes compared in treatment and control groups, before and after an intervention with no consideration of covariates (VIA #6, McK #1).	5 - Outcomes examined in a treatment group only, before or after an intervention. (VIA #7, McK #2)
Cross-sectional data, collected post-intervention**	N/A - Not possible to conduct a RCT post-intervention	6 - Outcomes compared in treatment and control groups, at one point in time post-intervention, with covariates considered through matching. (VIA #4 and #5, McK #6)	7 - Outcomes compared in treatment and control groups, at one point in time post-intervention, with no consideration of covariates. (VIA #6, McK #6)	8 - Outcomes examined in a treatment group only, at one point in time post-intervention. Descriptive profile with no generalizability. (VIA #7, McK #4)

*The intervention could be the adoption of certification, preparation for certification, or post-certification practice changes. **This category only includes data collected by independent researchers. Post-intervention monitoring data collected by VSS fall in Bloc 3.

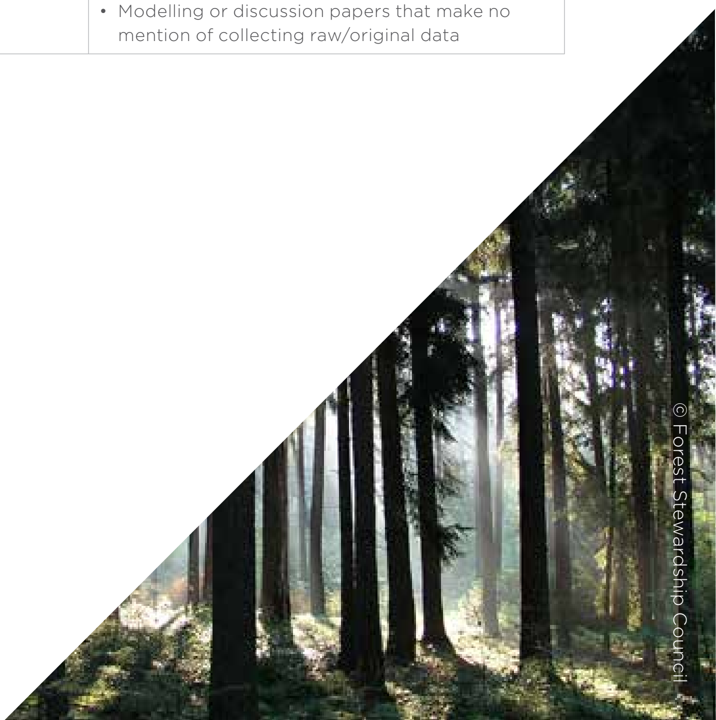
2. Inclusion and Exclusion Criteria: PICOs (Population, Intervention, Type of Control, and Outcomes)

POPULATION	Include...	Exclude...
For terrestrial systems: Area of land subject to or affected by standards or certification (e.g., certified farm, certified group, or portion of a landscape with a high concentration of certified operations). For marine systems: Fishery or species	Studies that are about VSS impacts on conservation outcomes related to the relevant populations	Studies that are about change in people (e.g. farmer, fisher, worker), rather than about land, species or operations linked to conservation impacts
INTERVENTION	**Include...**	**Exclude...**
Becoming or being certified or verified by recognised VSS[7] in agriculture, forestry, marine fisheries or aquaculture. These can be global, national or regional schemes.	Studies that discuss impacts of voluntary sustainability standards (do not have to be members of ISEAL) in the forestry, agriculture, marine fisheries or aquaculture sectors. ('Agriculture sector' includes livestock and excludes wild harvesting)	Studies that discuss the effects of: • ISO standards • Organic standards (unless in combination with another VSS that is included in the scope of the search) • Food safety standards incl. Kosher and Halal, or other standards that address sustainability only incidentally or not at all • Protected Geographic origin certificates • Moratoriums/ government policies • Farmer field schools • Rating systems and eco-labels that lack the characteristics of VSS • Payment for ecosystem services initiatives e.g. REDD • Standards that rely only on self-assessment for assurance • Wild harvesting
COMPARATOR	**Include...**	**Exclude...**
Absence of intervention, as defined above	Studies falling into all evidence typologies in BLOC1 and into 3 categories in BLOC2 (See Annex 1) • Outcomes of treatment and control groups compared, before and after an intervention, with random assignment to each group. • Outcomes of treatment and control groups compared, before and after an intervention, with covariates considered through matching. • Outcomes of treatment and control groups compared, at one point in time post-intervention, with covariates considered through matching.	Studies falling into any other evidence typologies Studies that only compare two different certifications to each other (and not to a non-certified control), unless there is overlap in multiple certification (e.g. RA vs FT = reject. RA + FT vs FT = accept as can determine the effects of RA)

7 / We adopted the following definition of VSS developed by Jeff Milder and Deanna Newson: VSS schemes are: (1) owned by non-state actors including civil society (hence 'voluntary') and (2) developed in partnership with a range of non-governmental actors such as civil society groups and businesses. (Systems that are developed and owned solely by a company or industry association without meaningful civil society involvement are not considered VSS)

VSS:
1. Define normative requirements and set expectations for organizational management, practice, or behaviour (hence "standards")
2. Address one or more sustainability topics (hence "sustainability")
3. Include a verification or assurance system (systems that rely only on self-assessment are not considered VSS)
4. Include a mechanism for market recognition and/or differentiation (which implies the possibility of market incentives and advantages for compliance). This mechanism can be B2B or B2C, and can include labelling or other approaches.
5. Provide market incentives to altering production processes towards more sustainable ones

OUTCOME	Include...	Exclude...
Conservation outcomes within scope are those listed in 'Include' column and similar types of conservation outcomes. Generally outcomes may be measured directly or through scientifically appropriate proxies	**Ecosystem integrity** • Deforestation • Reforestation, afforestation, or forest restoration • Conservation or restoration of non-forested natural ecosystems • Set asides of HCV or riparian buffer or other conservation areas • Measures of habitat configuration (e.g., patch size, fragmentation, connectivity) • Changes in seabed habitats, including effects on substratum, geomorphology and biota • Extent or change in extent of native vegetation cover • Level or change in vegetation cover quality (e.g., plant species diversity or assemblage) • (Reduction in) fire occurrence • Ecosystem services provision from terrestrial ecosystems (e.g. carbon storage) • Changes in ecosystem structure and function (trophic relationships, biodiversity etc **Biodiversity** • Assemblages of terrestrial wildlife (e.g., species richness, abundance., or assemblage metrics of wildlife taxa) • Conservation of focal species (e.g., rare, endangered, or keystone species) • Stock heath and stock status of target species; fishing mortality, stock recovery • Bycatch reduction / status of non-target species in fisheries **Pollution and waste** • Agrochemicals • Water quality • Air quality	This study will not include changes in behavior (e.g., "practice adoption"), except where the changes in behaviour or practices are identical or tantamount to outcomes within scope (e.g., cessation of deforestation can be considered both a practice and an outcome). We will exclude outcomes related to aquatic ecosystems.
OTHER	Include...	Exclude...
	Empirical studies	• Theoretical studies or models • Modelling or discussion papers that make no mention of collecting raw/original data



3. Paper Summaries and Quality Assessment

In the screening process for this review, only studies that fit in particular evidence types were retained for inclusion in our report (see Annexes 1 and 2). Included studies were also subjected to a quality assessment. Studies were assessed in first instance on the following attributes:

- Whether the design was appropriate to capture desired outcomes as stated in objectives of the study (i.e., study addressed research questions)
- If in practice the studies considered confounding factors and how
- Sample size

Further considerations for quality assessment included a modification of suggested screening practices from other articles (e.g., Mupepele et al., 2016), and overall aimed to answer the following questions:

- Were data collected and analyses used sufficient to answer the question(s) asked?
- Were assumptions made throughout the study development valid?
- Was the sample representative of the population and what was selection procedure?
- How was (were) the control group(s) assigned (e.g., propensity score matching or researcher opinion)?
- Were drop-outs reported and followed up?
- Were there statistical analyses carried out? Were these appropriate and did they include a measure of error?
- Was the assessor able to understand statistics to the degree that conclusions could be evaluated?
- Were conclusions drawn in line with the data and analyses presented?
- Was there any obvious risk of bias or other sources of subjectivity (e.g. who paid for the study or if samples were all from one company)?

Studies were assessed by at least two working group members and information was extracted and compiled for each study in an Excel datasheet. This database is an extension of the one that was developed for coding through the systematic search and inclusion/exclusion decision process, which included other information on each study that contained information on specific location (spatial coordinates); sample size and criteria for unit selection; impact(s) measured and metric(s); methods for data collection, unit selection, treatment assignment, and data analysis; and statistical analyses used, among other information.

Studies were ranked as follows:
1 = strong on all of our required dimensions; **1.5** = good, has limited but more complete effort at careful controls; **2** = good but limits; **3** = useful if less strong; **4** quite weak.

Individual paper assessments rankings for the forestry sector

- FSC was considered in all studies – while PEFC in 2 and CERTFOR and JSP each were considered in one.
- Covariates were considered in all studies, using a variety of methods from controls in a GLM or panel regression to various forms of matching (qualitative, propensity score matching, synthetic control method, sometimes pre-analysis).
- 10 were included in the quality assessment since they met the above-mentioned inclusion criteria.

Blackman et al. 2018: This article rigorously examined effects of FSC certification on rates of forest cover reduction in Mexico using remote-sensing data. Authors dealt with several issues including possible contamination of the control group by including firms that may have lost certification and considered time for decision-making for suspension to termination transition. Main result is changes in forest cover after certification. Considers years in certification and aimed to capture temporal effects of FSC certification adoption by including anticipatory effects (before certification), annually after FSC, cumulative on total certified time, and time since first certification. Tested for equal trends assumption and other robustness checks. Overall, no differences in forest cover attributable to FSC and notes non-statistical significant trends in results of FSC associated with increased forest clearing. ***RANK = 1***

Griscom et al. 2014: This study developed a methodological framework that combined field-data and modeling to optimize research efforts to assess carbon emissions associated with several logging activities in Indonesia. It considered slope and standing stocks as major factors that affect how logging occurs and thus resulting carbon emissions. Additionally, they compared emissions performance in 3 certified and 6 no-certified concessions. An attempt to control for confounding factors was done through an ordination exercise on non-certified concessions from which to pull out units on same clusters as certified units to serve as comparable counterfactual units. Ordination used information on area, forest structure, mean slope angles and elevations; logging intensity was treated as a covariate in their analyses. They found differences between certified and non-certified concessions only for skidding-related emissions being lower in FSC certified units. But this difference cannot be fully attributable to FSC. Different logging histories across sites influence size class distribution of all species and as such, affect commercial volumes and management/logging decisions and their consequences in terms of committed and realized emissions. Because there is no limit to timber extraction but timber harvesting is contingent on cutting diameter limits, this difference may influence how each company plans and realizes timber harvesting and the resulting emissions implications from the way they log. ***RANK = 2***

Heilmayr and Lambin 2016: This remote-sensing data paper demonstrates the impact that individual and joint (in all combinations) adoption of FSC, CERTFOR, and JSP, including none adopted, on the rates of avoided forest substitution in company-owned land in Chile. It also carried out analyses of potential spillovers to neighboring company-owned areas, which were not significant. Overall, they found that joint adoption of the three interventions achieved higher reduction of forest conversion to plantations. And that FSC certified companies had lower historic deforestation rates when compared to companies adopting the other two interventions, either jointly or individually. Also, that FSC certified companies had lower forest substitution rates than any of the other two interventions. But a spatial lag model showed that joint adoption of the three interventions had higher avoidance of forest substitution. ***RANK = 1***

Kalonga et al. 2016: This study examines the impacts of different forest management regimes on tree and seedling richness, density, and diversity in Tanzania. The regimes are community-based FSC certification, open access in village forests, and state forests. They had 2 units in each treatment. They qualitatively constructed a counterfactual using information on potential confounding factors (distance between pairs of villages: not too close to have spillovers but not too far to create variation across physiographic and other aspects). They used PRA /focus groups to obtain information useful to cross-validate and explain differences found in biodiversity outcomes through recall of history of threats and management practices. They found higher tree diversity, density, and richness in FSC when compared to the other management regimes. Results for seedlings were obscured due to interactions between risks posed by access and fire and seedling proxies used. They conclude they cannot fully attribute differences observed to FSC exclusively because of contrasting logging histories across sites. They also recognise that they did not have non-certified community-managed forests that were non-certified, so the certification treatment is confounded by the management regime. ***RANK = 3***

Miteva et al. 2015: This remote-sensing data article reports research on certification impacts looking at the village level as the unit of sampling and certification is granted at the forest management unit level. Treatment category was assigned to a village being certified when it contained at least one certified unit, independently of area within village and area covered within certified treatment, both of which may affect proportion of forest cover change (influences outcome of interest). It is unclear whether they had more than one certified unit in same village. The study found FSC certification decreased deforestation rates by 5% and increased forest fragmentation (perforation) rates by 4%. The study did not consider history of certification status when assigning treatment category (units may have lost certification at time of study affecting # villages included as controls). The study also reported a decrease in air pollution on certified villages when compared to non-certified (31%) but could not offer a plausible mechanism, especially given that fire occurrence did not vary between certified and non-certified villages. This is the first application of remote-sensing data analyses to address impact evaluation questions of FSC certification. ***RANK = 1.5.***

Medjibe et al. 2013: This field-intensive study aimed at comparing outcomes of forest management in certified and non-certified units through intensive sampling in each one site for treated and control locations in Gabon. Because there was not random allocation of units into treatment and control groups this non-experimental study attempted to select a comparison unit based on qualitative matching. Still variables that may affect the outcomes of interest were different, which could obscure magnitude of differences found (concession area,

workers, logging intensity). Differences before and after logging were determined for a range of variables (gap area and # trees damaged as a function of harvested tree DBH; # trees damaged/unit skidtrail length expressed on a per area and on a per volume basis; and changes in post-logging species composition). Although suggestive differences in time changes exist for sampled variables for each treatment, the study design does not offer sufficient degrees of freedom to statistically attribute effects to certification adoption. It does, however, illustrate possible methods to use when attempting to answer certification questions on an improved replicated study design. ***RANK = 3***

Panlasigui et al. 2018: This remote-sensing data paper analyses the impact of different types of forest management in Cameroon: protected areas, concessions establishment, and certification through time and amongst management categories. For FSC certification, they found that rates of forest loss significantly but very slightly decreased (0.02%; $p<0.05$) over the 2000-2013 period only in one out of 4 regions. They use panel regressions for all comparisons based on panel data on forest cover change. For FSC, they compare certified and non-certified concessions (now starting with pixel data for all interventions, concessions for robustness for the FSC test, and now also the parallel trends assumption is tested directly). ***RANK = 1***

Rana and Sills 2018: This remote-sensing data article illustrates the use of an innovative methodology, the synthetic control method, to assess deforestation (changes in tree cover loss) impacts of FSC in 3 case-studies, one/country in Gabon, Brazil, and Indonesia. In small sample sizes, confidence intervals become bigger so it is hard to use statistical inference. They established statistical significance at 80% is low (alpha= 0.2). This is a possibly reasonable significance threshold for a proof-of-concept article that aimed to illustrate the use of a methodology and its limitations. Differences between certified and noncertified varied through time in one country with a higher tree cover loss in Brazil, reduced tree cover loss in Indonesia, and no difference in Gabon. ***Study quality rank = 1.5***

Rico et al. 2018: This remote-sensing data paper analyses the impact of different types of forest management categories in Peru: protected areas, concessions establishment, and certification through time and amongst management categories. They found that rates of forest loss significantly but slightly decreased (1%; $p<0.05$) over the 2000-2013 period only in one out of 3 regions. They use panel regressions for this specific comparison based on panel data on forest cover change across certified and non-certified concessions. They performed robustness checks to test whether differences in concession area affected outcomes but do not describe/provide results. ***RANK = 1***

Villalobos et al. 2018: This article rigorously examined the impacts of FSC and PEFC certification's contribution to conservation through degradation avoidance in Sweden. Proxies for degradation were the post-felling preservation of environmentally important areas and establishment of additional set-asides (EIAs, established by the Swedish Forestry regulations), number of trees and of high stumps across groups of certified and comparable non-certified forest owners as reported by state agencies. The article demonstrated that the majority of those managers in treated and control groups does not comply with environmental regulations. Authors also found no effect of FSC on avoiding degradation or none of other possible deleterious conservation effects of forest management. ***RANK = 1***

Table 4.1 Ranking of included forestry papers

AUTHORS	STUDY TITLE	RANKING
Blackman et al. 2018	Does eco-certification stem tropical deforestation? Forest Stewardship Council certification in Mexico	1
Griscom et al. 2014	Carbon emissions performance of commercial logging in East Kalimantan, Indonesia	2
Heilmayr and Lambin 2016	Impacts of nonstate, market-driven governance on Chilean forests	1
Kalonga et al. 2016	Forest certification as a policy option in conserving biodiversity: An empirical study of forest management in Tanzania	3
Medjibe et al. 2013	Logging Concessions Compared in Gabon: Changes in Stand Structure, Tree Species, and Biomass	3
Miteva et al. 2015	Social and environmental impacts of forest management certification in Indonesia	1.5
Panlasigui et al. 2018	Impacts of certification, uncertified concessions, hunting zones and protected areas on forest loss in Cameroon, 2000 to 2013	1
Rana and Sills 2018	Does certification change the trajectory of tree cover in working forests in the tropics? An application of the synthetic control method of impact evaluation	1.5
Rico et al. 2018	Logging concessions, certification and protected areas in the Peruvian Amazon: forest impacts from combinations of development rights and land-use restrictions	1
Villalobos et al. 2018	Has forest certification reduced forest degradation in Sweden?	1

Individual paper assessments rankings for the marine/aquaculture sectors

- MSC and ASC were the only two certification schemes considered in the marine/aquaculture sectors that were included in this study
- In all the studies, except one (ASC, 2016), all available data were used (e.g. all US fisheries (both certified and uncertified) for which data existed were included in the analyses). Given the extensiveness of these studies, covariate matching was not considered necessary.
- Only 6 papers were included in the quality assessment given their compliance with all inclusion criteria. The extra papers that were included in our assessment of marine fisheries for reasons described in the main report were not subjected to quality assessment.

MSC 2016: This study examined all the biomass and fishing pressure exerted on MSC certified and non-certified fisheries in Europe to better understand stock health. Stock health was comparable for fisheries before and after certification as well as with and without certification, thus giving two controls to work with. All available stock status data provided by the International Council for the Exploration of the Seas (ICES), the supplier of information and management advice to the European Commission, was used. Given that all European stocks (for which data exists) was used, co-matching was not performed. The results show the changes in fishing pressure and biomass of European stocks between 2000 (prior to MSC certification) and 2014 for certified and non-certified fisheries. The median, upper quartile, lower quartile, maximum and minimum values were presented as well as all outliers. ***RANK = 1***

MSC 2017: The 2017 report is an extension of the MSC, 2016 analysis (described above) to not only include European fisheries but eight other regions. This included East Canada, West Canada, the USA, Africa, non-EU European countries, Countries in the EU, New Zealand, Australia and the Indian Ocean. Here, the health of marine populations (stocks) targeted by MSC certified and uncertified fisheries around the world were compared. For each region, recent biomass of the stock is compared to data from 2000, prior to MSC certification. Again, all available data from a variety of official sources was used (from large global databases and governmental databases). In some cases, sample sizes were small given the low numbers of certified fisheries in some regions (e.g. only three in South Africa). However, this small sample size is not a reflection of poor analysis but simply a representation the low certification in some regions. The median, upper quartile, lower quartile, maximum and minimum values were presented as well as all outliers. ***RANK = 1***

WWF Austria et al. 2016: Here, interviews at ASC and non-certified Pangasius and shrimp farms in Vietnam were undertaken to compare the use of chemicals, drugs and fossil fuels between certified and noncertified farms. It was found that ASC certified farms have lower levels of each of these inputs than uncertified farms. Small sample sizes (between 5 and 9 farms) meant that outcome could not be statistically significant but still informative. There were also limited attempts at matching control to treatment farms (although certified and uncertified farms occurred in the same geographic regions and farmed the same species). This study focused more on costs and benefits of certification and so focused on the input costs of chemicals for example. However, outcomes still show a reduction in chemical, drug and fuel use. ***RANK = 3***

Nhu et al. 2016: This study conducts an assessment of the environmental benefit of applying certification schemes on Pangasius production by comparing ASC certified and noncertified farms in Vietnam. This is done using a statistically supported Life Cycle Analysis. The study focused on both resource-related (water, land and total resources) and emissions-related (global warming, acidification, freshwater and marine eutrophication) categories. Data uncertainty analysis was then conducted using the Monte Carlo (MC) method, with 1000 iterations which is a sufficient but not excessive sample size. Appropriate statistical tests were subsequently performed to determine whether the differences between the environmental impacts of the two farming systems were significant. All outcomes were significant, save for freshwater eutrophication which was inconclusive. ***RANK = 1***

Selden et al. 2016: A comparison of MSC certified and non-certified fisheries in the United States was performed to understand the difference between gear impacts, marine mammal bycatch and discards. All available data on US fisheries was utilized. Gear types per fishery was collected using the RAM database (a compilation of stock assessment results for commercially exploited marine populations from around the world) and bycatch and discards data was compiled from multiple fisheries and databases. This included data from the U.S. National Bycatch Report, the 2013 update to the National Bycatch Report, the 2012 NMFS List of Fisheries (LOF), and NOAA Marine Mammal Stock Assessment Reports. Where there were mismatches in data, the following adjustments were made for the following reasons; both the discard and the marine mammal bycatch data define fisheries based on region and gear type and/or target species. This means that one certified fishery may be comprised of several NMFS fisheries, each of which may have different MMPA Category designations. In the NMFS northeast region many fisheries are multispecies, and thus defined without reference to a target species. This primarily affected the classification of the spiny dogfish fishery. Fisheries which landed spiny dogfish in 2010 using gears defined as eligible in the MSC unit of certification were considered. This is potentially problematic but considered to not negatively affect overall outcomes. ***RANK = 1***

Gutierrez et al. 2012: Here, the authors analyse 45 certified and 179 noncertified stocks using data from the RAM Legacy Stock Assessment Database, which represents the largest global stock assessment database available at the time. Where newer stock assessments were available, these were included and updated. For each stock, the ratio of the current biomass to the biomass at MSY and the current fishing exploitation (F) to the Fmsy was calculated. When estimates of one or both reference points were not available, they were estimated by fitting Schaefer surplus production models. In addition, 25 stocks (counted among the 179) that a pre-assessment had suggested would fail the Principle 1 of the MSC's Fisheries standard. These were defined as 'non-recommended stocks'. The analysis found that the ratio (Bcurrent/BMSY) is significantly different between certified and noncertified fisheries ($p < 0.005$). Furthermore 74% of certified stocks are above sustainable target biomass levels, compared with 44% of noncertified stocks. Additionally, 82% of certified stocks had current exploitation rates that are expected to maintain the stocks at BMSY or allow for rebuilding to BMSY compared with 65% of noncertified stocks. ***RANK = 1***

Table 4.2 Quality assessment of marine fisheries and aquaculture papers

AUTHORS	STUDY TITLE	RANKING
MSC 2017	Global Impacts Report 2017	1
MSC 2016	Global Impacts Report 2016	1
WWF Austria et al. 2016	Lessons Learned from conducting a cost - benefit analysis for Aquaculture Stewardship Council certified farms in Vietnam: The business case to illustrate value of certification through case studies of ASC certified farms	3
Nhu et al. 2016	Environmental impact of non-certified versus certified (ASC) intensive Pangasius aquaculture in Vietnam, a comparison based on a statistically supported LCA	1
Selden et al. 2016	Evaluating seafood eco-labeling as a mechanism to reduce collateral impacts of fisheries in an ecosystem-based fisheries management context	1.5
Gutierrez et al. 2012	Eco-label conveys reliable information on fish stock health to seafood consumers.	1

Individual paper assessments rankings for the agricultural sector

- RA and RSPO are most commonly studied VSS.
- Methods vary considerably depending on the outcome of interest
- 3 papers that went through quality assessment were ultimately determine not to meet our selection criteria (controlling for covariates non-existent or extremely weak). Those are not listed here. Another 16 papers went through the quality assessment and are discussed in the final report.

Azhar et al. 2015: Considered RSPO in Peninsular Malaysia in a remote-sensing study of 'landscape heterogeneity', though in practice that meant comparing mean size and number of forest patches. The sample size was 70 treatment and 70 control. A weakness of the paper is that it included no discussion of how matching was done, beyond controlling for elevation and distance to coast. The study found that large plantations have low patch numbers, whether RSPO-certified or not. Smallholders, without RSPO, have more patch numbers. ***RANK = 2***

Carlson et al. 2018: Evaluated deforestation rates in Kalimantan and Sumatra, Indonesia from 2001 to 2015 using annual satellite data and a comprehensive RSPO data set. This is a strong study with large sample size (varies by analysis) that controlled for differences between certified and noncertified plantations using propensity score matching. They find a 1/3 drop in deforestation yet it is in plantations mostly already deforested prior to certification (a selection problem). Impact results are driven by Kalimantan. The paper does not officially state a result for fires as the authors found that it was not possible to distinguish the effects of certification from other effects. ***RANK = 1***

Cattau et al. 2016: Considered RSPO and fires in Indonesia during 2012-15 using remote sensing data. That timing itself indicates a weakness – pointed out in Carlson – which is that the time periods are not pre- versus post-certification timing. Matching considered road density and area. Fire reduction is found for low fire likelihoods, i.e., non-peatlands in wetter years. They tried to get 'escaped fires' too but those were too few in number. ***RANK = 1.5***

Caudill and Rice 2016: Assessed the non-volant mammalian fauna and their associated habitat requirements in 23 sites at one point in time representing forest, Bird Friendly (BF) shade, conventional shade and sun coffee habitats. Matching of site selection controlled for proximity to roadways, water, residential areas and elevation, but not farm/forest remnant size, which is very important in biodiversity measurement. Mammal abundance and richness were measured (segregated by size class; small, medium, large). Tree species richness and tree density were also measured. Overall, BF coffee habitats had the highest species density and abundance of mammals, although not always statistically significantly higher than the other habitat types. There was no significant difference in the estimated tree species richness among the different habitat types. ***RANK = 2***

Everage and Ingersoll 2013: This report contains results of numerous studies including one that assessed biodiversity using a Likert scale index based on visual observation of canopy diversity and surveys of coffee farmers in Mexico at one point in time. These measures are weaker than direct observation of biodiversity. Their findings indicate that certified farms had significantly less degraded sparse canopy and significantly greater dense canopy area than control group. Neither certified or noncertified farms had very dense canopy. The report does not provide some important information such as which certification scheme was assessed or some details about study in question. For this reason, it is less useful. ***RANK = 2.5***

Haggar et al. 2017: Study looked at RA, Fair Trade, UTZ, CAFÉ Practices, and non-certified farms across the coffee growing regions of Nicaragua. Farms were propensity-score matched by altitude, area of coffee and farmer education to ensure comparability between noncertified and certified farms. Sample was 81 non-certified farms and between 35 and 48 of each certification. Farms under all certifications had better environmental characteristics than noncertified for some indicators, but none were better for all indicators. Farm certification had a highly significant effect on the Margalef index of tree diversity, with RA, FT, and UTZ all performing better than non-certified and C.A.F.E. practices. ***RANK = 1***

Hardt et al. 2013: evaluated coffee certification in Brazil by comparing changes in relatively complicated measures of the landscape structure (for biodiversity) but also deforestation before as well as nine years after the beginning of a certification process. The study size is small with only 5 treatment groups, farms with RA/SAN certification, and little detail is provided about how matching is done. They note that certified farms had a different conservation profile at the start of the process, complicating comparisons. Using a control group of noncertified farms and the surrounding landscape as a reference, nonetheless the authors looked at coffee-growing areas in two landscapes—savanna and Atlantic forest— and they find no significant differences in deforestation rates between certified and noncertified farms in either landscape. With the small sample size, this is an informative but not definitive study.
RANK = 2.5

Hughell and Newsom 2013: This report pulls together four different studies that looked at many different environmental and ecological variables on coffee farms in Colombia, including water quality, percent shade cover on the farm, and arthropod abundance. A sample size of over 50 control and 50 treatment farms was used, though the number of farms included in each analysis differed. To establish their samples, Hughell and Newsom randomly selected 36 RA certified farms in each of the two regions, and then, as a control, paired each certified farm with the nearest noncertified farm of a similar size, topography, and elevation. They found that percent shade cover was higher in certified farms, but there was no significant difference with non-cert farms. Arthropod richness was significantly higher on certified farms than on non-certified farms. While values increased on certified farms, there was no significant differences for measures of arthropod abundance or diversity. ***RANK = 1-2*** (differs by outcome)

Komar 2012: Examined resident and migratory forest-specialist bird species survivorship and site fidelity across a habitat disturbance gradient of natural forest, forest fragments, RA certified coffee farms, technified coffee farms, and open pasture lands. Sample size was 10 for each control and treatment option. The study controlled for elevation, distance between study sites, and distance to roads, but not plot size. Tree species richness per transect varied significantly across all treatments (certified significantly higher than technified coffee and pasture, but significantly lower than forest fragments and natural forests). There was no significant difference for mean survivorship for resident birds between certified compared to natural forest, forest fragments, pasture or technified coffee habitats. There was no significant difference for resident species site fidelity across habitat types. For both site fidelity and survival of migrant birds, certified was significantly higher than technified coffee and statistically similar to forest fragments, natural forests and pasture. Results indicate that natural forest, forest fragments and certified coffee had similar abundances of migratory birds, and each of these habitats presented significantly higher mean counts of migrants than technified coffee and open areas, resulting in higher abundance, site fidelity and apparent survivorship. For resident generalist birds, the habitat types were similar.
RANK = 2.5

Lentijo et al. 2013: This is a study that was actually designed to examine the effects of a participatory bird census in Colombia rather than certification. However, the sample design does also allow for comparison of RA certified farmers with non-RA certified farmers, both of which did not participate in the bird census. Selection of non-participants in the bird census was done randomly from a list. A posteriori comparison of treatment and control groups was done to check for bias. That comparison is relevant for our purposes here. For yes/no questions about conservation behaviors, RA groups were significantly more likely to preserve forest fragments and natural vegetation present on the farm and vegetation in streams compared with the baseline group. The results are based on farmer interviews only.
RANK = 3

Noojipady et al. 2017: study RSPO and fires in Indonesia and Malaysia. They used time series of satellite data to estimate the spatial and temporal patterns of fire-driven deforestation on and around oil palm plantations, starting with information on over 100 certified plantations and analysing satellite data to fund noncertified plantations. The process of matching is not clear. There is some discussion of selection because if certified have less forest, they also have less forest on which to have new fire. ***RANK = 2***

Rueda and Lambin. 2013: This study of 86 farms (43 matched pairs) found that certified farmers were more likely to have worked to protect watersheds through fencing and restoration. They assessed difference in number of tree species per hectare for RA certified and non-certified coffee farms in Colombia through farmer interviews. The study matching procedure attempted to control for covariates, within what was possible with data limitations, and then checked a posteriori to see if the matching process had in fact controlled for covariates. This study tested for possible selection bias by matching a variety of biophysical, socioeconomic, ecosystem conservation, and land expansion factors.

The research included interviews and field observations.
RANK = 1 - 2 (varies by outcome)

Rueda et al, 2015. This study used remote sensing data on forest canopies to examine coffee farms (all visible without cloud cover) in the Colombian eastern Andes. The study objective was to understand whether RA certification led to enhanced tree cover and landscape connectivity. Certified farms were matched with the closest non-certified farms of similar size (which they knew from a previous 2013 study would control effectively for covariates) and change between 2003 (pre-certification for all but 15 farms) and 2009 was observed. As some farms were only certified towards the end of this period, the authors concluded that impacts of certification are likely underestimated in their analysis. ***RANK = 1.5***

Takahashi and Todo, 2013/2014/2017: Published three papers all about the differences between RA certified and noncertified coffee production, concerning the same forested areas of Ethiopia. In their 2013 paper, which relied on remote-sensing data that was collected in 2005 and again in 2010, the authors found that deforestation rates were lower in forests that had certified coffee production than in forests without coffee. Forests with non-certified coffee, by contrast, did not show a significantly lower deforestation rate than forests without coffee. They concluded that, in the context of forest-grown coffee, certification significantly

Table 4.3 Quality assessment of agriculture papers

AUTHORS	STUDY TITLE	RANKING
Caudill and Rice, 2016	Do Bird Friendly (R) coffee criteria benefit mammals? Assessment of mammal diversity in Chiapas, Mexico	2
Everage and, Ingersoll, 2013	The COSA Measuring Sustainability Report	2.5
Haggar et al., 2017	Environmental-economic benefits and trade-offs on sustainably certified coffee farms	1
Hughell and Newsom 2013	Impacts of Rainforest Alliance certification on coffee farms in Colombia	1-2
Komar 2012	Are Rainforest Alliance Certified coffee plantations bird-friendly	2
Rueda et al. 2017	Responding to globalization: impacts of certification on Colombian small-scale coffee growers	1.5
Takahash and Todo (2017)	Coffee Certification and Forest Quality: Evidence from a Wild Coffee Forest in Ethiopia	1/1.5
Azhar et al. 2015	Promoting landscape heterogeneity to improve the biodiversity benefits of certified palm oil production: Evidence from Peninsular Malaysia	2
Carlson et al. 2018	Effect of oil palm sustainability certification on deforestation and fire in Indonesia	1
Cattau et al. 2016	Effectiveness of Roundtable on Sustainable Palm Oil (RSPO) for reducing fires on oil palm concessions in Indonesia from 2012 to 2015	1.5
Hardt et al. 2013	Does certification improve biodiversity conservation in Brazilian coffee farms?	2.5
Lentijo and Hostetler, 2013	Effects of a participatory bird census project on knowledge, attitudes and behaviours of coffee farmers in Colombia	3
Rueda and Lambin 2013	Eco-certification and coffee cultivation enhance tree cover and forest connectivity in the Colombian coffee landscapes	1-2
Takahashi and Todo, 2014	The impact of a shade coffee certification program on forest conservation using remote sensing and household data	1/1.5
Takahashi and Todo, 2013	The impact of a shade coffee certification program on forest conservation: A case study from a wild coffee forest in Ethiopia	1
Noojipady et al. 2017	Managing fire risk during drought: the influence of certification and El Niño on fire-driven forest conversion for oil palm in Southeast Asia	2

reduced the probability of deforestation. In their 2017 paper they assessed the difference in forest quality between RA certified and non-certified coffee production systems. Forest degradation in the matched control areas was significantly larger than that of the treatment group, indicating that the forest quality was preserved in forest areas around the certified coffee areas compared with the natural forest areas under same environmental conditions. Furthermore, the difference between the treatment and control groups grows as the buffer area increases to the 25 m to 50 m range. The results demonstrate that in forest areas within a 100 m radius, degradation was significantly alleviated for certified areas. These positive spillover effects may be due to economic incentives for certified producers. The authors conclude that forest coffee certification system has a positive impact on preventing forest degradation not only in the certified area but also in the surrounding forest regions. ***RANK = 1*** (1.5 for some outcomes - quality result in 2017; characteristics result in 2014)

4. Systematic Mapping of Agriculture VSS Evidence: Experimental Visualization of Results

Background

The Moore-Meridian working group that produced this report also worked on a second related project – a systematic map of agriculture VSS evidence and a visualization exercise to show the results of the identified studies. The goal was to conduct a systematic search for all relevant research undertaken between 1990 and 2018 on the effects of agricultural sustainability standards on conservation (e.g. biodiversity or deforestation) and then to produce a visualization of study results for an informed but non-academic public. The working group wanted to determine how to present these results in a way that reduces risk of misinterpretation of the study findings and highlights knowledge gaps, where research is missing or scarce.

Systematic mapping exercises do not normally include visual presentations of study findings. The visual presentation of findings falls somewhere between a systematic mapping exercise (where the end goal is to generate a coded database of studies and perhaps a visual representation of what topics and locations the evidence base covers), and an evidence assessment exercise that aims to draw conclusions about what the results of multiple studies tell us. There is little academic consensus about how to do a credible visualization of results. For this reason, we call the visualization we produced an experiment.

To elaborate the visualization approach, our working group consulted with communication and visualization experts[8] who advised on communication best-practices and pitfalls, helped generate possible approaches to visualizing results, and provided feedback on proposed approaches.

We do hope that the final result, as well as the many and long conversations that went into the design of this visualization pilot, will contribute to shape future efforts to graphically depict research results in an accurate and transparent manner and make them more easily accessible, understandable, and visible.

Scope

The agriculture systematic mapping exercise reviewed research undertaken between 1990 and 2018 and included a wider range of study types than those we have reviewed and discussed in this report. In particular, the systematic mapping exercise included study types that compared treatment and control groups without controlling for covariates (see Annex 1 for study typology)[9].

For the visualization presented here, we have included only those studies from the systematic mapping exercise that were published between 2011 and 2018 (to align to the timeline of this report). These include all of the agricultural VSS studies discussed in this report, as well as others published in this time period but that, because of their research design, did not qualify for review in this report.

Our initial experimental effort only covered agricultural VSS, but similar visualizations could be produced to depict the forestry, marine, and aquaculture VSS studies that are covered in this report.

Overview of the visualization

What each shape represents Each hexagon on the visualization represents one result from one study. In many cases, a single study contained multiple relevant results. For example, a study might report different results for different study locations (regions or countries), or might report results for different conservation measures (e.g. bird biodiversity and plant biodiversity). In those cases, each result is shown in a separate hexagon, and is separately numbered.

What the colors represent The results were grouped into outcome categories deemed to be of importance to conservation, such as biodiversity, deforestation, or habitat connectivity, among others. These categories are the same categories used in this report. In the visualization pilot, hexagons related to the same outcome category are given the same color. The first figure (Figure 1) presents results for biodiversity and

8 ⁄ To elaborate the visualization approach, our working group consulted with communication and visualization experts (mentioned in the Acknowledgements) who advised on communication best-practices and pitfalls, helped generate possible approaches to visualizing results, and provided feedback on proposed approaches.

9 ⁄ As was the case for this report, the systematic mapping exercise involved a systematic search for articles, a process to determine whether articles met the inclusion criteria for the systematic map, and then coding of a database with information about the studies, including their study type, using the study typology included in Annex 1 to this report.

deforestation. Other outcomes, for which there were fewer results, are on the second figure (Figure 2).

How results are categorized on the visualization
The visualization presents research results along two dimensions – study strength and the directionality of the result, including information on whether the result was found to be statistically significant.

Study strength The first dimension -- the vertical axis – categorizes studies by the design of the study that produced the results. The research design is a first, incomplete proxy for the 'strength' of a study. The following types of studies were included in each category on the vertical access (see Annex 1 for study type codes):

- **Most conclusive** - Includes results of systematic reviews and meta-analyses (study type 1)
- **Conclusive** - Includes studies with experimental and quasi-experimental designs that compare changes over time in treatment and control groups, controlling for covariates. (study type 2 and 3)
- **Less conclusive** - Includes studies that compare treatment and control at one period of time, controlling for covariates. (study type 6)
- **Suggestive only** - Includes studies that compare treatment and control without controlling for covariates. (study type 7 or 8))

The first three categories of studies are discussed in our report. The fourth is not in this report but is included on the visualization, reporting obtained from the information gathered through the systematic map search effort, for studies published between 2011 and 2018.

Study typology is an imperfect proxy for study strength because a poorly implemented study of a strong evidence type might in the end provide less useful information than a strong study of a weaker evidence type. Acknowledging this, we conducted quality assessments of all studies that were included in our report, as described in the report.

As more systematic search and quality assessment efforts increase through time, it may be possible to combine information on study design with expert quality assessments to find a more sophisticated approach to classifying the strength of studies on the vertical axis. . We have not attempted that here as our visualization includes studies for which we did not do a quality assessment.

Directionality of results The second (the horizontal axis of the figures) is the result itself – whether certified areas were found to be better (positive), worse (negative), or not different from comparison groups.

Results where the treatment group was significantly different from the control group were included in the "negative" or "positive" columns of the visualization, depending if the treatment group performed better ("positive") or worse ("negative") than the control group. For example, if tree species diversity was significantly higher on the treatment group, this result was put in the "positive" column. If pollution was significantly lower on the treatment group, this result was also put in the "positive" column.

Between these two are the 'no difference' results. These are results that were not statistically significant and thus considered indistinguishable from zero. For some of the studies in the 'just suggestive categories', no different results are results where the author concludes there was no difference, even if formal statistical analysis was not conducted.

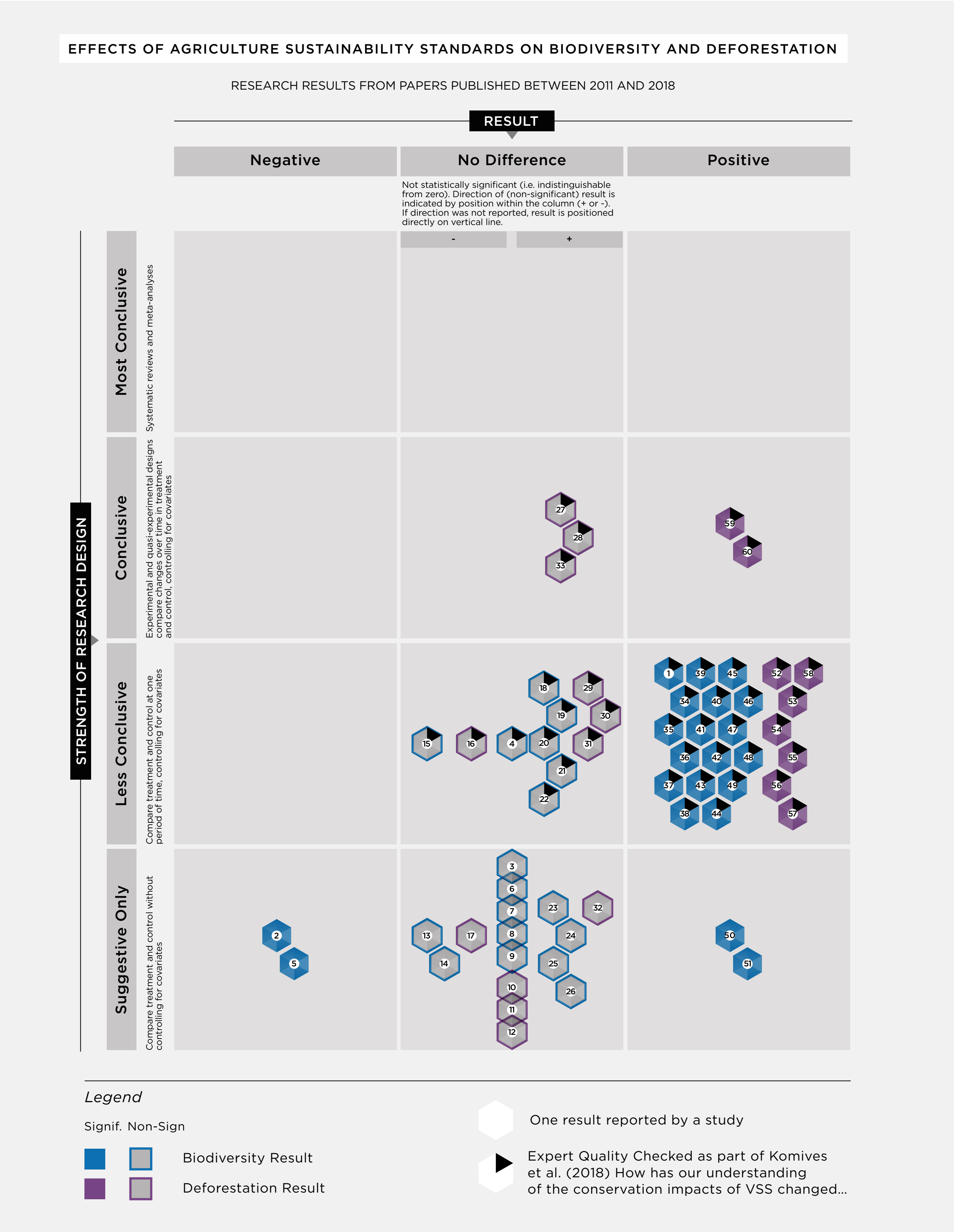
We colored these 'no difference' results gray to indicate a non-significant result. For example, the researcher may have found a positive result but found that that result was not statistically significant. The outline color of these grey hexagons indicates the type of outcome addressed by the study.

In general, all of the 'no difference' results can be taken to mean that there was no difference between certified entities and their controls. This is how the non-significant results are discussed throughout our report.

Nonetheless, the expert team decided to also state the directionality of the non-significant results whenever this trend was reported in a particular study. . Small sample sizes and other design and implementation weaknesses of many studies could generate non-significant results. Large number of negative non-significant results from multiple weak studies could suggest a negative tendency, which would then need to be assessed through stronger-designed studies.

The position of the hexagon within the "no difference" column conveys the directionality of the non-significant result. In other words, the direction of a (not significant) result is indicated by the position of the hexagon within the column (+ or -). If direction was not reported, these grey hexagons are positioned directly on the vertical line.

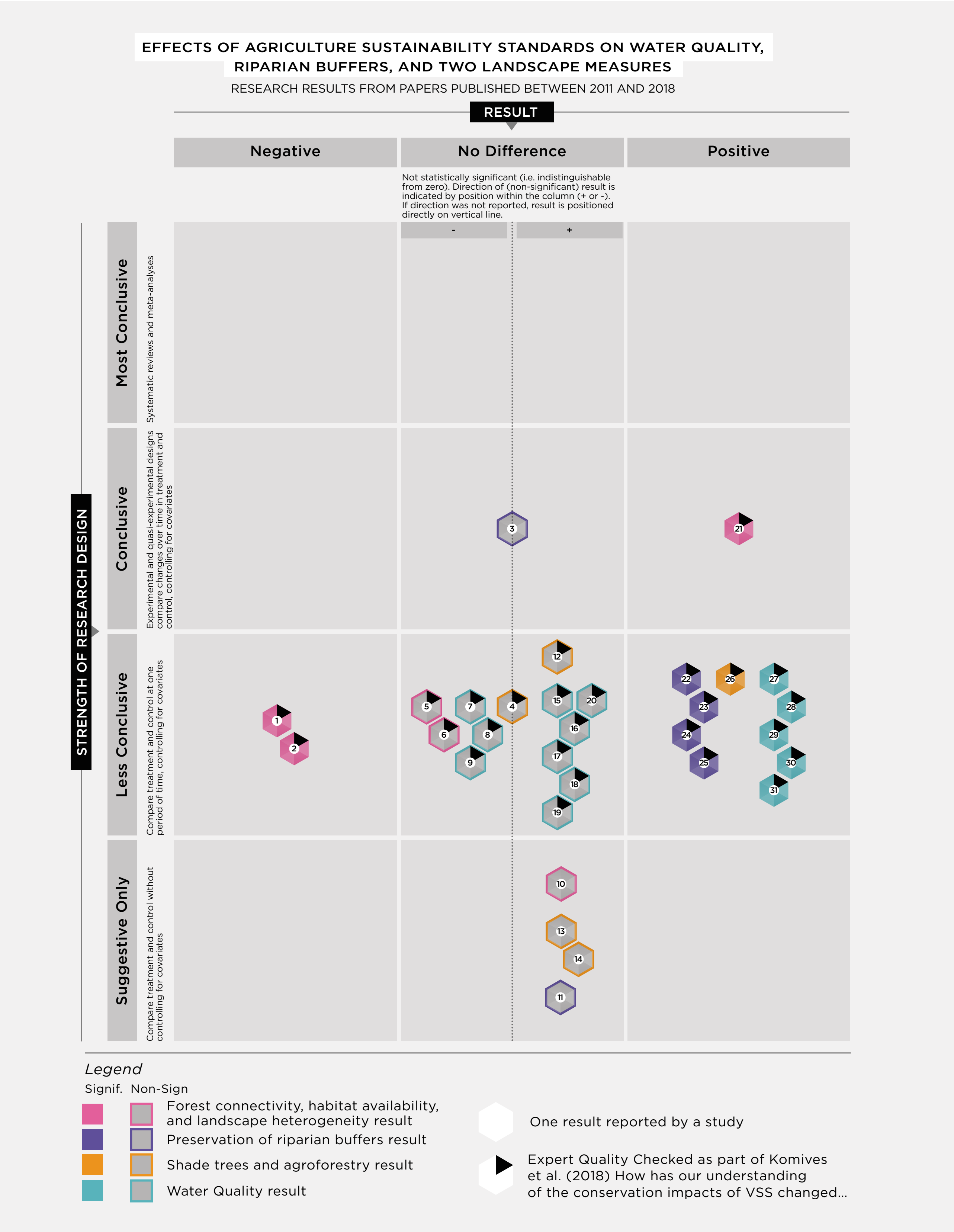
Figure 1: Effects of agricultural VSS on biodiversity and deforestation



MANY STUDIES HAVE MORE THAN ONE POLYGON IN THIS VISUALIZATION. E.G. ONE STUDY COULD HAVE HAD A BIODIVERSITY RESULT RELATED TO A CHANGE IN PLANT SPECIES, AND ANOTHER RELATED TO A CHANGE IN BIRD SPECIES. ONE STUDY COULD HAVE A BIRD RESULT IN ONE STUDY SITE AND ANOTHER BIRD RESULT IN A DIFFERENT SITE.

1. Noojipady, P., Morton, D.C., Schroeder, W., Carlson, K.M., Huang, C., Gibbs, H.K., Burns, D., Walker, N.F., Prince, S.D. (2017) Managing fire risk during drought: the influence of certification and El Niño on fire-driven forest conversion for oil palm in Southeast Asia, Earth System Dynamics...Indonesia, RSPO, Palm oil
2. Bose, A., Vira, B., Garcia, C. (2016) Does environmental certification in coffee promote "business as usual"? A case study from the Western Ghats, India, Ambio...India, Rainforest Alliance, Coffee
3. Meijaard, E., Morgans, C., Husnayaen, Abram, N. K., Ancrenaz, M. (2017) An impact analysis of RSPO certification on Borneo forest cover and orangutan populations...Malaysia/Indonesia (Borneo), RSPO, Palm oil
4. Komar, O. (2012) Are Rainforest Alliance Certified coffee plantations bird-friendly...El Salvador, Rainforest Alliance, Coffee
5. Ochieng, B.O., Hughey, K.F.D., Bigsby, H. (2013) Rainforest Alliance Certification of Kenyan tea farms: a contribution to sustainability or tokenism?, Journal of Cleaner Production...Kenya, Rainforest Alliance, Tea/Roobios
6. Ochieng, B.O., Hughey, K.F.D., Bigsby, H. (2013) Rainforest Alliance Certification of Kenyan tea farms: a contribution to sustainability or tokenism?, Journal of Cleaner Production...Kenya, Rainforest Alliance, Tea/Roobios
7. Philpott, S.M., Bichier, P., Robert, R., Greenberg, R. (2007) Field-Testing Ecological and Economic Benefits of Coffee Certification Programs, Conservation Biology...Mexico, Fairtrade, Coffee
8. Philpott, S.M., Bichier, P., Robert, R., Greenberg, R. (2007) Field-Testing Ecological and Economic Benefits of Coffee Certification Programs, Conservation Biology...Mexico, Fairtrade, Coffee
9. Smith Dumont, E., Gnahoua, G. M., Ohouo, L., Sinclair, F. L., Vaast, P. (2014) Farmers in Cote d'Ivoire value integrating tree diversity in cocoa for the provision of ecosystem services, Agroforest. Syst....Ivory Coast, Rainforest Alliance and UTZ, Cocoa
10. Meijaard, E., Morgans, C., Husnayaen, Abram, N. K., Ancrenaz, M.(2017) An impact analysis of RSPO certification on Borneo forest cover and orangutan populations...Malaysia/Indonesia (Borneo), RSPO, Palm oil
11. Fenger, N.A., Bosselmann, A.S., Asare, R., de Neergaard, A. (2017) The impact of certification on the natural and financial capitals of Ghanaian cocoa farmers, Agroecology and Sustainable Food Systems...Ghana, Rainforest Alliance, Cocoa
12. Smith Dumont, E., Gnahoua, G. M., Ohouo, L., Sinclair, F. L., Vaast, P. (2014) Farmers in Cote d'Ivoire value integrating tree diversity in cocoa for the provision of ecosystem services, Agroforest. Syst....Ivory Coast, Rainforest Alliance and UTZ, Cocoa
13. Bellamya, A.S., Svenssonb, O., van den Brink, P.J., Tedengren, M. (2016) What is in a label? Rainforest-Alliance certified banana production versus non-certified conventional banana production, Global Ecology and Conservation...Costa Rica, Rainforest Alliance, Banana
14. Bellamya, A.S., Svenssonb, O., van den Brink, P.J., Tedengren, M. (2016) What is in a label? Rainforest-Alliance certified banana production versus non-certified conventional banana production, Global Ecology and Conservation...Costa Rica, Rainforest Alliance, Banana
15. Komar, O. (2012) Are Rainforest Alliance Certified coffee plantations bird-friendly...El Salvador, Rainforest Alliance, Coffee
16. Cattau, M. E., Marlier, M. E., DeFries, R. (2016) Effectiveness of Roundtable on Sustainable Palm Oil (RSPO) for reducing fires on oil palm concessions in Indonesia from 2012 to 2015, Environmental Research Letters...Indonesia, RSPO, Palm oil
17. Borden, M. (2016) Social and Environmental Correlates of Palm Oil Certification in Indonesia...Indonesia, RSPO, Palm oil
18. Caudill, S. A., Rice, R. A. (2016) Do Bird Friendly (R) coffee criteria benefit mammals? Assessment of mammal diversity in Chiapas, Mexico, PLoS ONE...Mexico, Smithsonian bird friendly, Coffee
19. Caudill, S. A., Rice, R. A. (2016) Do Bird Friendly (R) coffee criteria benefit mammals? Assessment of mammal diversity in Chiapas, Mexico, PLoS ONE...Mexico, Smithsonian bird friendly, Coffee
20. Everage, L., Ingersoll, C. (2013) The COSA Measuring Sustainability Report...Mexico, Coffee
21. Hughell, D., Newsom, D. (2013) Impacts of Rainforest Alliance certification on coffee farms in Colombia...Colombia, Rainforest Alliance, Coffee
22. Hughell, D., Newsom, D. (2013) Impacts of Rainforest Alliance certification on coffee farms in Colombia...Colombia, Rainforest Alliance, Coffee
23. Fenger, N.A., Bosselmann, A.S., Asare, R., de Neergaard, A. (2017) The impact of certification on the natural and financial capitals of Ghanaian cocoa farmers, Agroecology and Sustainable Food Systems...Ghana, Rainforest Alliance, Cocoa
24. Haggar, J., Jerez, R., Cuadra, L., Alvarado, U., Soto, G. (2012) Environmental and economic costs and benefits from sustainable certification of coffee in Nicaragua, Food Chain...Nicaragua, Rainforest Alliance, Coffee,
25. Ingram V., Waarts Y, Ge L., van Vugt S., Wegner L., Puister-Jansen L., Ruf F., Tanoh R. (2014) Impact of UTZ certification of cocoa in Ivory Coast...Ivory Coast , UTZ, Cocoa
26. Philpott, S.M., Bichier, P., Robert, R., Greenberg, R. (2007) Field-Testing Ecological and Economic Benefits of Coffee Certification Programs, Conservation Biology...Mexico, Fairtrade, Coffee
27. Hardt, E., Borgomeo, E., dos Santos, R. F., Pinto, L. F. G., Metzger, J. P., Sparovek, G. (2015) Does certification improve biodiversity conservation in Brazilian coffee farms?, Forest Ecology and Management...Brazil, Rainforest Alliance, Coffee
28. Hardt, E., Borgomeo, E., dos Santos, R. F., Pinto, L. F. G., Metzger, J. P., Sparovek, G. (2015) Does certification improve biodiversity conservation in Brazilian coffee farms?, Forest Ecology and Management...Brazil, Rainforest Alliance, Coffee
29. Cattau, M. E., Marlier, M. E., DeFries, R. (2016) Effectiveness of Roundtable on Sustainable Palm Oil (RSPO) for reducing fires on oil palm concessions in Indonesia from 2012 to 2015, Environmental Research Letters...Indonesia, RSPO, Palm oil
30. Cattau, M. E., Marlier, M. E., DeFries, R. (2016) Effectiveness of Roundtable on Sustainable Palm Oil (RSPO) for reducing fires on oil palm concessions in Indonesia from 2012 to 2015, Environmental Research Letters...Indonesia, RSPO, Palm oil
31. Rueda, X., Thomas, N.E., Lambin, E.F. (2015) Eco-certification and coffee cultivation enhance tree cover and forest connectivity in the Colombian coffee landscapes, Reg Environ Change...Colombia, Rainforest Alliance, Coffee
32. Borg, J., Selmer, J. K. (2012) From Ghana to Magnum ice cream: tracking down the organisation of sustainable cocoa product chains...Ghana, Rainforest Alliance, Cocoa
33. Carlson, K. M., Heilmayr, R., Gibbs, H. K., Noojipady, P., Burns, D. N., Morton, D. C., Walker, N. F., Paoli, G. D., Kremen, K. (2018) Effect of oil palm sustainability certification on deforestation and fire in Indonesia, PNAS...Indonesia, RSPO, Palm oil
34. Caudill, S. A., Rice, R. A. (2016) Do Bird Friendly (R) coffee criteria benefit mammals? Assessment of mammal diversity in Chiapas, Mexico, PLoS ONE...Mexico, Smithsonian bird friendly, Coffee
35. Caudill, S. A., Rice, R. A. (2016) Do Bird Friendly (R) coffee criteria benefit mammals? Assessment of mammal diversity in Chiapas, Mexico, PLoS ONE...Mexico, Smithsonian bird friendly, Coffee
36. Caudill, S. A., Rice, R. A. (2016) Do Bird Friendly (R) coffee criteria benefit mammals? Assessment of mammal diversity in Chiapas, Mexico, PLoS ONE...Mexico, Smithsonian bird friendly, Coffee
37. Caudill, S. A., Rice, R. A. (2016) Do Bird Friendly (R) coffee criteria benefit mammals? Assessment of mammal diversity in Chiapas, Mexico, PLoS ONE...Mexico, Smithsonian bird friendly, Coffee
38. Caudill, S. A., Rice, R. A. (2016) Do Bird Friendly (R) coffee criteria benefit mammals? Assessment of mammal diversity in Chiapas, Mexico, PLoS ONE...Mexico, Smithsonian bird friendly, Coffee
39. Caudill, S. A., Rice, R. A. (2016) Do Bird Friendly (R) coffee criteria benefit mammals? Assessment of mammal diversity in Chiapas, Mexico, PLoS ONE...Mexico, Smithsonian bird friendly, Coffee
40. Everage, L., Ingersoll, C. (2013) The COSA Measuring Sustainability Report...Mexico, Coffee
41. Everage, L., Ingersoll, C. (2013) The COSA Measuring Sustainability Report...Mexico, Coffee
42. Haggar, J., Soto, G., Casanovesb, F., Virginio, E.M. (2017) Environmental-economic benefits and trade-offs on sustainably certified coffee farms, Ecological Indicators...Nicaragua, Fairtrade, Rainforest Alliance, UTZ, Coffee
43. Hughell, D., Newsom, D. (2013) Impacts of Rainforest Alliance certification on coffee farms in Colombia...Colombia, Rainforest Alliance, Coffee
44. Komar, O. (2012) Are Rainforest Alliance Certified coffee plantations bird-friendly...El Salvador, Rainforest Alliance, Coffee
45. Komar, O. (2012) Are Rainforest Alliance Certified coffee plantations bird-friendly...El Salvador, Rainforest Alliance, Coffee
46. Komar, O. (2012) Are Rainforest Alliance Certified coffee plantations bird-friendly...El Salvador, Rainforest Alliance, Coffee
47. Rueda, X. & Lambin, E. (2013) Responding to globalization: impacts of certification on Colombian small-scale coffee growers, Ecology and Society...Colombia, Rainforest Alliance, Coffee
48. Rueda, X. & Lambin, E. (2013) Responding to globalization: impacts of certification on Colombian small-scale coffee growers, Ecology and Society...Colombia, Rainforest Alliance, Coffee
49. Takahashi, R., Todo, V. (2017) Coffee Certification and Forest Quality: Evidence from a Wild Coffee Forest in Ethiopia, World Development...Ethiopia, Rainforest Alliance, Coffee
50. Haggar, J., Asigbaase, M., Bonilla, G., Pico, J., Quilo, A. (2015) Tree diversity on sustainably certified and conventional coffee farms in Central America, Biodivers Conserv...Nicaragua, Rainforest Alliance, Coffee
51. Haggar, J., Asigbaase, M., Bonilla, G., Pico, J., Quilo, A. (2015) Tree diversity on sustainably certified and conventional coffee farms in Central America, Biodivers Conserv...Nicaragua, Rainforest Alliance, Coffee
52. Cattau, M. E., Marlier, M. E., DeFries, R. (2016) Effectiveness of Roundtable on Sustainable Palm Oil (RSPO) for reducing fires on oil palm concessions in Indonesia from 2012 to 2015, Environmental Research Letters...Indonesia, RSPO, Palm oil
53. Rueda, X., Thomas, N.E., Lambin, E.F. (2015) Eco-certification and coffee cultivation enhance tree cover and forest connectivity in the Colombian coffee landscapes, Reg Environ Change...Colombia, Rainforest Alliance, Coffee
54. Takahashi, R., Todo, V. (2013) The impact of a shade coffee certification program on forest conservation: A case study from a wild coffee forest in Ethiopia, Journal of Environmental Management, Ethiopia, Rainforest Alliance, Coffee
55. Takahashi, R., Todo, V. (2014) The impact of a shade coffee certification program on forest conservation using remote sensing and household data, Environmental Impact Assessment Review...Ethiopia, Rainforest Alliance, Coffee
56. Takahashi, R., Todo, V. (2017) Coffee Certification and Forest Quality: Evidence from a Wild Coffee Forest in Ethiopia, World Development...Ethiopia, Rainforest Alliance, Coffee
57. Rueda, X. & Lambin, E. (2013) Responding to globalization: impacts of certification on Colombian small-scale coffee growers, Ecology and Society...Colombia, Rainforest Alliance, Coffee
58. Rueda, X. & Lambin, E. (2013) Responding to globalization: impacts of certification on Colombian small-scale coffee growers, Ecology and Society...Colombia, Rainforest Alliance, Coffee
59. Carlson, K. M., Heilmayr, R., Gibbs, H. K., Noojipady, P., Burns, D. N., Morton, D. C., Walker, N. F., Paoli, G. D., Kremen, K. (2018) Effect of oil palm sustainability certification on deforestation and fire in Indonesia, PNAS...Indonesia, RSPO, Palm oil
60. Carlson, K. M., Heilmayr, R., Gibbs, H. K., Noojipady, P., Burns, D. N., Morton, D. C., Walker, N. F., Paoli, G. D., Kremen, K. (2018) Effect of oil palm sustainability certification on deforestation and fire in Indonesia, PNAS...Indonesia, RSPO, Palm oil

Figure 2: Effects of VSS on other conservation outcomes



MANY STUDIES HAVE MORE THAN ONE POLYGON IN THIS VISUALIZATION. E.G. ONE STUDY COULD HAVE HAD A BIODIVERSITY RESULT RELATED TO A CHANGE IN PLANT SPECIES, AND ANOTHER RELATED TO A CHANGE IN BIRD SPECIES. ONE STUDY COULD HAVE A BIRD RESULT IN ONE STUDY SITE AND ANOTHER BIRD RESULT IN A DIFFERENT SITE.

1. Azhar, B., Saadun, N., Puan, C. L., Kamarudin, N., Aziz, N., Nurhidayu, S., Fischer, J. (2015) Promoting landscape heterogeneity to improve the biodiversity benefits of certified palm oil production: Evidence from Peninsular Malaysia, Global Ecology and Conservation...Malaysia, RSPO, Palm oil
2. Azhar, B., Saadun, N., Puan, C. L., Kamarudin, N., Aziz, N., Nurhidayu, S., Fischer, J. (2015) Promoting landscape heterogeneity to improve the biodiversity benefits of certified palm oil production: Evidence from Peninsular Malaysia, Global Ecology and Conservation...Malaysia, RSPO, Palm oil
3. Kuit, M., van Rijn, F., Minh Tu, V. T., Van Anh, P. (2013) The sustainable coffee conundrum: A study into the effects, cost and benefits of implementation modalities of sustainable coffee production in Vietnam...Vietnam, UTZ, Coffee
4. Hughell, D., Newsom, D., 2013, Impacts of Rainforest Alliance certification on coffee farms in Colombia...Colombia, Rainforest Alliance, Coffee
5. Azhar, B., Saadun, N., Puan, C. L., Kamarudin, N., Aziz, N., Nurhidayu, S., Fischer, J. (2015) Promoting landscape heterogeneity to improve the biodiversity benefits of certified palm oil production: Evidence from Peninsular Malaysia, Global Ecology and Conservation...Malaysia, RSPO, Palm oil
6. Azhar, B., Saadun, N., Puan, C. L., Kamarudin, N., Aziz, N., Nurhidayu, S., Fischer, J. (2015) Promoting landscape heterogeneity to improve the biodiversity benefits of certified palm oil production: Evidence from Peninsular Malaysia, Global Ecology and Conservation...Malaysia, RSPO, Palm oil
7. Hughell, D., Newsom, D. (2013) Impacts of Rainforest Alliance certification on coffee farms in Colombia...Colombia, Rainforest Alliance, Coffee
8. Hughell, D., Newsom, D. (2013) Impacts of Rainforest Alliance certification on coffee farms in Colombia...Colombia, Rainforest Alliance, Coffee
9. Hughell, D., Newsom, D. (2013) Impacts of Rainforest Alliance certification on coffee farms in Colombia...Colombia, Rainforest Alliance, Coffee
10. Fenger, N.A., Bosselmann, A.S., Asare, R., de Neergaard, A. (2017) The impact of certification on the natural and financial capitals of Ghanaian cocoa farmers, Agroecology and Sustainable Food Systems...Ghana, Rainforest Alliance, Cocoa
11. Ochieng, B.O., Hughey, K.F.D., Bigsby, H. (2013) Rainforest Alliance Certification of Kenyan tea farms: a contribution to sustainability or tokenism?, Journal of Cleaner Production...Kenya, Rainforest Alliance, Tea/Roobios
12. Haggar, J., Sotob, G., Casanovesb, F., Virginio, E.M. (2017) Environmental-economic benefits and trade-offs on sustainably certified coffee farms, Ecological Indicators...Nicaragua, Fairtrade, Rainforest Alliance, UTZ, Coffee
13. Fenger, N.A., Bosselmann, A.S., Asare, R., de Neergaard, A. (2017) The impact of certification on the natural and financial capitals of Ghanaian cocoa farmers, Agroecology and Sustainable Food Systems...Ghana, Rainforest Alliance, Cocoa
14. Haggar, J., Jerez, R., Cuadra, L., Alvarado, U., Soto, G. (2012) Environmental and economic costs and benefits from sustainable certification of coffee in Nicaragua, Food Chain...Nicaragua, Rainforest Alliance, Coffee
15. Hughell, D., Newsom, D. (2013) Impacts of Rainforest Alliance certification on coffee farms in Colombia...Colombia, Rainforest Alliance, Coffee
16. Hughell, D., Newsom, D. (2013) Impacts of Rainforest Alliance certification on coffee farms in Colombia...Colombia, Rainforest Alliance, Coffee
17. Hughell, D., Newsom, D. (2013) Impacts of Rainforest Alliance certification on coffee farms in Colombia...Colombia, Rainforest Alliance, Coffee
18. Hughell, D., Newsom, D. (2013) Impacts of Rainforest Alliance certification on coffee farms in Colombia...Colombia, Rainforest Alliance, Coffee
19. Hughell, D., Newsom, D. (2013) Impacts of Rainforest Alliance certification on coffee farms in Colombia...Colombia, Rainforest Alliance, Coffee
20. Hughell, D., Newsom, D. (2013) Impacts of Rainforest Alliance certification on coffee farms in Colombia...Colombia, Rainforest Alliance, Coffee
21. Hardt, E., Borgomeo, E., dos Santos, R. F., Pinto, L. F. G., Metzger, J. P., Sparovek, G. (2015) Does certification improve biodiversity conservation in Brazilian coffee farms?, Forest Ecology and Management...Brazil, Rainforest Alliance, Coffee
22. Lentijo, G. M., Hostetler, M. E. (2013) Effects of a participatory bird census project on knowledge, attitudes and behaviours of coffee farmers in Colombia, Environ Dev Sustain...Colombia, Rainforest Alliance, Coffee
23. Lentijo, G. M., Hostetler, M. E. (2013) Effects of a participatory bird census project on knowledge, attitudes and behaviours of coffee farmers in Colombia, Environ Dev Sustain...Colombia, Rainforest Alliance, Coffee
24. Rueda, X. & Lambin, E. (2013) Responding to globalization: impacts of certification on Colombian small-scale coffee growers, Ecology and Society...Colombia, Rainforest Alliance, Coffee
25. Rueda, X. & Lambin, E. (2013) Responding to globalization: impacts of certification on Colombian small-scale coffee growers, Ecology and Society...Colombia, Rainforest Alliance, Coffee
26. Haggar, J., Sotob, G., Casanovesb, F., Virginio, E.M. (2017) Environmental-economic benefits and trade-offs on sustainably certified coffee farms, Ecological Indicators...Nicaragua, Fairtrade, Rainforest Alliance, UTZ, Coffee
27. Hughell, D., Newsom, D. (2013) Impacts of Rainforest Alliance certification on coffee farms in Colombia...Colombia, Rainforest Alliance, Coffee
28. Hughell, D., Newsom, D. (2013) Impacts of Rainforest Alliance certification on coffee farms in Colombia...Colombia, Rainforest Alliance, Coffee
29. Hughell, D., Newsom, D. (2013) Impacts of Rainforest Alliance certification on coffee farms in Colombia...Colombia, Rainforest Alliance, Coffee
30. Hughell, D., Newsom, D. (2013) Impacts of Rainforest Alliance certification on coffee farms in Colombia...Colombia, Rainforest Alliance, Coffee
31. Hughell, D., Newsom, D. (2013) Impacts of Rainforest Alliance certification on coffee farms in Colombia...Colombia, Rainforest Alliance, Coffee

September 2018

SUPPLY CHAIN SUSTAINABILITY RESEARCH FUND CONTACT
Robyn Paulekas

Meridian Institute
PO Box 1829,
105 Village Place,
Dillon, CO 80435,
USA

Tel: +1 970.513.8340

AUTHOR WORKING GROUP CONTACT
Kristin Komives

ISEAL Alliance
The Green House,
244-254 Cambridge
Heath Road,
London, E2 9DA,
United Kingdom

Tel: +44 (0) 20 3246 0066