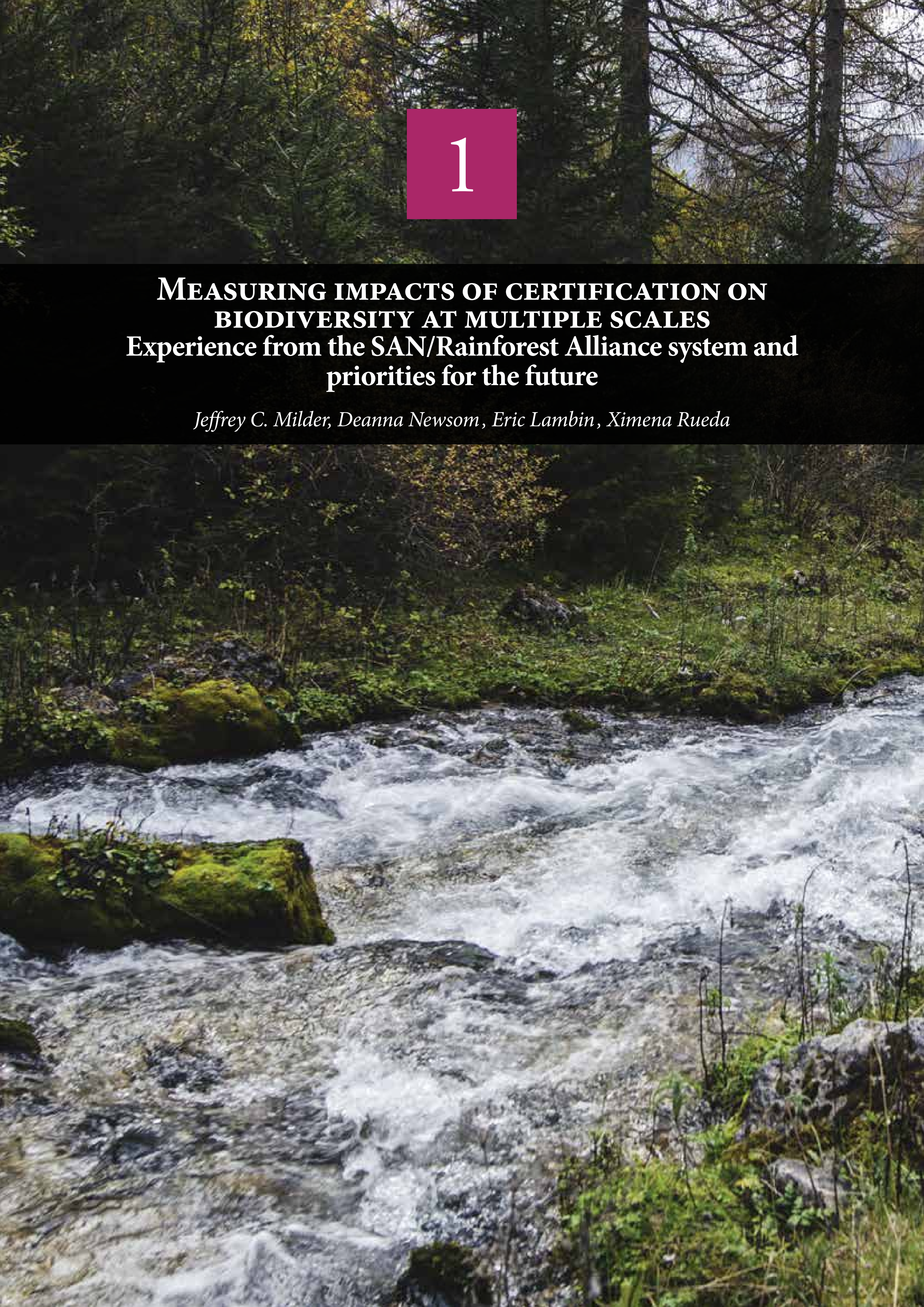


1

Measuring impacts of certification on biodiversity at multiple scales
Experience from the SAN/Rainforest Alliance system and priorities for the future

Jeffrey C. Milder, Deanna Newsom, Eric Lambin, Ximena Rueda

1

Measuring impacts of certification on biodiversity at multiple scales: experience from the SAN/Rainforest Alliance system and priorities for the future

Jeffrey C. Milder[1,2], *Deanna Newsom*[1],
Eric Lambin[3,4], *Ximena Rueda*[5]

Abstract

Voluntary certification standards (VCS) in agriculture and forestry typically include the protection of biodiversity among their objectives or requirements. This outcome is advanced through a range of mechanisms, from prohibitions on destroying certain types of natural ecosystems to requirements to conserve native species co-occurring in production systems to controls on negative externalities that can harm biodiversity, such as polluted runoff. Conservation results may be achieved at a range of scales—from smallholder farms to large landscapes—and as either direct or indirect consequences of implementing VCS. These myriad considerations point to the need for nuanced evaluation frameworks to understand effects on biodiversity across large certification portfolios while also evaluating the causality of VCS interventions on changes in biodiversity attributes. Here, we synthesize experience and perspectives from the VCS community to present a generalized assessment framework for understanding effects of VCS on biodiversity. We then use the example of the Sustainable Agriculture Network (SAN)/Rainforest Alliance certification system to illustrate the application of this framework to an agricultural standard covering about 3.5 million hectares and 1.2 million producers across 42 countries. The framework integrates evidence from multiple data sources, including basic attributes of certified operations, data on the adoption of biodiversity-friendly practices as revealed by annual audits, and research studies assessing biodiversity outcomes at farm and landscape level. Based on experience from evaluating biodiversity effects of the SAN/Rainforest Alliance system from 2011-2014, we reflect critically on challenges, opportunities, and future priorities for evaluating and improving the biodiversity conservation benefits of VCS more broadly.

Keywords: biodiversity, certification, evaluation, impact assessment, sustainable agriculture

1 Evaluation & Research Program, Rainforest Alliance, New York, NY, USA
2 Department of Natural Resources, Cornell University, Ithaca, NY, USA
3 School of Earth Sciences and Woods Institute for the Environment, Stanford University, Stanford, CA, USA
4 Earth and Life Institute, Universite Catholique de Louvain, Louvain-la-Neuve, Belgium
5 School of Management, Universidad de los Andes, Bogotá, Colombia

Introduction

Most of the major voluntary certification standards (VCS) in agriculture and forestry include the protection of biodiversity among their objectives or requirements (International Trade Centre, 2016). For instance, a review of 12 agricultural VCS indicated that all 12 included requirements for habitat protection; 10 prohibited clearance of certain land-cover types; 9 specified criteria for priority habitat areas; 8 addressed impacts to threatened species; and 7 included measures to address invasive species (UNEP-WCMC, 2011). Most standards also address natural resource management issues that can directly or indirectly affect biodiversity, such as water pollution, soil erosion, agrochemical use, waste management, and greenhouse gas emissions.

The recent rapid uptake of VCS in key sectors linked to biodiversity loss, such as various internationally traded tropical agricultural commodities (Potts et al., 2014), suggests that VCS could play an important role in mitigating leading global biodiversity threats. But there has historically been a dearth of evidence on the actual impacts of VCS adoption on biodiversity conservation and other social and environmental outcomes (Blackman et al., 2011). However, over the past few years, this situation has begun to change as VCS systems have upgraded their monitoring and evaluation (M&E) capabilities, as the VCS community has collaborated to define and implement best practices in M&E and impact assessment (e.g., ISEAL Alliance, 2014), and as researchers and others have focused on increasing both the quantity and rigor of studies evaluating the outcomes and impacts of VCS (Steering Committee, 2012).

In this paper, we review recent developments in the evaluation of biodiversity impacts of VCS, analyze the suitability of current approaches, and recommend further actions and investments to fill evidence gaps such that decision-makers can effectively understand and further improve biodiversity impacts of VCS. We begin by presenting a generic framework for evaluating biodiversity impacts of VCS. Next we illustrate how this framework has been applied in the case of the Sustainable Agriculture Network (SAN)/Rainforest Alliance certification scheme. Reflecting on this example, we highlight challenges, opportunities, and priorities for the VCS community and researchers to strengthen the base of decision-relevant evidence on biodiversity impacts of VCS.

A generalized framework for evaluating biodiversity impacts of VCS

Under the Convention for Biological Diversity (CBD), biodiversity is construed broadly to include the diversity of living organisms on Earth as well as the terrestrial and aquatic ecosystems of which they are part (United Nations, 1992). In addition, through its strategic plan and Aichi Biodiversity Targets, the CBD recognizes a range of actions necessary to conserve biodiversity, including not only the protection of species and habitats but also a reduction in threats such as pollution and invasive species (CBD, 2010). Consistent with this multi-faceted approach, biodiversity conservation impacts of VCS may be delivered and evaluated in three main ways (Milder et al., 2015):

1. **Conserving existing natural ecosystems and their associated biodiversity:** many VCS prohibit the destruction or conversion of certain natural ecosystems (e.g., natural forests or High Conservation Value areas) and some require certified operations to manage or restore on-site ecosystems to protect or enhance their biodiversity value. Some standards help conserve

species by prohibiting hunting, wildlife capture, or collection of endangered plants.

2. **Improving the conservation value of production systems and landscapes:** many VCS encourage or require conservation-friendly management of production systems, through measures such as maintaining or enhancing tree canopy cover, protecting or restoring other native vegetation, and conserving riparian zones. These actions can contribute to enhancing conservation value in the "matrix" of working lands that are increasingly recognized as critical for species survival (Perfecto et al., 2009).
3. **Reducing off-site environmental impacts:** nearly all VCS address agronomic and natural resource management issues such as soil erosion, water conservation, water quality, nutrient management, and pesticide use. Such requirements can mitigate threats to aquatic biodiversity from eutrophication, sedimentation, or hydrologic alteration, and threats to all biodiversity from toxic chemicals and (less directly) climate change.

Each of these "impact pathways" may be evaluated at three different levels, ranging from the most direct to the least so: i) adoption of specific best management practices (BMPs) or VCS requirements associated with the impact area; ii) documentation of proximate outcomes at the level of individual certified entities (e.g., farms, forest management units, farmer or forest owner groups, or mills); or iii) documentation of broader outcomes at the level of landscapes, watershed, or regions (Milder et al., 2015).

Taken together, the three impact pathways and three results levels define a three-by-three "evaluation matrix" that can be used to characterize evaluation or research efforts assessing biodiversity impacts of VCS.

Case example: the SAN/Rainforest Alliance certification system

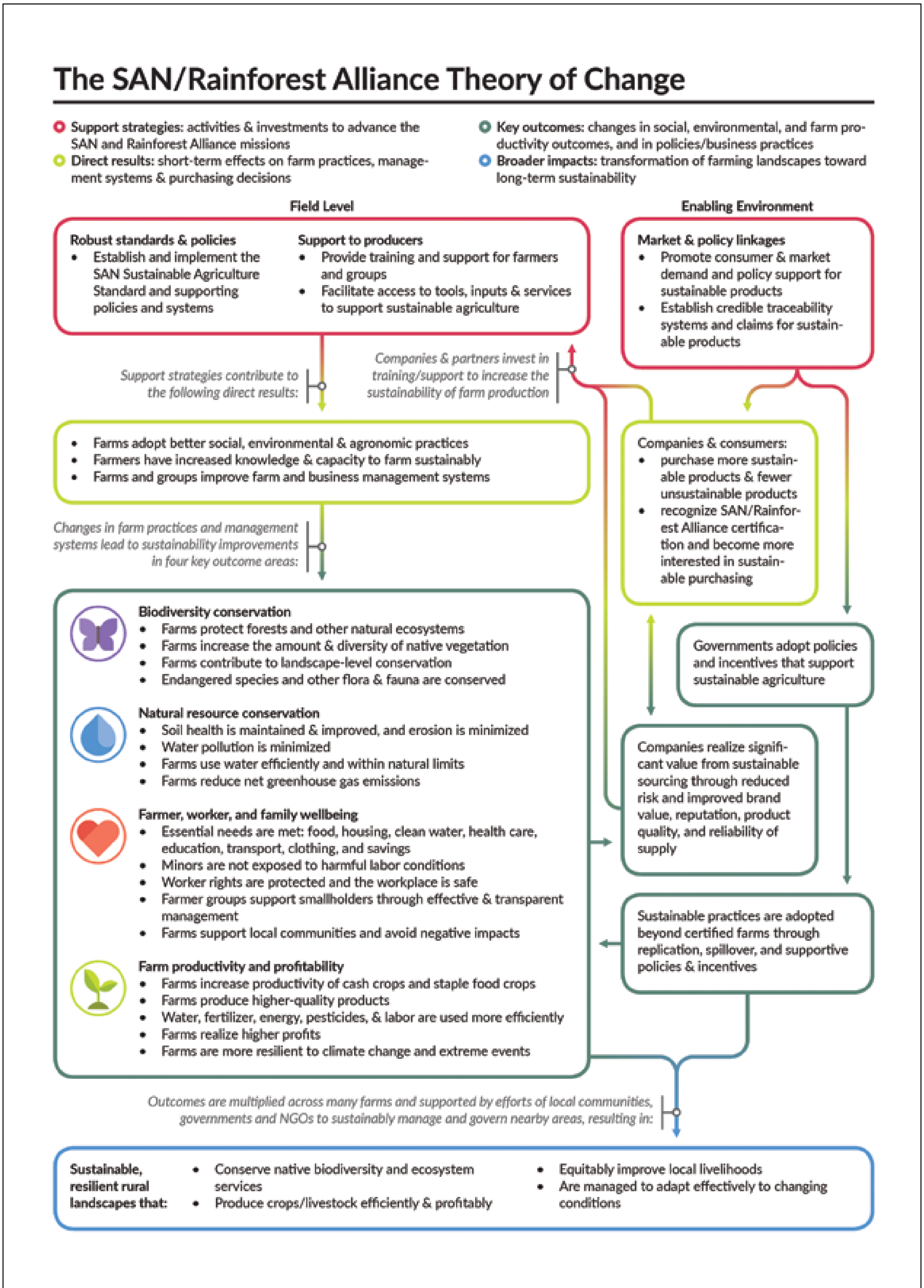
To illustrate the application of this generalized evaluation framework in practice, in this section we present a case example of monitoring and impact assessment for the SAN/Rainforest Alliance certification system. SAN/Rainforest Alliance is a global certification scheme that is currently applied on about 3.5 million hectares across 42 mostly tropical and sub-tropical countries. The scheme certifies more than 100 different crops, with coffee, cocoa, tea, and bananas comprising the largest land areas and numbers of producers.

Conservation hypotheses: the Theory of Change

A "theory of change" is a logical framework that defines the means by which an organization or project aims to achieve specific objectives and outcomes through targeted sets of activities or investments. Conservation scientists and practitioners increasingly recognize the value of using a theory of change or results chains to define the intended consequences and hypothesized impact pathways of any given conservation intervention (Margoluis et al. 2013). The ISEAL Alliance also identifies the development of a theory of change as an important step for certification schemes to clarify intended outcomes and establish an M&E framework (ISEAL Alliance, 2014).

The SAN/Rainforest Alliance Theory of Change (Figure 1) defines a four-step results chain in which support strategies (e.g., standard setting, standard implementation, and field support) lead to results at three different levels, mirroring the three results levels identified in the generic framework presented above. These levels include direct results (changes in

Figure 1. The SAN/Rainforest Alliance Theory of Change.



farm management practices, knowledge and capacity), key outcomes (farm-level changes in social, environmental, and farm performance), and broader impacts (transformation of farming landscapes toward long-term sustainability). At the outcome level, the Theory of Change defines four thematic outcome areas: i) biodiversity conservation; ii) natural resource conservation; iii) farmer, worker and family wellbeing; and iv) farm productivity and profitability.

Indicators and means of measure

The M&E framework for the SAN/Rainforest Alliance certification system derives from the Theory of Change, and includes one or more specific indicators related to each of the outcomes (Milder & Newsom, 2015; see Table 1). Practice adoption indicators (Table 1, middle column) are useful for evaluating hypotheses about whether support strategies (e.g., training and certification) lead to changes in the implementation rates of biodiversity-related BMPs. In some cases, when practice adoption is itself equivalent to—or prima facie evidence of—conservation outcomes, data from practice adoption indicators may be used to infer outcome-level results. For instance, the requirement of the SAN standard to conserve all natural ecosystems can be considered both a practice and an outcome. In a less direct way, the adoption of management practices such as increased agroforestry tree cover and protection of riparian buffers may serve as a plausible proxy for biodiversity conservation outcomes where there is strong and consistent evidence that these practices deliver conservation benefits compared to alternative practices (Bhagwat et al., 2008; De Beenhouwer et al., 2013; Newsom et al., 2012; Tscharntke et al., 2015).

Moving further along the results chain, outcome indicators (Table 1, right column) are used to test whether certification (via the implementation of biodiversity-related BMPs) leads to farm- or landscape-level outcomes in landscape composition and structure, sustained populations of native plant or animal species, improvements in water quality downstream of certified farms, or other parameters.

Within the M&E system, these indicators are generally evaluated in two different ways. First, some practice adoption data, along with basic information about each certificate, are collected across the entire certificate portfolio through the auditing mechanism. This approach is useful for documenting characteristics of certified entities and rates of adoption for various BMPs; however, it is limited by the scope of the certification audit and cannot furnish data on many outcome-level results. A second approach to evaluating indicators involves focused research to collect in-depth evidence on farm- and landscape-level outcomes from a subset of producers or landscapes. Studies may collect data on practice adoption, farm-level outcomes, landscape-level outcomes, or all three. Farm-level studies often compare certified production units to non-certified or pre-certified ones, or apply a difference-in-differences approach to track changes on both certified and non-certified operations. While focused research is better able than routine portfolio-wide monitoring to attribute observed results to certification interventions, it is resource intensive and therefore realistically can generate direct evidence on only a small portion of the certification portfolio.

Evaluation results

To date, the M&E system has furnished evaluation results related to all three of the biodiversity impact pathways described in the generalized framework. Here, we summarize this body of evidence and the range of evaluation approaches used to generate it.

Of the three biodiversity impact pathways identified in the framework, evidence on the conservation value of production systems and landscapes is the most abundant. Studies of all four of the largest SAN-certified crops—coffee,

Table 1. SAN/Rainforest Alliance monitoring and evaluation system indicators related to biodiversity conservation and natural resource conservation outcomes specified in the Theory of Change.

Theory of Change results theme	SAN/Rainforest Alliance M&E system indicators	
	Intended to be assessed for all certificates through auditing and traceability processes	Intended to be assessed for a sample of certified operations, or as part of impact studies
Biodiversity: Farms protect forests and other natural ecosystems	Land area under conservation management, by location and management objective Conformance with key SAN criteria, by crop and location	Rate of ecosystem destruction or restoration compared to surrounding areas Water quality and habitat quality characteristics in aquatic natural ecosystems
Biodiversity: Farms increase the amount and diversity of native vegetation	Conformance with key SAN criteria, by crop and location	Quantity and diversity of on-farm vegetation
Biodiversity: Farms contribute to landscape-level conservation	Conformance with key SAN criteria, by crop and location	Changes in landscape composition and structure following certification
Biodiversity: Endangered species are protected and all native flora and fauna are conserved	Conformance with key SAN criteria, by crop and location	Presence, abundance, or survivorship of species in key taxa around certified farms
Natural resources: Soil health is maintained and improved, and erosion is minimized	Conformance with key SAN criteria, by crop and location	Adoption of specific practices to foster soil conservation and health Fertilizer application rates relative to crop requirements Sediment load in receiving water bodies on or near certified farms
Natural resources: Water pollution is minimized	Conformance with key SAN criteria, by crop and location	Chemical and biological properties of receiving water bodies on or near certified farms
Natural resources: Farms use water efficiently and within natural limits	Conformance with key SAN criteria, by crop and location	Quantity of irrigation water used per unit crop produced (irrigated crops only)
Natural resources: Farms reduce net greenhouse gas emissions	Conformance with key SAN criteria, by crop and location	Estimates of net GHG emissions based on existing calculator tools

tea, cocoa and bananas—found that certified farms performed better than their non-certified neighbors with regard to the creation and maintenance of riparian buffer zones and the retention and planting of shade trees. Compared to non-certified farms, SAN-certified farms had more tree species per hectare (Rueda & Lambin, 2013) and a higher rate of tree cover increase (Rueda et al., 2015), although one study found no difference in canopy cover or the number of emergent trees (Komar, 2012). Time series analysis showed increased BMP adoption for many conservation-related topics but not all. Coffee, tea, and cocoa farms (but not banana farms) increased uptake of BMPs related to creation and maintenance of riparian buffers (criterion 2.5) and creation of buffers between areas of chemical use and natural ecosystems (criterion 2.6) (Milder & Newsom, 2015). Nearly all coffee farms were in compliance (or came into compliance) with shade cover and tree species diversity parameters for agroforestry shade canopies, but few cocoa farms complied with these parameters. All three studies that have examined the contribution to conservation value at a landscape level found positive effects of SAN/Rainforest Alliance certification (Rueda et al., 2015; Hardt et al., 2015; Takahashi & Todo 2013). For instance, Rueda and colleagues (2015) found that dense forest cover had increased in landscapes with substantial areas of certified agriculture, improving forest connectivity.

Several studies have examined the role of certification in reducing off-site environmental impacts. Eight studies found that SAN-certified farms implemented BMPs in the following areas at higher rates than nearby non-certified farms: wastewater treatment, downstream water quality monitoring, agrochemical reduction and safe storage, and the use of soil analytics to guide fertilizer and agrochemical application. Two studies documented certificate-level outcomes related to off-site environmental impacts: one found that SAN-certified coffee farms had better wastewater treatment and less erosion around water sources than non-certified farms (Haggar et al., 2012) while another documented improved water quality as indicated by the Stream Visual Assessment Protocol and levels of pollution-intolerant aquatic species (Hughell & Newsom, 2013). Time series analysis based on audit data revealed improvements in the uptake of BMPs related to off-site environmental impacts, including practices addressing wastewater treatment, water quality monitoring of discharge, water quality monitoring of downstream water bodies, and the rotation and reduction of agrochemical use.

The impact pathway with the least amount of independent research is the conservation of existing natural ecosystems and their associated biodiversity. However, with regard to this pathway, the SAN standard includes strict requirements to protect existing natural ecosystems. These requirements are reflected in audit data by the full adoption of SAN criterion 2.1 (which requires the protection and restoration of natural ecosystems) and criterion 2.2 (which prohibits the destruction of natural ecosystems).

In terms of the quantity of evidence at each level of the results chain, the largest amount of evidence relates to direct results (i.e., implementation of BMPs): a total of nine studies have compared rates of adoption of biodiversity-related BMPs on SAN-certified vs. non-certified farms, while time-series trajectories of practice adoption have been evaluated for 219 certificates across 13 countries. Seven studies evaluated farm-level outcomes (focused mainly on coffee farms) while three studies examined landscape-level biodiversity outcomes (focused exclusively on landscapes with certified coffee farms).

Evaluation gap analysis

Using the structure of the generalized evaluation framework, we conducted a gap analysis to characterize the level of evidence on effects of SAN/Rainforest Alliance certification on different biodiversity-related results and to identify gaps and priorities for future work (Table 2). Overall, there is relatively strong evidence that certified farms implement best management practices at higher rates than non-certified farms, and that rates of adoption increase over time on certified farms (see Milder & Newsom, 2015, for a review of the literature on BMP adoption rates for certified and non-certified coffee, cocoa, tea and banana farms). Given that this evidence already exists – but that studies on BMP adoption can offer at best proxy information on biodiversity outcomes – we consider further research on BMP adoption to be a lower priority for future research.

In contrast, field research that compares farm-level outcomes on certified and non-certified farms is a higher priority. Due to the relatively high costs of field research, long time periods sometimes necessary to observe effects, and other methodological issues (described in the following section), there remains insufficient evidence on farm-level outcomes for all biodiversity-related themes. Research priorities include investigating

Table 2. Gap analysis characterizing the level of evidence on effects of SAN/Rainforest Alliance certification on different biodiversity-related results and identifying priorities and suggested topics for future evaluation and research work. "C/NC" signifies a comparison of certified versus non-certified scenarios. The stated priority levels are based on the authors' assessment of the quality of the existing base as well as the expected utility of additional research results. Under the impact pathway "conserving existing natural ecosystems," BMP adoption and certificate-level outcomes are merged because best practices required by certification (e.g., conserve natural ecosystems) are equivalent to certificate-level outcomes (e.g., existing natural ecosystems are protected from destruction or degradation).

<table>
<tr><th rowspan="2">Impact pathway</th><th colspan="3">Evaluation level</th></tr>
<tr><th>Adoption of BMPs</th><th>Documentation of farm-level outcomes</th><th>Documentation of landscape-level outcomes</th></tr>
<tr><td>Conserving existing natural ecosystems</td><td colspan="2">Existing level of evidence: medium

Priority for future work: medium

Suggested topics:

C/NC comparisons of the size, configuration, health, and management regime of on-farm natural ecosystem set-asides</td><td>Existing level of evidence: low

Priority for future work: high

Suggested topics:

Spatially explicit analysis of contribution of on-farm natural ecosystems to composition and structure of habitats at landscape level

Spatial explicit analysis of changes to landscape composition and structure in landscapes with C/NC</td></tr>
<tr><td>Improving the conservation value of production systems & landscapes</td><td>Existing level of evidence: high

Priority for future work: low

Suggested topics:

C/NC comparison of BMP adoption for crops and regions with less evidence</td><td>Existing level of evidence: low

Priority for future work: high

Suggested topics:

C/NC comparisons of on-farm native vegetation (including canopy cover), riparian zones, water quality

C/NC comparisons or time-series analysis of populations of key species on and around farms</td><td>Existing level of evidence: low

Priority for future work: high

Suggested topics:

C/NC comparisons of off-farm impacts including based on potential spillover effect

C/NC comparisons of aggregate effects of many clustered certified areas on populations of key species</td></tr>
<tr><td>Reducing off-site environmental impacts</td><td>Existing level of evidence: high

Priority for future work: low

Suggested topics:

C/NC comparison of BMP adoption for crops and regions with less evidence</td><td>Existing level of evidence: low

Priority for future work: high

Suggested topics:

C/NC comparisons related to soil erosion, water quality, and pesticide effects</td><td>Existing level of evidence: low

Priority for future work: high

Suggested topics:

research to quantifies or model aggregate effects of certification-related BMPs on water quality, water flow, or pesticide effects</td></tr>
</table>

on-farm and downstream water quality and aquatic biodiversity, soil health, on-farm native vegetation (including shade canopy cover), and on-farm faunal populations. Spatially explicit assessments of landscape-level outcomes, where feasible, are also a high priority. Examples include evaluating the effects of certification on landscape composition and structure (e.g., functional connectively of natural ecosystems) and modeling effects of certification-related BMPs on watershed health (e.g., water quality or water flow at larger scales, where there are concentrations of certified farms).

Challenges, opportunities, and recommendations for improving impact evaluation

Reflecting on experience from the SAN/Rainforest Alliance case example as well as other trends in M&E for VCS and broader developments in data availability and analysis, in this section we identify salient challenges, opportunities, and priorities for improving the evaluation of biodiversity effects of VCS more broadly.

Spatial location of certified units

Recent years have witnessed a revolution in the availability of geographical and environmental data. A global, historical Landsat archive is now on-line through the Google Earth Engine; new Earth observation satellites are being launched, recording data at ever greater spatial, temporal and spectral resolutions; new web-based service platforms facilitate access to these data at low cost; crowdsourcing of local data enhances the validation of global scale maps; and global-scale value-added products such as forest cover change, above-ground biomass, forest fires, and deforestation in places tied to commodity supply chains are freely accessible on-line, notably through the Global Forest Watch web site. For some of these data, time series covering a couple of decades are available, thus allowing measurement of land cover dynamics since the adoption of certification. The high spatial resolution of some of the remote sensing data facilitates evaluation research that links socio-economic data with land cover data at the level of individual decision-making units (i.e., farms).

The impact pathway with the least amount of independent research is the conservation of existing natural ecosystems and their associated biodiversity.

To exploit these rich data to evaluate farm- and landscape-level outcomes of VCS will require precise location information for each farm unit to identify samples of certified and non-certified farms for comparison. Ideally, these data should include exact farm boundaries and an identification of each cultivated plot, especially when certification does not concern the entire production of a farm. While almost all certification programs recognize the utility of spatial data on certified operations, each program is on its own path to acquire that information. For example, the Rainforest Alliance currently publishes an online map with point data for SAN/Rainforest Alliance certificates (http://www.rainforest-alliance.org/work/impact/map) while the Roundtable on Sustainable Palm Oil, with its focus on larger plantations, requires public disclosure of concession boundary data. Other schemes release no spatial data publicly, but have contributed information toward compiling a global atlas of certified lands (http://www.conservation.cam.ac.uk/collaboration/eco-certification-tropical-crops).

Selection bias and rigorous counterfactual

It is well-known that the incremental benefits of VCS relative to a pre-certification baseline may be limited by selection effects—i.e., the phenomenon that producers already conforming to the requirements of certification standards may have stronger incentives to participate in such programs as the cost of compliance is lower for them than for laggards. Rigorous assessment of the effects of certification is complicated by this potential selection bias (Crosse et al., 2011). Research designs with a rigorous counterfactual can help avoid bias in estimates of the impacts of certification (Blackman & Rivera, 2011). Methods

such as propensity score matching may be used when the researcher is not able to conduct randomized designs by selecting *a priori* which farms will be certified or not certified.

Attribution challenges

Assessment of the effects of VCS is also complicated by confounding factors, such as technological progress and market changes that may improve or diminish producers' performance over time for reasons unrelated to certification (Crosse et al., 2011). Moreover, not all differences between farmers in the treatment and control groups are observed. For these reasons, the most rigorous impact evaluation method is based on a matched difference-in-differences design (Heilmayr & Lambin, 2016; Van Rijsbergen et al., 2016; Ruben & Fort, 2012; Ibanez & Blackman, 2016). After matching a sample of treatment and control units – for instance through the use of propensity score matching – the average effect of the VCS on BMP adoption rates or ecosystem change parameters is measured for certified farms. Longitudinal data are used to control for unobserved, time-invariant characteristics of the farming units. Panel methods are then used to calculate the difference between treatment and control groups in time differences in performances, therefore controlling for confounding factors that affect both groups. This approach creates quasi-experimental results (Ferraro & Miranda, 2014). It requires time series of observations for both certified and non-certified farms and, in particular, baseline data to represent the pre-certification situation. Baseline data have rarely been measured as program evaluation is often an after-thought in resource-scarce VCS programs. The method is heavily dependent on untested assumptions and thus reporting on credibility checking of those assumptions has been recommended (Lampach & Morawetz, 2016).

Cumulative and interactive effects, and additionality

VCS are never implemented in an institutional void: rather, they interact with other public and private policies and incentives that also influence decision-making by producers. These multiple policies often work in synergy but sometimes work at cross-purposes. For instance, governments create enabling conditions for private governance to be more effective and scale-up; the state provides extension services that support (or may undermine) implementation of BMPs required by certification standards; governments set up the rule of law and define land rights; and governments remove (or institute) bottlenecks in supply chains and influence consumer awareness and expectations for businesses (Lambin et al., 2014). It is also common for farmers to obtain multiple certifications to hedge their risk and sell at the highest price available at any given time. Often, evaluation programs ignore these interactions between governance regimes and assess the effectiveness of a particular VCS as if it were implemented in isolation. In reality, success stories may be attributed to multiple, independent programs reinforcing each other in complex policy mixes, e.g., following a "carrot-and-stick" configuration. Accounting for the policy ecosystem in which VCS are implemented is therefore essential, although challenging.

These policy interactions raise the question of additionality – that is, whether VCS are producing additional impact on ecosystem conservation beyond a "business-as-usual" scenario.

These policy interactions raise the question of additionality – that is, whether VCS are producing additional impact on ecosystem conservation beyond a "business-as-usual" scenario. This is not necessarily the same as asking whether VCS improve conservation outcomes: in contexts where the prevailing trajectory is one of declining ecosystem health (as it frequently is in tropical production landscapes), interventions to maintain existing conservation status may be construed as bringing additional benefit compared to business-as-usual. Measuring additionality becomes more complex when considering the underlying landscape matrix. As certified farms are part of a landscape mix of conservation areas, forest remnants and other land uses, the effects of VCS on biodiversity enhancement depend greatly on the spatial configuration of the landscape matrix and the proximity of the farms to core reserves and other landscape features.

Conclusions

As VCS schemes have upgraded their M&E systems and the scientific community has increased the quantity and rigor of research on impacts of VCS, the once-accurate refrain that environmental effects of VCS are largely unknown no longer holds true. Certainly, the evidence base is far from complete: as with most other conservation interventions, effects of VCS are difficult to generalize across diverse settings and crops, as the evidence base that does even come close to sampling fully from these disparate contexts suggests. Nevertheless, as illustrated by the case of the SAN/Rainforest Alliance certification scheme, mutually corroborating evidence related to some sets of biodiversity-related results have begun to emerge when portfolio-wide evidence from internal M&E systems is combined with more detailed research studies, including those with credible counterfactuals. Such evidence provides a foundation upon which future evaluation and research efforts can build in a targeted way to fill key gaps such as those defined in this paper.

More broadly, the time is ripe for a large-scale, systematic effort to monitor and adaptively manage VCS to create a feedback loop toward continuous improvement and increased effectiveness of these private instruments for environmental governance. Such a system is now possible based on recent progress in the availability of environmental data and in the sophistication of evaluation methods based on counterfactuals. In its broad outline, such a global system would overlay property boundaries of all certified operations on high spatial resolution environmental databases such as Global Forest Watch. It would track over time environmental indicators such as forest-cover change, biomass, ecosystem-level biodiversity, and landscape connectivity. Panel analyses would compare trends in these indicators from a baseline – ideally, the year of adoption of VCS for every unit – to a random matched sample of farms that do not participate to the VCS. The very large number of observations would allow analyzing impacts of VCS in diverse policy contexts, therefore facilitating an analysis of interactions between multiple private and public policies. In a similar way that "big data" methods have revolutionized other fields, the large volumes of geo-referenced data on the adoption of VCS and its component BMPs at the level of production units has the potential to greatly improve the deployment, understanding, and policy support for sustainable production systems. A more systematic, large-scale evaluation would increase the effectiveness of VCS through better targeting of places and actors that are lagging in environmental performance, fine tuning of standards to local priority issues, rapid detection of areas with low compliance or additionality, and design of optimal policy mixes.

References

Bhagwat, S.A., Willis, K.J., Birks, H.J.B. & Whittaker, R.J. (2008). Agroforestry: a refuge for tropical biodiversity? *Trends in Ecology & Evolution, 23*, 261-267.

Blackman, A. & Rivera, J. (2011). Producer-level benefits of sustainability certification. *Conservation Biology 25*, 1176-1185.

CBD [Convention on Biological Diversity]. (2010). COP 10 Decision X/2: Strategic Plan for Biodiversity 2011-2020. Retrieved from https://www.cbd.int/decision/cop/?id=12268

Crosse, W., Newsom, D. & Kennedy, E. (2011). Recommendations for improving certification impacts. Appendix I. Toward sustainability: the roles and limitations of certification. Washington, DC: RESOLVE, Inc.

De Beenhouwer, M., Aerts, R. & Honnaya, O. (2013). A global meta-analysis of the biodiversity and ecosystem service benefits of coffee and cacao agroforestry. *Agriculture, Ecosystems & Environment, 175*, 1-7.

Ferraro, P.J. & Miranda J.J. (2014). The performance of non-experimental designs in the evaluation of environmental programs: A design-replication study using a large-scale randomized experiment as a benchmark. *Journal of Economic Behavior & Organization, 107*(Part A), 344–365.

Haggar, J., Cuadra, L., Alvarado, U., & Soto, G. (2012). Environmental and economic costs and benefits from sustainable certification of coffee in Nicaragua. *Food Chain, 2*: 24–41.

Hardt, E., Borgomeo, E., dos Santos, R. F., Pinto, L. F., 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). Effectiveness of collaborative versus confrontational market governance of deforestation: the case of Chile. *Proceedings of the National Academy of Sciences*, doi/10.1073/pnas.1600394113.

Hughell D. & Newsom D. (2013). Impacts of Rainforest Alliance certification on coffee farms in Colombia. New York: Rainforest Alliance.

Ibanez, M., & Blackman, A. (2016). Is eco-certification a win–win for developing country agriculture? Organic coffee certification in Colombia. *World Development, 82*, 14-17.

International Trade Centre. (2016). Standards Map. Retrieved from http://www.intracen.org/standardsMap

ISEAL Alliance. 2014. Assessing the impacts of social and environmental standards systems: ISEAL code of good practice, Version 2.0. London, U.K.: ISEAL Alliance.

Komar O. (2012). Are Rainforest Alliance Certified coffee plantations bird-friendly? Final technical report for UNDP/RA/GEF Biodiversity Conservation in

Coffee Project. Retrieved from http://www.rainforest-alliance.org/publications/komar-bird-study
Komives, K., Milder, J.C., Arbuthnot, M., Kennedy, E.T., & Stephenson, P.J. (2016). Progress and innovations in monitoring and improving the impacts of sustainability standards on biodiversity and livelihoods. *Policy Matters*, 21.
Lambin E.F., Meyfroidt P., Rueda X., Blackman A., Börner J. Cerruti P.O., Dietsch T., Jungmann L., Karsenty A., Lamarque P., Lister J., Walker N.F., & Wunder S. (2014). Effectiveness and synergies of policy instruments for land use governance in tropical regions. *Global Environmental Change, 28*, 129-140.
Lampach, N., & Morawetz, U. B. (2016). Credibility of propensity score matching estimates. An example from Fair Trade certification of coffee producers. *Applied Economics*, 1-11. doi: 10.1080/00036846.2016.1153795
Margoluis, R., Stem, C., Swaminathan, V., Brown, M., Johnson, A., & Placci, G. (2013). Results Chains: A Tool for Conservation Action Design, Management, and Evaluation. *Ecology and Society, 18*(3), Article 22. http://www.ecologyandsociety.org/vol18/iss3/art22/.
Milder, J. C., Arbuthnot, M., Blackman, A., Brooks, S. E., Giovannucci, D., Gross, L., Kennedy, E. T., Komives, K., Lambin, E. F., Lee, A., Meyer, D., Newton, P., Phalan, B., Schroth, G., Semroc, B., Van Rikxoort, H., & Zrust, M. (2015). An agenda for assessing and improving conservation impacts of sustainability standards in tropical agriculture. *Conservation Biology,* 29(2), 309–320. doi.org/10.1111/cobi.12411
Milder, J. C. & Newsom, D. (2015). 2015 SAN/Rainforest Alliance impacts report: evaluating the effects of the SAN/Rainforest Alliance certification system on farms, people, and the environment. New York: Rainforest Alliance and México D.F., México: Sustainable Agriculture Network.
Newsom, D., Kennedy, E., Miller, J., Bahn, V., & Adhikari, S. (2012). Testing a BMP-based approach for assessing gaps in certification impacts research. Appendix H. Toward sustainability: the roles and limitations of certification. Washington, DC: RESOLVE, Inc.
Perfecto, I., Vandermeer, J., & Wright, A. (2009). *Nature's matrix: Linking agriculture, conservation and food sovereignty*. London, U.K.: Earthscan.
Potts, J., Lynch, M., Wilkings, A., Huppe, G., Cunningham, M. & Voora, V. (2014). The state of sustainability initiatives review 2014. Winnipeg, Canada: International Institute for Sustainable Development and London, U.K.: International Institute for Environment and Development.
Ruben, R., & Fort, R. (2012). The impact of fair trade certification for coffee farmers in Peru. *World Development*, 40(3), 570-582.
Rueda, X., & Lambin, E.F. (2013). Responding to globalization: impacts of certification on colombian small-scale coffee growers. *Ecology and Society, 18*(3): 21.
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. doi.org/10.1007/s10113-014-0607-y
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.
Tscharntke, T., Milder, J. C., Clough, Y., Declerck, F., Waldron, A., Rice, R., & Ghazoul, J. (2015). Conserving biodiversity through certification of tropical agroforestry crops at local and landscape scales. *Conservation Letters, 8*(1), 14–23. doi.org/10.1111/conl.12110
van Rijsbergen, B., Elbers, W., Ruben, R., & Njuguna, S. N. (2016). The ambivalent impact of coffee certification on farmers' welfare: A matched panel approach for cooperatives in Central Kenya. *World Development, 77*, 277-292.
UNEP-WCMC. (2011). Review of the biodiversity requirements of standards and certification schemes. CBD Technical Series No. 63. Montreal, Canada: Secretariat of the Convention on Biological Diversity.
United Nations. (1992). Convention on Biological Diversity. Retrieved from https://www.cbd.int/convention/text