



An agenda for assessing and improving conservation impacts of sustainability standards in tropical agriculture

Jeffrey C. Milder,* ††††† Margaret Arbuthnot,† Allen Blackman,‡ Sharon E. Brooks,§
 Daniele Giovannucci,* † Lee Gross,†† Elizabeth T. Kennedy,* Kristin Komives,‡‡
 Eric F. Lambin,§§*** Audrey Lee,††† Daniel Meyer,‡‡‡ Peter Newton,§§§ Ben Phalan,****
 Götz Schroth,†††† Bambi Semroc,‡‡‡‡ Henk Van Rikxoort,§§§§ and Michal Zrust*****

*Rainforest Alliance, Evaluation & Research Program, 233 Broadway, 28th Floor, New York, NY, 10279, U.S.A.

†World Wildlife Fund, 1250 24th Street NW, Washington, D.C., 20037, U.S.A.

‡Resources for the Future, 1616 P Street NW, Washington, D.C., 20036, U.S.A.

§UNEP-WCMC, Cambridge CB3 0DL, United Kingdom

**Committee on Sustainability Assessment, Philadelphia, PA, 19147, U.S.A.

††EcoAgriculture Partners, 1100 17th Street NW, Suite 600, Washington, D.C., 20036, U.S.A.

‡‡ISEAL Alliance, 50-52 Wharf Road, London, N1 7EU, United Kingdom

§§School of Earth Sciences and Woods Institute for the Environment, Stanford University, 473 Via Ortega, Stanford, CA, 94305, U.S.A.

***Georges Lemaître Centre for Earth and Climate Research, Earth and Life Institute, Université Catholique de Louvain, Place Louis Pasteur 3, 1348, Louvain-la-Neuve, Belgium

†††Roundtable on Sustainable Palm Oil, Kuala Lumpur, Malaysia

‡‡‡Round Table on Responsible Soy Association, Florianópolis, Brazil

§§§International Forestry Resources and Institutions research network, School of Natural Resources and Environment, University of Michigan, 440 Church Street, Ann Arbor, MI, 48109, U.S.A.

****Department of Zoology, University of Cambridge, Cambridge, CB2 3EJ, United Kingdom

††††Rainforest Alliance, Lombardi 34, 6708LT, Wageningen, The Netherlands

‡‡‡‡Conservation International, 2011 Crystal Drive, Suite 500, Arlington, VA, 22202, U.S.A.

§§§§UTZ Certified, De Ruyterkade 6, 1013AA, Amsterdam, The Netherlands

*****Zoological Society of London, Regent's Park, Outer Circle, London, NW1 4RY, United Kingdom

Abstract: *Sustainability standards and certification serve to differentiate and provide market recognition to goods produced in accordance with social and environmental good practices, typically including practices to protect biodiversity. Such standards have seen rapid growth, including in tropical agricultural commodities such as cocoa, coffee, palm oil, soybeans, and tea. Given the role of sustainability standards in influencing land use in hotspots of biodiversity, deforestation, and agricultural intensification, much could be gained from efforts to evaluate and increase the conservation payoff of these schemes. To this end, we devised a systematic approach for monitoring and evaluating the conservation impacts of agricultural sustainability standards and for using the resulting evidence to improve the effectiveness of such standards over time. The approach is oriented around a set of hypotheses and corresponding research questions about how sustainability standards are predicted to deliver conservation benefits. These questions are addressed through data from multiple sources, including basic common information from certification audits; field monitoring of environmental outcomes at a sample of certified sites; and rigorous impact assessment research based on experimental or quasi-experimental methods. Integration of these sources can generate time-series data that are comparable across sites and regions and provide detailed portraits of the effects of sustainability standards.*

†††††email jmilder@ra.org

Paper submitted April 12, 2014; revised manuscript accepted July 25, 2014.

To implement this approach, we propose new collaborations between the conservation research community and the sustainability standards community to develop common indicators and monitoring protocols, foster data sharing and synthesis, and link research and practice more effectively. As the role of sustainability standards in tropical land-use governance continues to evolve, robust evidence on the factors contributing to effectiveness can help to ensure that such standards are designed and implemented to maximize benefits for biodiversity conservation.

Keywords: biodiversity, certification, evaluation, evidence-based conservation, monitoring, voluntary sustainability standards

Una Agenda para Evaluar y Mejorar los Impactos de Conservación de los Estándares de Sustentabilidad sobre la Agricultura Tropical

Resumen: *Los estándares de sustentabilidad y de certificación sirven para diferenciar y proporcionar reconocimiento de mercado a los bienes producidos de acuerdo con las buenas prácticas sociales y ambientales e incluyen típicamente a las prácticas para proteger a la biodiversidad. Dichos estándares han tenido un rápido crecimiento, incluso en comodidades de la agricultura tropical como el cacao, el café, la palma de aceite, la soya y el té. Dado el papel de los estándares de sustentabilidad en influenciar el uso de suelo en zonas críticas de biodiversidad, deforestación e intensificación agrícola, mucho podría ganarse de los esfuerzos para evaluar e incrementar la indemnización de conservación de estas estrategias. Para este fin, diseñamos una estrategia sistemática para monitorear y evaluar los impactos de conservación de los estándares de sustentabilidad agrícola y para usar la evidencia resultante para mejorar la efectividad de dichos estándares con el tiempo. La estrategia está orientada a partir de un juego de hipótesis y preguntas de investigación correspondientes sobre cómo los estándares de sustentabilidad están pronosticados para entregar beneficios de conservación. Estas preguntas se abordan a partir de datos de múltiples fuentes, incluyendo información común básica de auditorías de certificación; monitoreo en campo de resultados ambientales en una muestra de sitios certificados; e investigación de evaluaciones rigurosas de impacto con base en métodos experimentales o casi experimentales. La integración de estas fuentes puede generar datos de series de tiempo que son comparables a lo largo de sitios y regiones y proporcionan retratos detallados de los efectos de los estándares de sustentabilidad. Para implementar esta estrategia, proponemos colaboraciones nuevas entre la comunidad de investigadores de la conservación y la comunidad de estándares de sustentabilidad para desarrollar indicadores comunes y protocolos de monitoreo, fomentar la síntesis y el compartir los datos y enlazar con mayor efectividad la investigación y la práctica. Conforme el papel de los estándares de sustentabilidad en la gobernanza del uso de suelo continúa con su evolución, la evidencia fuerte de los factores que contribuyen a la efectividad puede ayudar a asegurar que dichos estándares son diseñados e implementados para maximizar los beneficios para la conservación de la biodiversidad.*

Palabras Clave: biodiversidad, certificación, conservación con base en evidencias, evaluación, monitoreo, normas voluntarias de sostenibilidad

Introduction

The expansion and intensification of agriculture to meet growing demand for food, feed, and fuel is a major cause of tropical biodiversity loss and ecosystem degradation (Laurance et al. 2014). These effects occur either directly, through the conversion of natural ecosystems to cropland or pastures and the degradation of on-farm habitats, or indirectly, through habitat fragmentation, water pollution or diversion, spread of invasive species, greenhouse gas emissions, and other off-farm environmental impacts. Agriculture may also disrupt a range of ecosystem services, from water cycle regulation to soil protection, that underpin food production and other aspects of human well-being.

Voluntary sustainability standards have emerged as a promising response to the serious challenges associated with tropical commodity agriculture, which include not

only ecological impacts, but also social issues such as child labor, land and water conflicts, and the perpetuation of severe poverty and inequality (Potts et al. 2014). Typically established by nonprofit organizations, defined through multistakeholder processes, and adopted voluntarily by producers and buyers in agricultural supply chains, such standards provide a nonstate governance approach to addressing sustainability concerns (Cashore et al. 2004). Sustainability standards serve to codify the practice of sustainable agriculture in ways that could support the widely held objective of increasing agricultural output while reducing agriculture's ecological impacts and supporting rural livelihoods (Foley et al. 2011; Garnett et al. 2013). Such standards also provide a market mechanism to convert demand for more sustainable agricultural products into farm-level incentives (COSA 2014).

Voluntary sustainability standards systems typically include 4 components: the standard itself, which identifies

putatively sustainable practices or outcomes; assurance systems, including certification audits, intended to assess conformance with the standard by specific production or business units; sustainability labels (ecolabels) that enable producers, brands, and retailers to differentiate their products in the eyes of consumers, intermediary buyers, government regulators, and civil society; and training and technical assistance activities to assist producers in achieving certification. Hereafter, we use the term *sustainability standards* to signify this entire set of linked components. We refer to farms, groups of farms, or other production units (e.g., a mill plus its supply shed) that have complied with sustainability standards as *certified entities*. We focus on third-party sustainability standards in which standard setting and assurance are carried out by independent parties unrelated to the entity being evaluated; however, the issues we consider are relevant to a broader set of sustainability standards initiatives, including standards set by industry associations and individual companies.

Most agricultural sustainability standards address biodiversity conservation and environmental quality. A review of 12 such standards showed 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 soil erosion, pesticide use, water pollution, waste management, and, increasingly, greenhouse gas emissions (Steering Committee 2012). The degree to which each standard delivers environmental benefits depends not only on its requirements, but also on how it is implemented in practice, including the quality and consistency of auditing.

Agricultural sustainability standards have seen rapid uptake in recent years and are now applied on tens of millions of hectares worldwide. From their position as a small niche only a decade ago, such standards now govern a major proportion of the global production of coffee (38% of all production), cocoa (22%), palm oil (15%), and tea (12%), and small but growing production segments for cotton (3%), sugar (3%), soybeans (2%), and bananas (3% of global production but 18% of exported production) (Potts et al. 2014). This growth has been accompanied and accelerated by the increased use of sustainability standards as proxies for acceptable practice within government and nonstate regulatory frameworks, including public-sector procurement policies and lending criteria by private and multilateral banks (Giovannucci et al. 2014). In addition to these regulatory and quasi-regulatory drivers, the business case for adopting sustainability standards may include increased net revenue or market access for producers, and increased market share, improved branding position, and decreased operational or reputational risks for companies (Levin 2012).

Despite the growing uptake of sustainability standards and their explicit conservation objectives, there is little evidence about their conservation impacts. A review of evaluation literature on sustainability certification in selected sectors found 26 peer-reviewed studies on bananas and coffee but only 8 of these addressed environmental results and none of these 8 were judged to have been designed rigorously, based on quantitative measures and a credible counterfactual scenario (Blackman & Rivera 2011).

This dearth of evidence poses risks for many stakeholders and a missed opportunity for conservation. First, companies, investors, and governments rely upon sustainability standards as a credible means to implement sustainability policies and commitments. Without robust evidence of impact, it is difficult to know whether these commitments are being fulfilled and translated into progress for biodiversity conservation and other objectives. Second, sustainability standards are rapidly evolving, but the lack of data on impacts impedes continuous improvement and leads to fragmented policy messages that are not rooted in solid evidence (Newton et al. 2013). Third, better evidence on conservation impacts—and their relationship to social, economic, and productivity impacts—is needed to improve the cost-benefit calculus and business case for producers and businesses to consider sustainability standards.

In short, better evidence on impacts could help position sustainability standards as a powerful adaptive management framework in support of global sustainability—that is, a structured set of initiatives under which key tenets of efficient, equitable, and conservation-friendly agriculture are widely applied, monitored, and analyzed in consistent ways, with the resulting evidence used to revise standards systems and improve their benefits over time. Conversely, a continued lack of evidence could undermine support for sustainability standards (especially the more demanding and costly standards), causing businesses to select less demanding standards or less transparent approaches to sustainable sourcing. A manifest risk is that sourcing programs billed as sustainable might satisfy public demand for green products without actually delivering conservation benefits.

Efforts to generate robust evidence on sustainability standards could draw on recent advances in evidence-based conservation, including best practices for systematic monitoring, evaluation, and adaptive management of conservation interventions (Pullin & Knight 2003; Ferraro & Pattanayak 2006). Several features of sustainability standards make them both compelling and tractable as subjects for collaborative, systematic evaluation following an evidence-based conservation paradigm. First, sustainability standards present features of both consistency and variation that make them well suited to longitudinal, cross-sectional, and comparative evaluation

designs. Sustainability standards can be considered overall as a similar set of nonstate governance interventions, and individual standards are usually applied consistently due to the codified nature of standards. However, there is considerable variation among agricultural sustainability standards, which presents opportunities to compare the effectiveness of different approaches. Second, sustainability standards have been adopted widely across many contexts, and ongoing expansion presents opportunities to compare adoption and nonadoption trajectories. Third, most standards systems can generate consistent, time-series data through audit processes. Finally, coordination presents strong opportunities to improve efficiency, reduce duplication, and enhance data interpretability and cross-context learning associated with the growing, but now largely separate, set of evaluation efforts emerging from the research and NGO communities.

Although both the need and the opportunity for systematic evaluation and improvement of agricultural sustainability standards are clear, it now remains to turn these possibilities into reality. We propose an approach and operational strategy for doing so.

Key Considerations

Conservation Hypothesis

Conservation interventions may be described according to a theory of change or results chain that logically connects activities to results (i.e., outputs, outcomes, and impacts) (Mascia et al. 2014). Such results-based approaches are now widely used to characterize the causal relationships and assumptions that underpin conservation effectiveness (Margolis et al. 2013).

We recognize 3 main ways by which agricultural sustainability standards may deliver conservation benefits: conserving existing natural ecosystems and their associated biodiversity; improving the conservation value of farms and production landscapes through restoration and conservation-friendly management; and reducing off-site environmental impacts of agriculture, such as water pollution (Fig. 1). Each of these results may be delivered at the scale of individual certified entities or, more broadly, across landscapes and regions through aggregate or indirect effects.

Many of these results are hypothesized to be achieved in a sequential and cumulative manner: direct results such as improved farm management practices support positive changes in the local environment (intermediate results), which in turn contribute to benefits over larger spatial and temporal scales (broader results) (Fig. 1). Sustainability standards also interact reciprocally with other state and nonstate governance approaches to influence standard-setting processes, demand for sustainably

produced products, and the adoption of sustainability standards and their component land use and farm management practices (Steering Committee 2012; Eberlein et al. 2014). This results chain provides an overarching metahypothesis around which specific research and evaluation questions can be designed to interrogate each hypothesized linkage (small right-pointing horizontal arrows in Fig. 1). The stepwise logic supports robust evaluation design by clarifying the intermediate factors that influence the achievement of ultimate impacts (Miteva et al. 2012). The relative importance of each result, and the associated causal linkages, may differ among commodities and contexts.

Specific Demands for Evidence

Efforts to evaluate and improve conservation impacts of sustainability standards will hold greater value if they address key stakeholder priorities. A widely held priority—shared by companies, civil society, governments, agricultural financiers, and consumers—is to document avoidance of the worst environmental practices, such as destruction of mature forests and other ecosystems of high conservation value. Such evidence is needed to demonstrate adherence to various regulations and voluntary commitments, such as the Equator Principles and the Consumer Goods Forum zero net deforestation pledge (Brown & Zarin 2013). Similarly, many companies have made sustainability commitments around issues such as water conservation, pollution mitigation, and increasing on-farm habitat (e.g. Unilever 2010) and require credible monitoring to report on these. Although companies are currently using sustainability standards as a proxy for demonstrating progress, stakeholders are increasingly demanding more direct evidence. Information on conservation outcomes may also inform consumer behavior, although such information may have limited reach unless amplified by marketing campaigns (e.g., to highlight favorable evidence) or activist campaigns (e.g., company scorecards or name-and-shame tactics that shine a spotlight on negative evidence).

Different actors prioritize different evidence in different contexts. For commodities where climate change or land degradation threaten future supply (e.g., cocoa, coffee, and tea), farmers, companies, and industry groups seek information on the condition of natural resources that underpin productive farming systems. Certification for developing country markets is incipient, and stakeholders may emphasize legal compliance, social and environmental management systems, and food safety over demonstrated conservation outcomes (Newton et al. 2014). Especially for crops that can be grown in agroforestry systems, the composition, abundance, and persistence of native species assemblages are of particular interest to conservationists (Tschardt et al. 2014).

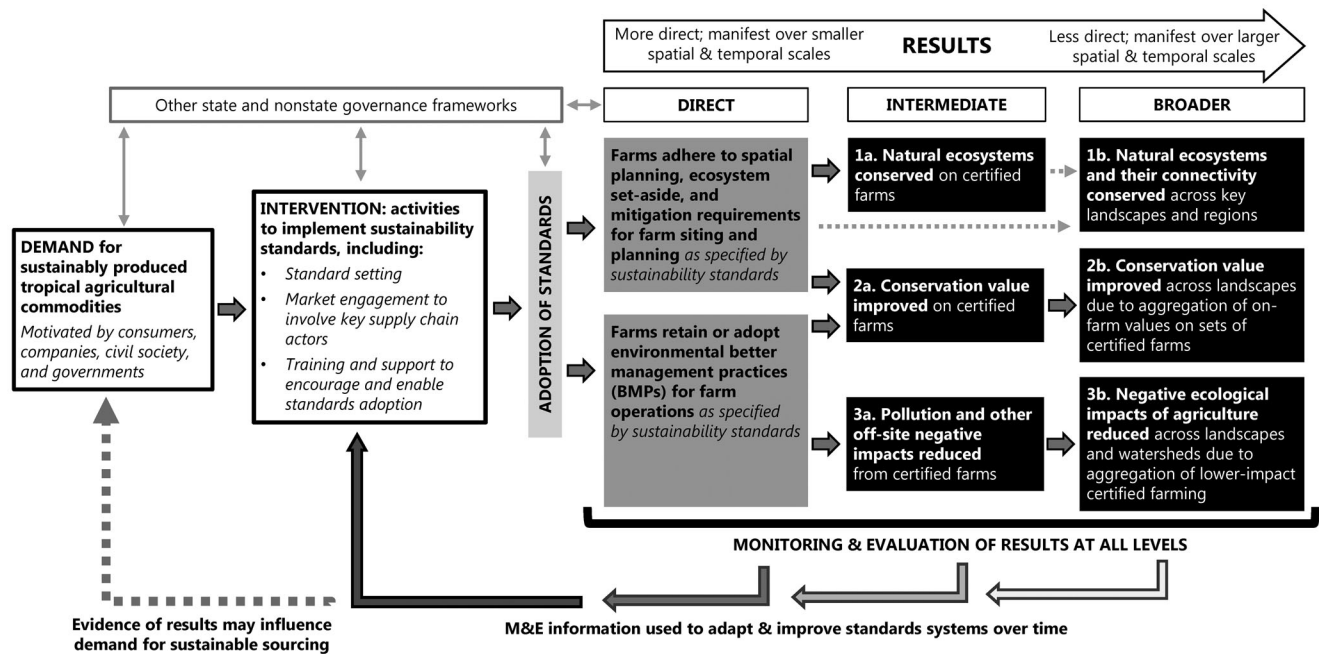


Figure 1. Generalized results chain illustrating the ways in which agricultural sustainability standards are hypothesized to deliver conservation benefits and reduce conservation threats. Black rectangles elaborate the 3 sets of potential conservation benefits identified in the text. Feedback arrows at the bottom indicate that monitoring and evaluation of results at all levels should inform the adaptation and improvement of standards systems over time.

Sustainability standards bodies seek information on impacts to understand and improve the effectiveness of their standards over time; to increase the credibility of sustainability claims; and to assert competitive advantage based on positive results. Finally, donors and international institutions that support sustainability standards seek information on the effects of their investments on key conservation and human development objectives.

Methodological Challenges and Remedies

To develop a robust assessment framework, it is important to recognize key limitations that have characterized most evaluations of sustainability standards to date (Table 1). Many such challenges can be largely addressed through improved design and execution of future monitoring and research (Table 1).

Other limitations are more difficult to address and may continue to impede the interpretation of evaluation findings. One challenge is to distinguish the role of sustainability standards in improving practices and outcomes versus simply recognizing or rewarding existing good management. If the most sustainable producers adopt sustainability standards whereas poorer performers avoid them, then the incremental benefit (additionality) may be limited. However, this issue requires a nuanced lens because in some cases simply maintaining good practices in the face of pressures to expand or intensify agriculture

may constitute success. In other cases, producers make major improvements in preparation for certification, but their results are not captured if precertification baselines have not been established. An additional challenge is that each sustainability standard is actually a bundle of interventions (e.g., specific requirements of the standard) that may differ by context and change over time as standards are revised. Accordingly, when evaluating sustainability standards, it is important to select the appropriate independent variable or variables and understand the extent to which evaluation findings may be legitimately generalized across standards, contexts, and time.

Framework for Assessing Conservation Impacts of Agricultural Sustainability Standards

Based on the preceding considerations, we devised a framework for a coordinated effort to monitor, evaluate, and adaptively manage the conservation impacts of agricultural sustainability standards. The framework is oriented around 8 research questions (Table 2) designed to test the causal logic and associated hypotheses in Fig. 1. These questions are aligned with many of the top 100 important questions for the future of agriculture (Pretty et al. 2010) and biodiversity conservation (Sutherland et al. 2009), thereby linking evaluation of agricultural sustainability standards to these broader agendas. The

Table 1. Key limitations of the existing evidence base on impacts of agricultural sustainability standards and corresponding recommendations for future activities or approaches that could help improve the evidence base.

<i>Limitations of existing approaches and evidence base</i>	<i>Needs and recommendations for generating a more robust evidence base</i>
Evaluation literature is focused on isolated case study investigations limited to a few crops and geographies.	Apply a systematic and adequately dense sampling framework representing key crops and geographies.
Different studies use different methods, preventing comparative analysis: different treatment variables (interventions) different outcome variables (indicators) different, or unaccounted for, control variables (covariates) different data collection protocols.	Apply standard methods and indicators to facilitate generalization and comparison across contexts.
Studies evaluate only one dimension of sustainability (e.g., social, environmental, or economic).	Evaluate multiple dimensions in a given system to understand sustainability outcomes more holistically, including the realization of synergies or trade-offs.
Studies lack counterfactual comparisons.	Construct credible counterfactuals using experimental or quasi-experimental methods.
Short time frame of studies fails to reveal the trajectory of ecological and other outcomes, especially for slowly changing variables.	Conduct repeated monitoring of key variables to generate robust time series and panel data (including for counterfactual scenarios).
Scale of evaluations (e.g., certified farms) is not matched to the scale of key conservation outcomes (e.g., landscape, regional, or global).	Complement site-level evaluations with analysis at larger spatial and temporal scales to evaluate cumulative effects and interactions of farms with surrounding areas.
Studies lack transparency, are not conducted to high research standards, or are not appropriately interpreted relative to the aims and mechanisms of sustainability standards.	Use theory-based evaluation approaches; adhere to evaluation best practices; improve collaboration between researchers and standards bodies; and publish results in peer-reviewed journals.
Studies on environmental results focus on changes in practices, rather than outcomes, or use indicators of low credibility (e.g., farmer perception).	Select and quantitatively monitor changes in key environmental parameters that sustainability standards are hypothesized to influence.

questions are also defined to address stakeholder demands for evidence (identified above).

Research questions are posed in Table 2 in a generalized way; researchers would further specify these questions for specific evaluation studies. Individual evaluations will often focus on a specific sustainability standard or on a subset of the practices or outcomes required under that standard. For instance, question 3 in Table 2, related to the protection of on-farm conservation values, would be approached quite differently for a coffee standard that promotes diversified agroforestry systems than for a sugar standard that promotes efficient production monocultures with conservation set-aside areas. Collectively, the body of evidence from many such individual evaluations can clarify which practices (i.e., requirements of different sustainability standards) are most effective at delivering conservation benefits in different contexts. This information can elucidate the relative effectiveness of different standards at meeting different objectives and, more importantly, support all standard systems in becoming more effective over time. In addition to addressing the effects of individual sustainability standards or their component practices, the research questions also consider the cumulative and interactive effects of sustainability standards applied broadly across a sector or region, rela-

tive to the biophysical, spatial, and governance contexts.

To address these questions will require a monitoring and evaluation approach that combines breadth, depth, and integrative analysis. We therefore propose a hierarchical, nested framework that integrates data and learning at 3 levels (Fig. 2). At the broadest level (Fig. 2, level 1), system-wide monitoring across most or all of the certified estate is necessary to characterize sustainability standards at a basic level, understand variation across contexts, quantify aggregate results, and contextualize place-specific research. To move beyond basic characterizations and understand intermediate conservation outcomes (Fig. 1) requires more in-depth field monitoring that is too costly to conduct for every certified entity. Such sampled monitoring (Fig. 2, level 2) should be conducted across a stratified sample of sites to understand the effects of sustainability standards in different settings and thereby support more generalized conclusions and reporting of aggregate outcomes. Finally, focused research (Fig. 2, level 3) is needed to improve the credibility of evaluation findings, address existing methodological limitations (Table 1), and evaluate broader and indirect effects associated with sustainability standards. Below we describe each level of the framework, provide examples, and identify the most critical next steps.

Table 2. Key questions to be addressed by a coordinated approach to monitor and evaluate conservation impacts of agricultural sustainability standards.

<i>Research question^a</i>	<i>Key themes^b</i>	<i>Scope of investigation^c</i>	<i>References^d</i>
1. To what extent do sustainability standards lead to changes in on-farm environmental management practices?	soil and water, noncrop vegetation, agrochemicals, farm land use	certified lands	A79, A100, C95
2. Do sustainability standards help maintain or improve the on-farm natural resource base and associated ecosystem services to agriculture?	soil structure and fertility, water, pest control, climate resilience, pollination	certified lands	A8, A21
3. Do sustainability standards help maintain or improve on-farm biodiversity conservation values?	noncrop vegetation, shade canopy, riparian protection, conservation set asides	certified lands	A32, C41
4. When and to what extent are conservation outcomes synergistic or conflicting with productivity and profitability outcomes?	above themes combined with key economic and productivity results	certified lands	A15, A66, A67, A92, C40, C41
5. Do sustainability standards contribute to biodiversity-friendly landscapes and land-use mosaics?	habitat connectivity, natural ecosystem conservation and condition, population viability	certified lands in context	A19, C37, C40
6. To what extent do sustainability standards mitigate negative off-site environmental impacts or support ecosystem service flows to off-site beneficiaries?	water quality, water flow regulation, sedimentation, eutrophication, greenhouse gases	certified lands in context	A21
7. Do sustainability standards help reduce destruction and degradation of natural ecosystems and areas of biodiversity importance at landscape to regional scales?	protection of key areas	certified lands in context	C95
8. How do sustainability standards interact with other land use, governance, and market incentive systems and dynamics to influence the answers to the preceding questions?	governance, land-use planning and regulation, market incentives, agricultural supply chains	sustainability standards in context	

^a *Research questions about the effects of sustainability standards can be answered at multiple levels: effects of implementing specific requirements of one or more sustainability standards, such as the adoption of specific conservation-friendly management practices; effects of adopting a specific sustainability standard; or uptake of sustainability standards and certification more generally across a sector or locale. The research questions should be construed as including all these aspects, although the salience of each aspect differs from question to question.*

^b *Further suggested foci for each research question.*

^c *This column identifies whether each question needs to be answered by examining certified lands alone; certified lands within their biophysical, economic, or sociopolitical contexts; or entire sustainability standards systems within such contexts.*

^d *This column links the research questions to previously identified top 100 important questions for the future of global agriculture (questions with an A [Pretty et al. 2010]) and biodiversity conservation (questions with a C [Sutherland et al. 2009]).*

System-Wide Monitoring

System-wide monitoring (level 1) comprises data collected for every certified entity under a given sustainability standard. Such data collection is typically integrated into certification processes to achieve full data coverage. However, this means that level 1 data are limited to those that can be feasibly collected through certification applications (i.e., information self-reported by producers) and audits (which verify self-reported data and collect additional information to assess conformance with a stan-

dard). Key level 1 data include the location and land area of certified entities, area under production and conservation land uses, crop types and production systems, when and for how long certificate holders achieve and maintain certification, and whether a certified entity has multiple certifications. Such data enable basic characterization of certified lands and are critical for informing secondary analysis or research on broader impacts. For example, contributions to landscape conservation cannot be understood without knowing the location of certified entities.

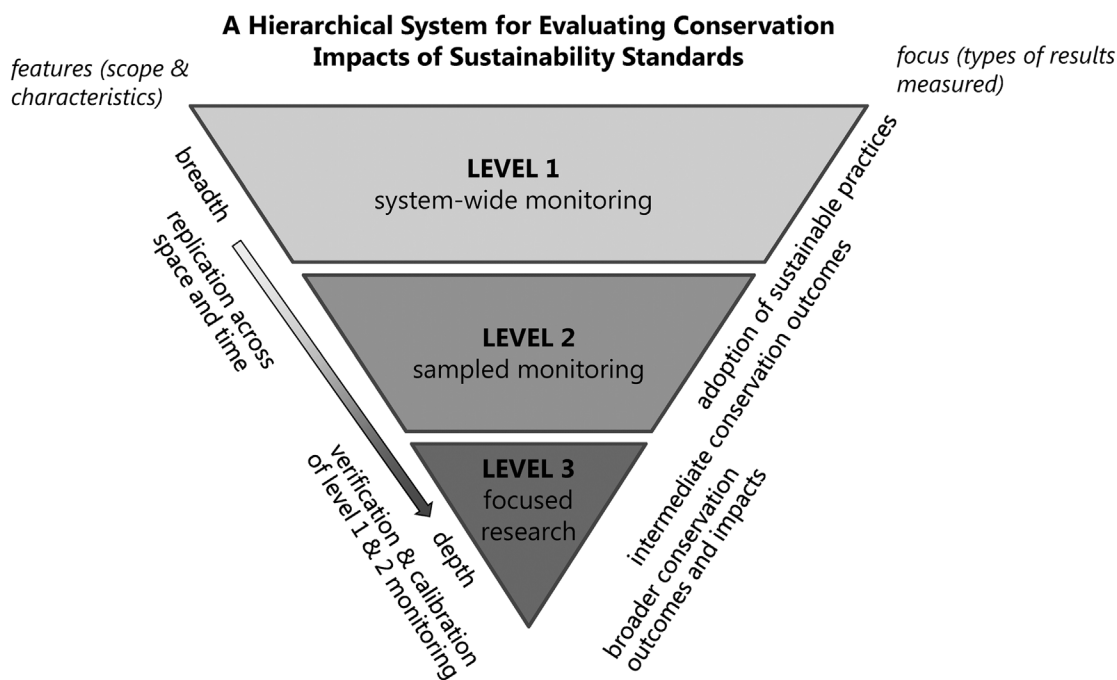


Figure 2. Proposed framework for a coordinated system to evaluate the conservation impacts of agricultural sustainability standards. The system aligns evaluation themes, indicators, and methods to integrate data collected system-wide (i.e., for all certified entities; level 1), data collected using comparable methods at a representative sample of certified operations (level 2), and focused research conducted using experimental and quasi-experimental designs as well as methods to understand broader contributions to, and indirect effects on, conservation outcomes (level 3). Each level is explained further in the text.

System-wide monitoring can also be used to collect data on farm performance relative to specific practices or outcomes identified in a standard, to the extent that auditors evaluate these parameters to assess conformance. For example, Starbucks' Coffee and Farmer Equity (C.A.F.E.) Practices sustainability standard and verification program has used audit data to track farm practices over several years, thereby informing adaptive management by highlighting potential improvements to the standard and priorities for farmer training (Thomas et al. 2012). In sustainability standards that are strictly practice based, audit data would be restricted to information on practice adoption. However, some standards also include outcome-based requirements. For example, compliance to the Bonsucro sugar standard considers outcome-level data on greenhouse gas emissions, water use efficiency, and chemical oxygen demand of agricultural runoff, among others (Bonsucro 2011).

At level 1, the priority is to achieve complete data coverage, consistency, and comparability for a core minimum data set that is feasible to collect through certification processes and that spans different sustainability standards and contexts. This could be achieved by defining common indicators and guidelines for data collection to facilitate data comparability. If data are to be aggregated across multiple sustainability standards and made

available to the research community (as we recommend), policies on data confidentiality, spatial precision (e.g., blurring farm location data to protect individual farmer identity while still enabling spatial analysis), and terms of use will also be needed.

Many processes necessary for the establishment of a minimum data set are already underway. For example, the ISEAL Alliance (a membership organization for sustainability standards) requires its members to develop credible monitoring and evaluation systems and encourages the collection of basic common information, such as the location of certified entities (ISEAL Alliance 2010). However, additional work is needed to better incorporate conservation-related indicators, support all major standards systems to participate in a common level 1 data collection effort, and define mutually agreeable policies for data quality, confidentiality, and sharing of data with the research community. These efforts will help redefine the certification audit as not solely a conformance assessment tool but also one that facilitates monitoring and evaluation.

Sampled Monitoring

Effective monitoring of intermediate conservation outcomes (level 2) usually requires going beyond the scope

of certification audits. Given resource constraints, this is best achieved through a stratified sampling approach designed to represent key axes of variability most likely to affect conservation outcomes, such as crop type, geographic setting, local regulatory context, identity of the sustainability standards applied, and type of production system (e.g., plantation vs. smallholder group). This approach recognizes that outcomes are likely to vary by context; therefore, data on multiple certified entities per combination of context variables can help elucidate effectiveness and the factors that influence it. Level 1 data provide a basis for establishing stratified samples.

Sampled monitoring may assess indicators such as land-cover composition, tree diversity and vegetation structure, presence of natural ecosystem set asides and restoration zones, and water quality of streams within certified units. Depending on the conservation values of interest, it may be more appropriate to monitor change through the use of proxy variables (e.g., habitat quality metrics) than direct measures (e.g., populations of species of concern). Pairing sampled monitoring of intermediate results with system-wide monitoring of direct results (e.g., practice adoption) is especially valuable for testing the intermediate hypothesized linkages within the results chain.

Sampled monitoring conducted on certified properties without control sites is useful for tracking change trajectories and generating suggestive evidence of the effects of certification. Adding control groups can provide greater confidence about the causality of any observed effects. With or without a control group, sampled monitoring should be conducted on a time-series basis, for instance, at baseline and every 2–3 years thereafter. This interval is generally suitable to detect changes associated with initial adoption of standards, subsequent improvements or regressions, and the effects of periodic standard revisions. Monitoring can be conducted by auditors who have received additional training on the relevant field methods (as a supplement to the audit process itself) or by local researchers, students, or paraprofessionals.

Several initiatives have defined sets of environmental indicators at the sampled monitoring level, including work by the Committee on Sustainability Assessment (COSA 2014); the SAI Platform, representing the food industry (Elferink et al. 2012); and the Zoological Society of London, focusing on high conservation value areas in the context of oil palm production (Zrust et al. 2013). In addition, several calculator tools are available that model environmental outcomes by integrating farmer-provided data on land management practices with environmental context data. For instance, the Field to Market tool estimates outcomes related to water pollution, greenhouse gases, and biodiversity, which can be analyzed at the farm level or aggregated across supply chains or watersheds (Field to Market 2013). Each of these initiatives offers useful tools and insights, but none has yet been applied

widely to evaluate intermediate conservation outcomes of sustainability standards.

Two sets of activities are needed to advance sampled monitoring. First, data collectors and data users need to move toward sets of common indicators and methods for tracking key conservation results. Doing so will increase data comparability and facilitate comparative analysis and meta-analysis. The indicator sets and tools mentioned above provide helpful starting points but may need to be expanded, customized, or streamlined to fit the specific contexts where sustainability standards are applied. Given differing interests of different stakeholders and the need for context-sensitive indicators, some metrics might be widely standardized whereas others will need to be adapted to specific crops or contexts.

Second, sampled monitoring needs to be operationalized, which may best be achieved through initiatives focused on regions where commodity production is a key conservation threat. These efforts would address specific questions nested within the overall framework (Fig. 1, Table 2), for instance, do sustainability standards applied to cocoa-producing areas in West Africa's Guinean Forests biodiversity hotspot support sustainable crop intensification and reduced deforestation and forest degradation? In this example, multiple sustainability standards working in cocoa in West Africa would conduct sampled monitoring at a network of sites to evaluate effects of their efforts, both individually and collectively. Monitoring could be conducted by local or regional institutions (e.g., universities) and coordinated under the auspices of research teams or grant-funded projects. In this way, impacts can be understood beyond the level of individual certified farms and linked to conservation priorities at larger scales.

Focused Research

Level 3 includes 2 types of focused research that require more specialized or in-depth methodologies than are typically possible at levels 1 and 2. First, to overcome methodological limitations (Table 1), there is a need for research that uses rigorous program evaluation techniques. Such techniques measure causal effects by comparing the performance of a group subject to an intervention (e.g., sustainability standards) to a counterfactual scenario, that is, a credible approximation of performance absent the intervention (Blackman & Rivera 2011). Counterfactuals may be established through experimental methods (randomized controlled trials) in which producers are randomly assigned to either a treatment group, which adopts sustainability standards, or a control group, which does not. However, this approach is uncommon due to logistical and ethical challenges of intentionally excluding some producers. It is more common to mimic an experimental design by allowing producers to interact with sustainability standards in the usual way and then

Table 3. Proposed initiatives to implement the coordinated approach for assessing and improving conservation outcomes of agricultural sustainability standards.

<i>Initiative or activity^a</i>	<i>Tangible products or consequences of this activity^b</i>	<i>Key groups to be involved</i>
Develop common indicators, data collection guidelines, and data policies for a core minimum data set for system-wide monitoring (level 1).	foundation for a collaborative system established based on best-practice monitoring approaches; conservation-related indicators effectively incorporated	standards bodies, convening body such as ISEAL Alliance, monitoring experts, funders
Support standards systems to implement system-wide monitoring (level 1).	time-series data on basic characteristics of all certified entities collected according to a common monitoring framework	standards bodies, convening body such as ISEAL Alliance, funders
Establish a data sharing platform (levels 1, 2, 3).	data management facility and functions (including clear data sharing policies and terms of use) that support use of data for analysis and research	standards bodies, neutral organization to host platform, data systems experts and other advisors, funders
Develop common indicators and metrics for sampled monitoring of conservation outcomes (level 2).	framework and field protocols for carrying out coordinated monitoring of key conservation results	data collection organizations (e.g., universities or research centers within focal regions), standards bodies, companies, researchers, governments, funders
Establish regional sampled monitoring initiatives in key regions where commodity production drives conservation threats (level 2).	comparable panel data from stratified samples of certified properties in regions where commodity production drives key conservation threats	data collection organizations, standards bodies, companies, researchers, governments, funders
Establish a research network to define a research agenda, coordinate research activities, and synthesize and communicate results (level 3).	research agenda guides project selection and design; data and knowledge sharing facilitates collaborative research and meta-analysis; new evidence supports adaptive management and continuous improvement of conservation benefits	researchers, standards bodies, convening organizations for the network, other groups interested in a stronger evidence base (e.g., civil society), funders

^aInitially, these activities could be developed and launched in a modular fashion through the groups indicated in the right column. Over time, the activities would coalesce into the integrated 3-tiered assessment approach described in the text.

^bSpecific products or results of value to “key groups to be involved” and to society at large that each activity will deliver.

use statistical techniques to identify a control group that is similar to the treatment group but did not adopt the standards. Under either approach, the control group’s average performance constitutes a credible counterfactual against which the treatment group is compared.

Second, level 3 research is needed to understand conservation impacts over larger spatial and temporal scales (i.e., results 1b, 2b, and 3b in Fig. 1) to which sustainability standards may contribute only partially or indirectly. Doing so requires analyzing certified lands in their spatial, biogeographical, political, institutional, and market contexts. For example, understanding impacts on land-cover change or species’ populations may require disentangling the effects of sustainability standards from those of other landscape dynamics or broader trends, such as climate change. Spatial analysis can clarify contributions of certified properties to landscape connectivity or conservation networks and, when paired with statistical methods, reveal whether sustainability standards are helping to arrest threats such as forest degradation or peatland burning. Level 1 data, such as the spatial location and basic characteristics of certified units, provide

important input to such research. Focused research is typically best conducted by independent researchers who possess specialized methodological expertise and whose work is less likely to be perceived as biased than monitoring associated with standards bodies or supply chain actors.

The priority at level 3 is not only to increase the amount of relevant research on sustainability standards, but also to target, align, and use this research more effectively to improve conservation benefits. Doing so requires 3 sets of activities. First, a priority research agenda should be codefined by researchers, sustainability standards organizations, and other stakeholders. As an example, the Roundtable for Sustainable Palm Oil (RSPO) and Royal Society recently collaborated to define a coordinated 5-year multidisciplinary research program on the impacts of RSPO certification, the results of which will support efforts to evaluate and improve RSPO’s sustainability standard (SEARRP 2012). Second, data management functions should be established within a neutral, trusted organization to archive level 1 and 2 monitoring data and make it available to researchers for legitimate purposes. Third,

an active network for research on conservation impacts of sustainability standards should be established to support researchers in aligning their work with the overall evaluation framework articulated in this article, facilitate meta-analysis and other data synthesis, and communicate research results more effectively to standards bodies and other stakeholders.

From Concept to Action

Experience to date suggests that coordinated monitoring, evaluation, and data management systems are certainly possible but are challenging to develop because they require agreement of many people and institutions regarding shared goals and processes, while addressing issues such as data management and access, confidentiality, and funding (Wollenberg et al. 2007; COSA 2014). Coordinated efforts are most likely to succeed if they advance in a modular fashion, leading with discrete initiatives that offer clear benefits and synergies for specific groups, such as precompetitive collaboration on common indicators or increased monitoring efficiency by consolidating duplicative effort in specific regions. Table 3 identifies a set of initiatives that should be feasible to launch over the next few years and that could coalesce into the fully formed 3-tiered approach advocated here.

New resources will be needed to deliver stronger evidence on the conservation impacts of sustainability standards. There is now a growing business case for such investment, as companies require new monitoring to demonstrate fulfillment of sustainability commitments and as sustainability standards and supply chain actors seek to leverage the value of existing data (e.g., certification audits and farmer record keeping) to better understand and improve supply chain performance. However, not all outcomes of importance for global biodiversity conservation are of high interest to the private sector—particularly those that manifest across wider landscapes or have less direct impact on future commodity supply. Rather, these are public goods that require government and donor support to safeguard—in this case, by contributing to the generation of a robust evidence base to assess and continuously improve such outcomes. Such support is strategic and critical: comparatively modest investment in monitoring systems for sustainability standards could help solidify and expand the role of such standards in leveraging market forces to deliver conservation benefits, whereas failure to invest could undermine the continued viability of such mechanisms.

Conclusions

Agricultural sustainability standards have gone well beyond niche status and now govern substantial propor-

tions of key tropical commodities implicated in biodiversity loss. But despite the increasing reach of sustainability standards, relatively little is known about their conservation impacts. Sustainability standards may therefore now be at a fork in the road. If robust evidence on results and mechanisms can be generated—thereby substantiating claims, identifying limitations, and supporting continuous improvement—then sustainability standards could become increasingly potent vehicles to translate public demand and corporate commitments for sustainability into real conservation benefits. However, if there remains little evidence to substantiate (or refute) standards' implied sustainability claims, then widespread skepticism and loss of credibility could follow. In this scenario, sustainability marketing campaigns (i.e., greenwashing) could crowd out the more transparent and independent standard-setting and auditing processes that are the hallmarks of third-party sustainability standards.

Given the potentially pivotal role of sustainability standards in addressing stubborn social and environmental problems, the development of a robust evidence base related to these standards must be considered as among the highest investment priorities for increasing the sustainability of tropical agriculture. We have demonstrated here that it is both possible and necessary to deliver such evidence by better coordinating a range of existing monitoring and research activities and supplementing these with targeted new initiatives. The private sector and sustainability standards bodies can support this work to a considerable degree, building on current efforts. However, in view of the public good value of benefits that sustainability standards may deliver, donors and governments must also invest in the supportive infrastructure—particularly robust monitoring systems—that is needed to leverage societal and corporate demand for sustainability to deliver biodiversity conservation and other benefits at full scale.

Literature Cited

- Blackman, A., and J. Rivera. 2011. Producer-level benefits of sustainability certification. *Conservation Biology* 25:1176–1185.
- Bonsucro. 2011. Bonsucro production standard, version 3.0. Available from <http://bonsucro.com/site/production-standard> (accessed December 2013).
- Brown, S., and D. Zarin. 2013. What does zero deforestation mean? *Science* 342:805–807.
- Cashore, B., G. Auld, and D. Newsom. 2004. *Governing through markets: forest certification and the emergence of non-state authority*. Yale University Press, New Haven, Connecticut.
- COSA (Committee on Sustainability Assessment). 2014. *The COSA measuring sustainability report: coffee and cocoa in 12 countries*. COSA, Philadelphia.
- Eberlein, B., K. W. Abbott, J. Black, E. Meidinger, and S. Wood. 2014. Transnational business governance interactions: conceptualization and framework for analysis. *Regulation & Governance* 8:1–21.
- Elferink, E., G. Kuneman, A. Visser, and E. van der Wal. 2012. Sustainability performance assessment of farming practices: guidelines

- for developers of quantitative monitoring tools. Available from <http://www.saipatform.org/activities/alias/SPA> (accessed December 2013).
- Ferraro, P. J., and S. K. Pattanayak. 2006. Money for nothing? A call for empirical evaluation of biodiversity conservation investments. *PLoS Biology* 4 DOI: 10.1371/journal.pbio.0040105.
- Field to Market. 2013. Fieldprint calculator. Available from <http://www.fieldtomarket.org/fieldprint-calculator> (accessed December 2013).
- Foley, J. A., et al. 2011. Solutions for a cultivated planet. *Nature* 478:337–342.
- Garnett, T., et al. 2013. Sustainable intensification in agriculture: premises and policies. *Science* 341:33–34.
- Giovannucci, D., O. von Hagen, and J. Wozniak. 2014. Corporate social responsibility and the role of voluntary sustainability standards. Pages 359–384 in C. Schmitz-Hoffmann, M. Schmidt, B. Hansmann, and D. Palekhov, editors. *Voluntary standards systems—a contribution to sustainable development*. Springer, Berlin.
- ISEAL Alliance. 2010. Assessing the impacts of social and environmental standards systems v1.0: ISEAL code of good practice. ISEAL Alliance, London.
- Laurance W. F., J. Sayer, and K. G. Cassman. 2014. Agricultural expansion and its impacts on tropical nature. *Trends in Ecology & Evolution* 29:107–116.
- Levin, J. 2012. Profitability and sustainability in palm oil production: analysis of incremental financial costs and benefits of RSPO compliance. WWF, Washington, D.C.
- Margoluis, R., C. Stem, V. Swaminathan, M. Brown, A. Johnson, G. Placci, N. Salafsky, and I. Tilders. 2013. Results chains: a tool for conservation action design, management, and evaluation. *Ecology and Society* 18:22. Available from: <http://www.ecologyandsociety.org/vol18/iss3/art22/>.
- Mascia, M., S. Pailler, M. L. Thieme, A. Rowe, M. C. Bottrill, F. Danielsen, J. Geldmann, R. Naidoo, A. S. Pullin, and N. D. Burgess. 2014. Commonalities and complementarities among approaches to conservation monitoring and evaluation. *Biological Conservation* 169:258–267.
- Miteva, D. A., S. K. Pattanayak, and P. J. Ferraro. 2012. Evaluation of biodiversity policy instruments: What works and what doesn't? *Oxford Review of Economic Policy* 28:69–92.
- Newton P., A. Agrawal, and L. Wollenberg. 2013. Enhancing the sustainability of commodity supply chains in tropical forest and agricultural landscapes. *Global Environmental Change* 23:1761–1772.
- Newton, P., H. N. Alves-Pinto, and L. F. Guedes Pinto. 2014. Certification, forest conservation, and cattle: theories and evidence of change in Brazil. *Conservation Letters* DOI:10.1111/conl.12116.
- Potts, J., M. Lynch, A. Wilkings, G. Huppe, M. Cunningham, and V. Voora. 2014. The state of sustainability initiatives review 2014. International Institute for Sustainable Development, Winnipeg and International Institute for Environment and Development, London.
- Pretty, J., et al. 2010. The top 100 questions of importance to the future of global agriculture. *International Journal of Agricultural Sustainability* 8:219–236.
- Pullin, A. S., and T. M. Knight. 2003. Support for decision making in conservation practice: an evidence-based approach. *Journal for Nature Conservation* 90:83–90.
- SEARRP (South East Asia Rainforest Research Programme). 2012. SENSOR: an integrated multi-disciplinary research programme for sustainability. Available from <http://www.searrp.org/sensor> (accessed February 2014).
- Steering Committee of the State-of-Knowledge Assessment of Standards and Certification. 2012. *Toward sustainability: the roles and limitations of certification*. RESOLVE, Washington, D.C.
- Sutherland, W. J., et al. 2009. One hundred questions of importance to the conservation of global biological diversity. *Conservation Biology* 23:557–567.
- Thomas, M., E. Baer, B. Semroc, and J. Sonenshine. 2012. C.A.F.E. practices results assessment, fiscal years 2011–2012. Conservation International, Arlington, Virginia.
- Tscharntke, T., J. C. Milder, G. Schroth, Y. Clough, F. DeClerck, A. Waldron, R. Rice, and J. Ghazoul. 2014. Conserving biodiversity through certification of tropical agroforestry crops at local and landscape scales. *Conservation Letters* DOI:10.1111/conl.12110.
- UNEP-WCMC. 2011. Review of the biodiversity requirements of standards and certification schemes. CBD Technical Series No. 63. Secretariat of the Convention on Biological Diversity, Montreal.
- Unilever. 2010. Unilever Sustainable Agriculture code. Available from <http://www.unilever.com/aboutus/supplier/sustainablesourcing> (accessed December 2013).
- Wollenberg, E., L. Merino, A. Agrawal, and E. Ostrom. 2007. Fourteen years of monitoring community-managed forests: learning from IFRI's experience. *International Forestry Review* 9:670–684.
- Zrust, M., et al. 2013. HCV threat monitoring protocol. Zoological Society of London, London.

