



Lessons learned for REDD+ from evaluations of GHG statements

Till Neeff and Donna Lee
For the Global Forest Observations Initiative

Cover photo: ©FAO/Hoang Dinh Nam

Foreword

The Global Forest Observations Initiative (GFOI) is an informal partnership to coordinate international support to developing countries on forest monitoring and greenhouse gas (GHG) accounting for the purposes of REDD+ and related fora. By providing a larger and more targeted package of support than any one partner can provide alone, GFOI collaboration seeks to help developing countries accelerate their progress in these areas.

Through the collaborative action of its partners, GFOI seeks to cover the breadth and depth of the tasks required to establish National Forest Monitoring Systems (NFMS) to underpin Measurement, Reporting and Verification (MRV) of GHGs and mitigation results. GFOI facilitates exchanges of information, sharing of resources, south-south collaboration and enables comparative advantages. The Initiative fosters a large and diverse network of practitioners to help address challenges, knowledge gaps and enable progress.

In recent years, many developing countries have made strong progress in the development of their NFMS and capabilities for GHG measurement and reporting, but knowledge of the 'V' in MRV remains quite low. This is likely because, to date, many countries have been focused on REDD+ readiness and implementation of REDD+ strategies and action plans. However, some countries are starting to move towards reporting their emissions reduction results, increasing the relevance of knowledge building on this topic.

We hope that improving the understanding of GHG evaluations will ultimately help to improve measurement and reporting processes and hence the broader MRV cycle. Furthermore, an increased understanding of GHG evaluations presents a unique opportunity for targeted capacity building where evaluations can serve as the basis for identifying data gaps, defining training needs and improving national processes. Evaluation processes can also connect closely to continuous improvement plans and become a core pillar in further developing MRV capacities.

We thank the lead authors for their work in driving this study and express our gratitude to the many partners who have made valued contributions and supported the report's design and execution. The study has drawn upon the significant experience and expertise of the GFOI network to ensure its findings represent a broad perspective of views. We hope that "Lessons learned for REDD+ from evaluations of GHG statements" can contribute to an improved understanding of the processes and approaches used for evaluating GHG accounts in the forest sector.

GFOI Leads Group, 26 January 2018

Acknowledgements

The authors would like to express their gratitude to the following people with deep understanding on the different evaluation schemes for having shared their expert knowledge on evaluations of REDD+ results or other types of quantified GHG statements for this report. We gratefully acknowledge their contribution, while noting that listing their name below does not imply agreement with the analyses in the report, which is the sole responsibility of the authors.

Andrés Espejo
Anthony Bennie
Carly Green
Chris Dragisic
Dan Doherty
Edwin Aalders
Evan Notman
Fred Stolle
Inge Jonckheere
Leticia Guimaraes
Maarten van der Eynden

Marco van der Linden
Marieke Sandker
Marius von Essen
Martin Herold
Naomi Swickard
Jan-Willem Martens
Pradeepa Bholanaath
Sam Hoffer
Sandro Federici
Sylvia Wilson
Thomas Harvey

Funding for this report has been provided by the Global Forest Observations Initiative. The report represents the views of the authors only, it does not necessarily represent the views of the GFOI partners.

Executive summary

Third-party evaluations of GHG statements have long been tools to build trust in statements made by GHG mitigation programs or efforts and to enable their use—against international commitments or voluntary targets—often with significant financial implications. Such evaluations can also provide feedback that enables learning and an opportunity for capacity building to improve GHG measurement and reporting. In meeting these objectives, there can be tension between evaluations designed to underpin the fungibility of GHG statements and those designed to retain flexibility to address varying country circumstances. This paper focuses on evaluations of GHG statements with a view to providing lessons for REDD+, as countries complete REDD+ readiness and move towards achieving results.

This paper is a comparative analysis of the ways in which GHG statements are evaluated. Such evaluations occur in various contexts ranging from technical assessments, whose key objective is to build measurement and reporting capacity while recognizing country circumstances, to verifications of GHG mitigation statements that are used to issue fungible credits that enter carbon markets. What we find is that *the objectives* largely drive a range of different characteristics found in evaluations of GHG statements. Moreover, *who* is designing and conducting the evaluation may also have an impact on its mandate and process—e.g. whether the evaluation emanates from a multilateral, country-driven venue or whether it was designed by an independent entity for the private sector. Understanding such contexts provides the backdrop to understanding why evaluations differ.

While evaluation processes tend to have similar procedural elements, their dynamics and the implications of the evaluation report can be quite different. Evaluations also use different terminologies; for example, under the UNFCCC, ‘reviews’ under the Kyoto Protocol can lead to adjustment of the estimated emission reduction, while ‘technical assessments’ for REDD+ reference levels provide capacity building recommendations. Some evaluations underlie funding decisions, while others feed into compliance systems (either generating offsets or verifying the achievement of targets). These differences can impact, for example, the extent to which an evaluation makes a pass-fail judgment or has the mandate to request corrective actions to the GHG report (e.g. to meet a target, issue a credit, or receive funding). By comparison, other types of evaluations aim to encourage and support countries in stepwise improvements.

Evaluation approaches also vary and lead to different evaluation reports. Some evaluations aim to provide a clear level of assurance on whether an emission reduction statement is correct. These typically employ the concept of materiality and ensure the evaluation is focused on areas that have the highest risk to the quantification of emission reductions. Such methods typically check primary data and issue summary statements that include verification of quantified amounts. By comparison, other evaluations (especially those focused on building capacity) are more qualitative in nature. Those that include evaluation of institutional arrangements, systems, and processes and that include in-country visits tend to provide stronger capacity building opportunities.

Who conducts the evaluation, how they are managed, and how much guidance is provided to evaluators also impacts outcomes of the evaluation. Auditing firms are driven by different interests than expert panels (typically used in multilateral settings): they typically are liable for the correctness of the evaluation and seek to avoid conflicts of interest. Such firms often use the International Organization for Standardization (ISO) GHG validation and verification standard and have in place quality management systems, compared to expert panels with less standardized approaches, where quality management can be more challenging. The level of guidance provided—both for how to measure and report emission reductions, as well as for how to conduct the evaluation—also varies

significantly among the different schemes we analyzed. We found that this can impact the comparability and quality of the evaluations, but that flexibility can be useful where the main objective is capacity building.

Our report collects lessons learned on what kinds of evaluations provide stronger capacity building opportunities compared to those that lead to stronger assurances of GHG statements and also offers considerations for the future:

Evaluations can help improve the quality of GHG reports over time, especially when they are part of an iterative process and when the evaluations focus on both systems and data. They are an underused opportunity for capacity building on GHG measurement and reporting, both for those who receive feedback and for those who provide it. Evaluations that assess not only data and information, but also the systems, processes and institutions that generate the data are the most useful for building capacity. Evaluations that are qualitative and flexible (i.e. likely providing less guidance) may also be more effective for capacity building—allowing experts to adjust recommendations based on national circumstances. Finally, evaluations are mostly likely to improve the quality of GHG reports over time when they are part of an iterative process with provisions for updating data and methodologies in response to evaluation reports.

The design of the evaluations can have an impact on the transparency, accuracy, consistency and, consequently, comparability of GHG statements; strengthening the assurance of GHG statements requires designing the evaluation accordingly. Incorporating materiality thresholds and providing a level of assurance is critical to the understanding of the accuracy of emission reduction statements. Risk-based approaches also can provide stronger clarity on the accuracy of statements (and lead to more accurate statements over time) compared to broad-based evaluations. Clear and detailed guidance and criteria, as well as quality management procedures, are also important for providing such assurances, but also for ensuring there is a level of comparability among GHG statements and consistency among evaluations. Finally, evaluations that are highly transparent not only avoid conflicts of interest, but also build trust on emission reduction statements.

Evaluations can be designed to maximize learning and ensure progress towards GHG reduction targets—over time, such objectives may need to converge in the processes surrounding the Paris Agreement. In the near term, there is a need to strike a balance between clarity and detail of guidance and criteria with flexibility to accommodate learning in the early stages. However, over time, there may be a need to increase alignment among schemes intending to build capacity and those designed to create fungibility among emission reduction units—particularly for those countries with quantified Nationally Determined Contributions or that wish to participate in trading of mitigation obligations (under Article 6) or other emerging market mechanisms. However, in the near term, a priority should be to build capacity in developing countries—both to measure and report forest-related GHG fluxes and also to participate in evaluation processes.

Table of Contents

1	Introduction.....	9
2	Differing objectives of evaluations.....	11
3	Procedural elements common to evaluations.....	15
4	Different evaluation approaches for different contexts.....	23
5	Profile of the evaluators.....	30
6	Lessons learned.....	36
	6.1 What type of evaluations lead to stronger assurances of GHG statements?.....	37
	6.2 What type of evaluations lead to stronger capacity building?.....	39
	6.3 Final considerations for the future.....	40
7	Sources.....	42

Acronyms

AF	Amazon Fund
aka	also known as
BUR	Biennial Update Report
CA	California
CARB	California Air Resources Board (ARB)
CDM	Clean Development Mechanism
CER	Certified Emission Reduction
CFP	Carbon Fund Participant
COP	Conference of the Parties
DOE	Designated Operational Entity
ER	Emission Reductions
ERPD	Emission Reductions Program Document
EU	European Union
EU ETS	European Union Emissions Trading Scheme
FMT	Facility Management Team (of the World Bank that manages the FCPF)
FREL/FRL	Forest Reference (Emission) Level (under the UNFCCC)
FCPF	Forest Carbon Partnership Facility
GCF	Green Climate Fund
GFOI	Global Forest Observations Initiative
GHG	Greenhouse gas
IDEAM	Institute of Hydrology, Meteorology and Environmental Studies (Colombia)
IPCC	Intergovernmental Panel on Climate Change
ISO	International Organization for Standardization
iTAP	independent Technical Advisory Panel (of the GCF)
KP	Kyoto Protocol
LULUCF	Land use, land-use change and forestry
MF	Methodological Framework (of the FCPF)
MRV	Measurement, Reporting and Verification
N/A	Not applicable
QM	Quality management
RBP	Results-based payments
REDD+	Reducing emissions from deforestation and forest degradation and the role of conservation, sustainable management of forests and enhancement of forest carbon stocks
REM	REDD+ Early Movers
tCO ₂ e	tonnes of carbon dioxide equivalent
TAP	Technical Advisory Panel (for the FCPF Carbon Fund)
TBD	to be determined
UNFCCC	United Nations Framework Convention on Climate Change
US	United States
USD	United States Dollar
VCS	Verified Carbon Standard
VCU	Verified Carbon Unit
VVB	Validation and Verification Body (for the VCS)

1 Introduction

Third-party evaluations have long been tools to build trust in statements made by GHG programs and to enable their use -- against international commitments or voluntary targets—often with significant financial implications. Such evaluations can also provide feedback that enables learning and an opportunity for capacity building to improve GHG measurement and reporting. In meeting those objectives, there can be tension between evaluations designed to underpin the fungibility of GHG statements and those designed to retain flexibility to address country circumstances.

This study looks into third-party evaluations, aiming to identify common elements and understanding differences in order to learn lessons for REDD+. It is written for those who undergo evaluations—to understand what to expect and why some are different than others. It may also inform those involved in capacity building on measuring and reporting of GHG emissions, those who provide results-based funding for REDD+, or those who want to understand what type of evaluations would be necessary to transact REDD+ related emission reductions. It may also be useful to those developing or implementing evaluations themselves, to better understand the range of process options and apply appropriate approaches for the objectives that the evaluation is serving.

To inform our analysis, we looked at a variety of evaluation ‘schemes’, for example the UNFCCC technical assessments / analyses of forest reference (emission) levels and REDD+ results, financing initiatives such as the FCPF Carbon Fund and the Guyana-Norway agreement, the Colombia REDD+ Early Movers program, the Kyoto Protocol, and several offset programs such as VCS, CDM and California’s offset program. We also considered the GCF’s emerging iTAP review, and (to a lesser extent) a few other schemes are mentioned throughout the text (e.g. the EU Emissions Trading Scheme, Australia’s Emission Reduction Fund). To understand the various evaluation schemes, we analyzed the guidance and criteria provided to GHG programs and evaluators, program documents and review reports, and conducted interviews with a range of experts involved in varying functions of such schemes.

This report is a comparative analysis of contexts that are not usually mixed, but all of which concern the evaluation of GHG statements. The comparison of such different contexts is immensely difficult, and there were mixed views on whether such a comparison is warranted—for example, some found that evaluations of GHG statements by country-level mitigation efforts should be kept separate from project level schemes. Others appreciated the comparison, but suggested a clear separation between evaluations of program design (including baselines or reference levels) and verifications of emission reduction statements. Those concerned with REDD+ related evaluation schemes—whether under the UNFCCC, for the GCF or the FCPF Carbon Fund—wanted to know how such evaluations compared to others, among other things to understand where there could be room for improvement. While we agree that *GHG programs and their statements* should not always be put into the same basket, we believe that an analysis of *their evaluations* uncovers a range of useful considerations that this report summarizes.

During interviews, we heard much targeted (and sometimes contradictory) criticism of the evaluation processes of the most prominent REDD+ schemes: the UNFCCC, the FCPF Carbon Fund, and the GCF’s forthcoming call for results-based REDD+. While some criticized an apparent ‘lack of teeth’ other found evaluations to be overly ambitious. Also, some found schemes to be overly regulated, while others found them ambiguous in guidance.

Our report begins by making an important distinction in clarifying the context and objectives of evaluations, why they are designed differently, and what their objectives are. The next section deconstructs the procedural

components of evaluation processes and identifies similarities and differences. This is followed by a discussion on what evaluators actually do during reviews or technical assessments, and then who they are. This information leads up to findings on what elements of evaluations contribute most to enhancing the quality of GHG statements and to developing country capacities for GHG measurement and reporting.

Finally, we recognize that what matters more than evaluations of GHG statements are the underlying mitigation actions. How best these are evaluated—through approaches that are flexible to country circumstances or stringent to allow market-based activity—varies by context. Because of this, we find it necessary to point out that our report should not be misunderstood to imply that more stringent evaluations are somehow ‘better’. Rather than making such judgments, our report aims to provide insight into the characteristics of different types of evaluations that achieve different objectives that are all needed to drive mitigation action to achieve the goals of the Paris Agreement.

Throughout this report, we use the generic term ‘evaluation’ to denote the various ways in which GHG statements are reviewed, assessed, or verified. Different GHG schemes use varying terminology for such evaluations, but there are some commonalities: ‘Verifications’ have been used in compliance and/or offsetting schemes when referring to the evaluation of emission reduction statements, although more recently this term has been used in the UNFCCC in a broader sense. ‘Reviews’ tend to be broader and can also refer to other kinds of GHG statements, such as the evaluation of GHG inventory report. ‘Technical assessments’ tend to be focused on producing a set of technical recommendations on improving GHG measurement and reporting to improve its quality. Such nomenclature, however, is not consistently used across various schemes—which can cause confusion when discussing evaluations. Similarly, there are different terms to denote what is evaluated. ISO refers to ‘GHG assertions’, under the UNFCCC countries make ‘submissions’ and under programs that generate carbon units, project design documents and monitoring reports are evaluated. Throughout the document, we refer to ‘GHG statements’ to generically refer to all of these. Finally, we refer to ‘schemes’ when speaking of the various contexts in which evaluations are undertaken (e.g. UNFCCC, FCPF, CDM, VCS, etc.).

2 Differing objectives of evaluations

Evaluation objectives vary substantially, and therefore need to be understood before evaluations can be usefully analyzed. Among the 'schemes' that we considered that include evaluations of GHG statements, objectives were diverse and include:

- Building capacity through (e.g. UNFCCC REDD+) technical assessments / analyses
- Payment for emission reductions that include title transfer (e.g. the FCPF Carbon Fund), or not (e.g. through bilateral arrangements or the GCF)
- Compliance with country-level emission reduction commitments (e.g. the Kyoto Protocol)
- Issuance of carbon credits for offsetting (e.g. VCS, CDM, California's Compliance Offset Program)

Evaluations that lead to the issuance of carbon credits or otherwise to title transfer understandably tend to be the most stringent. Often, these are conducted by auditing companies as pass-fail assessments and underlie certification of a specific quantity of achieved emission reductions. For example, under the CDM, successful projects generate Certified Emission Reductions (CERs), while a portion of projects fail validations resulting in an ineligibility to issue carbon credits. At the time of writing, the UNFCCC webpage includes records of 342 rejected and withdrawn CDM projects (around 4.4% of the total). According to a database by the Institute for Global Environmental Strategies, as of March 2017, issuance reviews had led to adjustments and rejections in excess of 16 million CERs. Similarly, the carbon credit registries for the VCS, issue Verified Carbon Units (VCUs) for projects, and projects participating in the California (CA) offset program generate (Air Resources Board) ARB offsets. The stringent reviews behind carbon credits has enabled their use in multiple schemes, for example most CERs can be imported into the EU Emissions Trading Scheme (EU ETS), although they were originally designed for the Kyoto Protocol.

Compliance-based systems (e.g. the Kyoto Protocol, the EU ETS, California's cap-and-trade) also tend to have stringent rules, and more in-depth assessment of GHG statements. We included the Kyoto Protocol (KP) in this study and note that its review process differs considerably from other evaluations under the UNFCCC, specifically those of Biennial Update Reports, their REDD+ results annexes, or FREL/FRLs. Included among the objectives of the multi-step evaluation process are assessments of whether Parties demonstrate capacity to account for emissions and whether they have met their target. The KP review may lead to 'adjustments' of a country's GHG statement. Review reports are forwarded to the KP Compliance Committee, that may take actions if a country is found in non-compliance.

On the other side of the spectrum are evaluation processes designed to build capacity in GHG measurement and reporting of forest-related fluxes. Most relevant to this paper is the UNFCCC technical assessment process of REDD+ forest reference (emission) levels (FREL/FRLs) and technical analysis of REDD+ 'results', i.e. emission reductions achieved. The FREL/FRL technical assessment process was developed by the UNFCCC COP through a negotiation among countries—what one interviewee suggested represents the "lowest common denominator". The assessment is mandated to be a "facilitative, non-intrusive exchange of information" that identifies 'areas of improvement'. It is 'expert-led', rather than using an auditing body or approach. There is no 'pass-fail' assessment, neither does the technical assessment endorse or approve the FREL/FRL. The COP has also stated (Decision 14/CP.19, para 15) that additional modalities for verification may be needed in order for results to be eligible for market-based approaches.

There are several bilateral arrangements that have provided payments for REDD+ emission reductions, and the evaluation processes they employ vary. Some schemes do not include a transfer of title when payments are made

for achieved emission reductions. Among them, the Guyana-Norway agreement used a verification process that most closely mirrored carbon crediting schemes: an auditing company accustomed to CDM and VCS verifications that applied an approach following the ISO 14064-3 standard. However, flexibility was also provided to take into account the piloting nature of the program, the need to build capacity, and the recognition that this was a development assistance program. The verification was seen as an opportunity for learning—not to penalize the country—and the goal was not only to provide Norway a level of confidence that they were paying for ‘real’ emission reductions, but also to build capacity on forest monitoring and reporting within Guyana, over a period of several years. Similarly, the REDD+ Early Movers (REM) program in Colombia operates more akin to a payment for ecosystem service, where no legal title to emission reductions is transferred. Quantification of emission reductions is based on the technically assessed UNFCCC FREL/FRL (except that its adjustment above the historical average is not recognized). Because of this, the Colombia REM program does not require a further evaluation of the reference level, but does include a verification of achieved emission reductions by an auditing company. Emission reductions are tracked in Colombia’s national REDD+ registry; those paid for by the REM program are retired and not transferred under the arrangement between Colombia and donor governments.

The FCPF Carbon Fund straddles the stringency required by offsetting or title-transfer agreements with the flexibility seen in bilateral and other ‘results-based payment’ schemes. An initial assessment of the Emission Reductions Program Document (ERPD) is carried out via an expert panel (i.e. the Technical Advisory Panel, or TAP) that has a reduced mandate compared to some other evaluation processes—i.e. the assessment it provides is not determinative, but only serves as advisory input into Carbon Fund Participants’ (CFPs) decision making on whether to include the ERPD into its portfolio. Because of this, while the TAP provides a pass-fail on each of the indicators for 37 criteria in the Methodological Framework (MF), there is no overall pass-fail assessment. Neither is the TAP assessment meant to generate capacity-building recommendations that countries would draw on to improve their forest measurement approach. Rather the methodological approaches for measurement and baseline setting are fixed once the ERPD is approved (unless CFPs request specific changes) because subsequent monitoring needs to retain full consistency of data, definitions and methods. Verification of emission reductions for which payment will be made (and title transferred) will be decided in future meetings, although is expected to follow an auditing approach and thus resemble more closely the example of CDM, VCS, CA offset program and the Guyana-Norway bilateral agreement, described above.

Unlike other REDD+ result-based finance approaches developed jointly between donors and forest countries, or in response to meeting the requirements of a multilateral funding instrument, the Amazon Fund (AF) was conceived, and largely set up, by the Brazilian government. The government then invited donors to fund emission reductions from forests in the legal Amazon. Its design reflects a nationally constructed system: drawing on Brazil’s system for monitoring change in the Amazon forest (i.e. PRODES) as data provider and the Brazilian development bank for fund management. The evaluation approach for claimed results is somewhat unique in that a Technical Committee, comprised of Brazilian nationals, with a more limited mandate than other evaluations, provides the attestation of the claimed results. However, we understand from an interview, that the Brazilian government is currently revising the current evaluation procedures and may in the future align more closely with the UNFCCC technical assessment / analysis process.

The case of the Green Climate Fund (GCF) illustrates how evaluation processes are often tailor-made to fit the mandates and process of their specific context. In October 2017, the GCF Board decided to pilot results-based payments (RBP) for REDD+ by launching a Request for Proposals. The pilot will use the GCF’s independent Technical Advisory Panel (iTAP, using LULUCF experts from the UNFCCC roster of experts) to review funding proposals by forest countries, using an agreed scorecard that draws on the UNFCCC technical assessment

(referenced above), as well as the technical analysis of REDD+ results included in the Biennial Update Report (BUR). The score received will determine a volume of eligible emission reductions for payment, although the GCF Board will make final decisions on any financial transfer. In addition, the scorecard contains pass-fail elements thus introducing a more stringent 'filter' beyond the UNFCCC evaluation processes.

Finally, we note that in some cases, mandates, roles and responsibilities are not entirely clear to everyone involved in evaluations. For example, interviews revealed confusion on the FCPF decision-making process that draws on two evaluations with some redundancy: the TAP assessment meant to advise Carbon Fund Participants decision making, and the Carbon Fund Participants' own review, conducted in parallel. Moreover, for the assessment itself, there is an apparent lack of clarity on the approach that should be taken—some see the mandate strictly as an assessment of the data and information provided against the MF while others see an opportunity to provide advice to countries on how to make improvements. This is particularly challenging in the case of the UNFCCC REDD+ technical assessments / analysis as the objective explicitly includes capacity building; some of our interviewees indicated confusion as to the distinction between conducting an evaluation and providing advice.

Table: Context and objectives of several evaluation schemes.

	Context of the evaluation	Objectives of the evaluation	Evaluation output and users
REDD+ UNFCCC	Agreed within the context of the UNFCCC COP, i.e. by all countries	Assess whether a FREL/FRL and subsequently reported REDD+ results are in line with COP guidance; in the case of the FREL/FRL, offer a facilitative, non-intrusive, technical exchange of information to support capacity building; in the case of the REDD+ results, increase the transparency of mitigation actions and their effects	Technical Assessment report (for the FREL/FRL) and a Technical Analysis report (for results contained in the BUR Annex); used by developing countries to build capacity and by the GCF as input into funding decisions
FCPF Carbon Fund TAP	Designed by Carbon Fund Participants (contributors)	Assess the ERPD against the MF to inform Carbon Fund Participants on whether to accept the program into the portfolio; a subsequent verification process for emission reductions is currently being designed	TAP assessment report with yes-no scoring of indicators, including on GHG measurement and reporting, used by CFPs as input into whether or not to admit a country into the funding portfolio
Colombia REDD+ Early Movers verification	Bilateral agreement between sovereign countries, partly based on the UNFCCC REDD+ technical assessment process	Assess whether Colombia's report on achieved emission reductions is consistent with internationally recognized approaches and with the UNFCCC FREL/FRL to enable issuance of emission reductions in the national REDD+ registry and the receipt of results-based payment	Verification report with an auditor statement to backup issuance of emission reductions in the national REDD+ registry; used by donors to the Vision Amazonia program, particularly those making results-based payments
GCF REDD+ RBP	Agreed by the GCF Board, comprised of an equal number of developed and developing countries	Assess funding proposals, based on the UNFCCC REDD+ technical assessment / analysis against a scorecard to calculate a volume of emission reductions eligible for GCF payment and to inform the GCF Board's funding decisions	iTAP scoring of funding proposals against a predetermined scorecard, using the UNFCCC technical assessment/analysis reports and any other supplementary information provided by the country the iTAP review is used by the GCF Secretariat to propose a funding decision to the Board

Norway-Guyana bilateral agreement	Bilateral agreement between two sovereign countries	Provide Norway with a professional and independent verification of the results reported by Guyana; used as basis for funding allocation	Publicly available verification report, including a verification statement, and summary of observations, clarification requests and corrective action requests, as well as an assessment of how public comments to Guyana's results report was taken into account.
Amazon Fund	Host country designed instrument	Provide assurance to donors that Amazon Fund investments are backed by GHG reductions	The AF Technical Committee validates the emission reduction estimates (proposed by the Brazilian MoE) through the proceedings from an annual meeting, to establish a fundraising ceiling for BNDES.
VCS	Decided by the VCS board	Verify project compliance with the VCS standard in order to Issue tradable carbon credits	Validation and verification reports; used by project proponents to request issuance of carbon credits, buyers of VCUs are indirect users (i.e. the verification provides confidence in the asset being purchased)
CDM	Agreed within the context of the UNFCCC COP, i.e. by all countries, with details decided by the CDM Executive Board	Verify emission reductions are real and additional and can be turned into Certified Emission Reductions that may be used for compliance under the Kyoto Protocol	Validation and verification reports; used by project proponents to request issuance of carbon credits, and indirectly by buyers of the CERs
CA offset program	State legislation	Ensure the credibility of offsets that enter CA's cap-and-trade system	Validation and verification reports; used by project proponents to request issuance of ARB offsets
Kyoto Protocol	Agreed within the context of the COP/MOP, i.e. by Parties to the Kyoto Protocol	Provide a technical assessment of the implementation of the KP; promote consistency and transparency; improve reporting; may lead to quantitative adjustment of estimates provided that are basis for assessing compliance	Review reports; users are the Compliance Committee, and all parties to the Kyoto Protocol

Key points:

- The objective and context for evaluations of GHG statements drive design choices; these may differ, and therefore must be understood.
- The nature of those designing the evaluation process (e.g. multilateral bodies, governments, independent entities including those representing the private sector) may also have an impact on the overall approach.
- Evaluations that lead to the disbursement of funding or to the creation of carbon credits tend to be pass-fail, whereas technical assessments chiefly produce technical recommendations designed to contribute to developing GHG measurement and reporting capacities.
- Mandates vary among evaluations, from merely providing input for donors' decision making to confirming the ability to issue carbon credits; sometimes mandates lack clarity, confusing the evaluation process.

3 Procedural elements common to evaluations

While evaluation processes can be structured in a variety of ways, they usually include similar procedural elements. Evaluations are typically structured around consideration of: (a) program design and GHG quantification approach (including baseline or reference level setting), and (b) reporting of achieved emission reductions. There are also common procedural elements: management of the evaluation, guidance on measurement and reporting of GHGs as well as for the evaluation itself, documentation, selection of evaluators and/or their accreditation, and the opportunity to seek redress. Since the process and their procedural elements are similar, evaluations are comparable to each other.

Evaluation processes: Evaluations are often split into an evaluation of the GHG program design and its approach to emissions and removals quantification and construction of reference levels (aka ‘validation’), and an evaluation of the quantification of emission reductions achieved (aka ‘verification’). For example, under the CDM and the VCS, an initial validation reviews the applicability of the chosen methodology, the baseline scenario and the monitoring plan, while the verification is the review of data on achieved emission reductions. Similarly, the UNFCCC process includes a technical assessment of the proposed FREL/FRL and a subsequent technical analysis of stated emission reductions. Some schemes allow these two steps to be collapsed into a single process (e.g. CA offset program does validation during the first verification, VCS also allows this). In some cases, a third step may involve a decision on certification of emission reductions, acceptance of GHG programs, applicable payments, issuance of carbon credits or compliance with targets.

Table: Overview of evaluation steps.

	Evaluation of a GHG program design and emission reduction quantification approach	Evaluation of emission reduction statements	Decision on acceptance of GHG programs, payments, issuance of carbon credits, or compliance with targets
REDD+ UNFCCC	Technical assessment of FREL/FRL	Technical analysis of BUR Annex detailing emission reductions	N/A (to UNFCCC), but going through the UNFCCC evaluation steps is a requirement to access GCF funding
Colombia REDD+ Early Movers	UNFCCC technical assessment (except that adjustment is disregarded)	Verification of stated emission reductions	Issuance of emission reductions in national REDD+ registry and retirement against results-based payments
FCPF Carbon Fund	TAP assessment of ERPD	TBD	Carbon Fund Participants make final decision
GCF REDD+ RBP	Scorecard review by the iTAP		Board decision for funding
Norway-Guyana bilateral	Review of methods and results done in one step by auditing company		Final review by Norway for payment
Amazon Fund	N/A	AF Technical Committee attests to the stated amount of reduced emissions by the Ministry of Environment	Donors may use the evaluation when determining if, and how much, to provide to the Fund

VCS	Validation of the Project Document, including the applicability of methodologies to establish the baseline	Verification of monitoring report (stated emission reductions)	Issuance of carbon credits (VCUs)
CDM	Validation of the Project Design Document, including the applicability of methodologies to establish the baseline	Verification of monitoring report (stated emission reductions)	Certification of emission reductions and issuance of carbon credits (CERs)
California offset program	Validation is combined with the first verification of emission reductions, i.e. the first verification also provides assurance that the offset project has been set up to conform to the requirements of the regulation and applicable Compliance Offset Protocol		Issuance of carbon credits (ARB offsets)
Kyoto Protocol	Review of initial reports (calculation of assigned amount) and Forest Management Reference Level	Review of annual reports	Determination of compliance with Kyoto Protocol targets

GHG program design and choosing an emission reduction quantification approach requires taking a number of involved decisions. For example, constructing REDD+ reference levels require decisions on the period of historical data used and how to project expected emissions in the absence of action. According to one of our interviewees, “these decisions are highly political”, rendering their evaluation potentially contentious, unless very clear criteria are defined. Once the approach is established, however, verifications of achieved emission reductions are more data driven, making the evaluation more straightforward.

In some instances, there is a final step—a decision to provide results-based finance or issue carbon credits or validate compliance with criteria. In some cases, this step immediately follows the verification. In other instances, particularly donor-funded instruments such as the FCPF Carbon Fund, there may be a more involved consideration by the donor government—in addition to, e.g. the TAP assessment—before a country may move to the next step (e.g. placement in a funding portfolio). This type of case-by-case decision making is less feasible where turnover is high, such as in the CDM and VCS, which churn through thousands of projects. In these cases, evaluations are more standardized, have a heavier focus on rules, and also include appeals procedures.

An example of a multi-step process is provided below that combines the UNFCCC REDD+ and Green Climate Fund’s processes. The first step is the technical assessment of the FREL/FRL, followed by a second step, the technical analysis of stated results (in the BUR Annex). These evaluation reports then feed into the GCF’s own evaluation process, where funding proposals are ‘scored’ based on the UNFCCC technical assessment / analysis reports (along with any additional information provided in the proposal) to translate the analyzed results into a volume of GCF-eligible emission reductions that may receive finance from the Fund. The final decision on funding is made by the GCF Board.

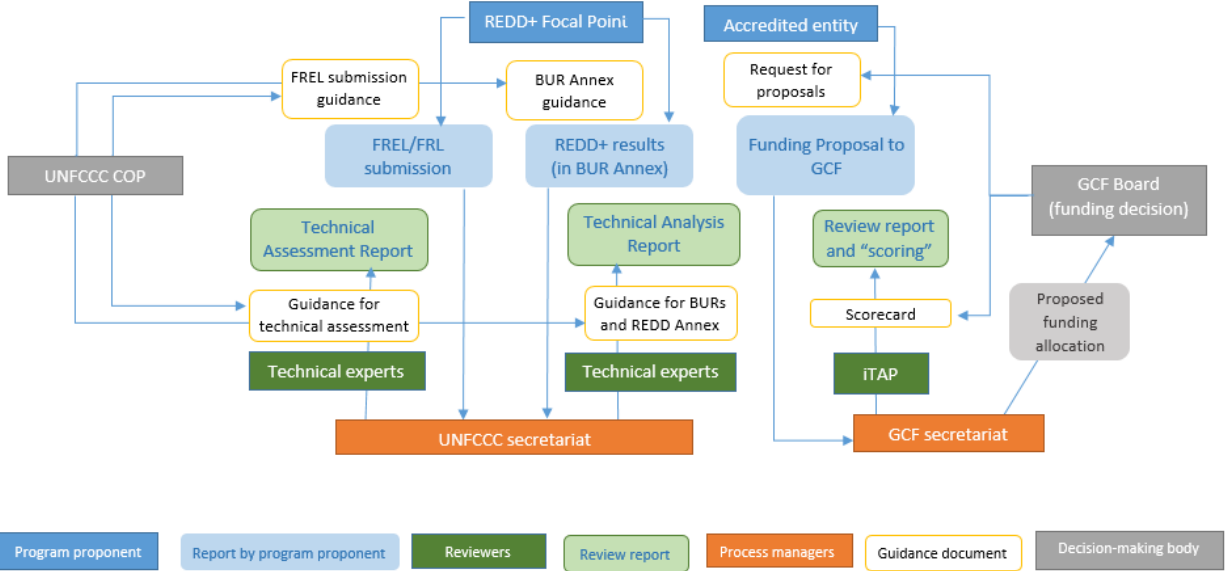


Figure: Flow-chart of the technical assessment / analysis for REDD+ under the UNFCCC and for the GCF results-based payments iTAP review.

Communication modalities: Evaluations may include interviews with technical staff responsible for GHG monitoring and reporting to understand data and methodological choices made and their context. The organization and format of such communication can make a difference in the dynamics, and thus, in the evaluation outcome. In the UNFCCC REDD+ technical assessments / analyses, all communication between countries and the technical experts is checked and relayed through the UNFCCC Secretariat. Other schemes, such as the FCPF TAP assessment, the validations and verifications for CDM / VCS / CA offset program and the Norway-Guyana review, have evaluators spend an extended time in-country communicating independently with the country team. In addition, some evaluation schemes include a mechanism for collecting public comments; for example, in the Guyana-Norway bilateral, the verification team directly and explicitly were required to consider such public comments.

In the UNFCCC process, communication between the UNFCCC Secretariat and the countries undergoing technical assessment of their FREL/FRLs includes multiple rounds of commenting on the draft assessment reports. One interviewee suggested that such exchanges could translate into an opportunity to influence the technical assessment outcome. In extreme cases, a country may hold up publication on the UNFCCC website of an evaluation report. By contrast, other evaluation processes require the party under review to use a formal process of complaints and appeals in case of disagreement.

Procedural elements: Albeit designed for different contexts and to achieve different results, the evaluations that we looked into had similar procedural elements. There is a governing body, GHG program design, GHG monitoring reports and evaluation reports, guidance and criteria for GHG quantification, as well as evaluation guidance; some (but not all) schemes require an approach for accreditation of evaluators, and in some cases complaint procedures. These functions are illustrated below.

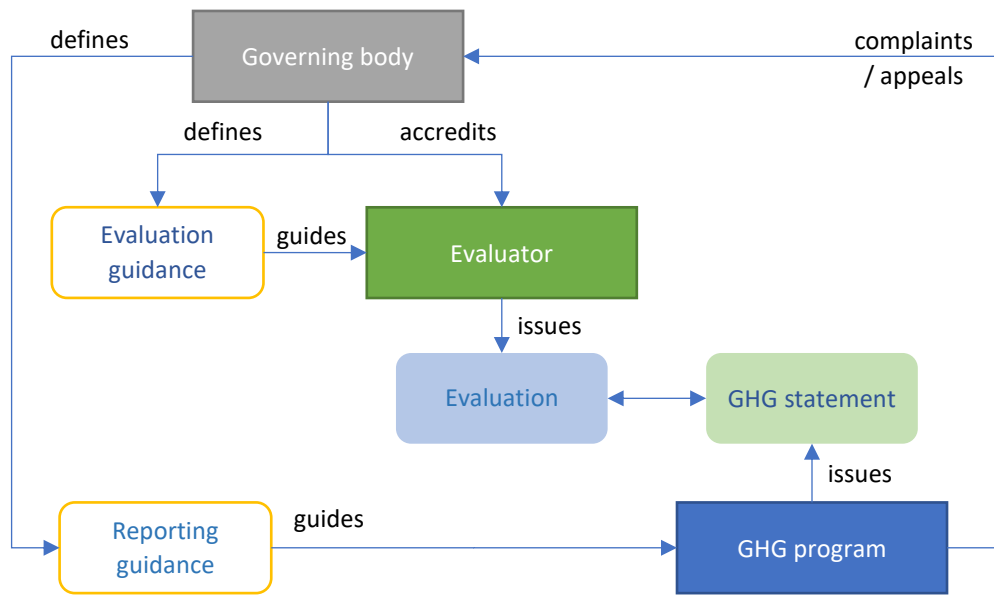


Figure: Procedural elements of evaluations.

Management: Expert-led evaluations, such as the UNFCCC technical assessments / analyses, the FCPF TAP assessment, and the GCF iTAP review require secretariats to facilitate the process; in some instances these bodies take an important role. The UNFCCC Secretariat manages the UNFCCC REDD+ assessments, assembling the expert teams, providing technical and procedural advice, carrying out quality control, and editing the evaluation reports—in part, to ensure a level of consistency across the evaluations. However, our interviews suggested that such functions were implemented to varying degrees across evaluations, and thus not conducive to ensuring consistency. Similarly, the World Bank’s Facility Management Team contracts TAP experts, manages the assessment process and takes a more active role than the Technical Advisory Panel itself in presenting the assessment results to the Carbon Fund Participants. During interviews, it was highlighted that the Facility Management Team also did not manage evaluations evenly across TAPs and also has multiple roles, including supporting FCPF countries in developing the ERPDs under assessment (resulting in, at time, conflicts of interest). Similarly, the GCF Secretariat is expected to manage the review process for REDD+ results-based finance, including finding and contracting members for the iTAP, conducting its own due diligence of the funding request, and shepherding the overall process. Where reviews use duly accredited auditing companies, as in the case of CDM, VCS, CA offset program and the Norway-Guyana agreement or else when using the ISO 14064 standard, management functions are more heavily covered in-house (e.g. within the auditing company).

Contracting of evaluators also may vary and can be undertaken by the GHG program, prospective funders or by the secretariats that manage the evaluation. For example, in the case of CDM and VCS, the project proponent contracts and funds the validator and verifier, as was done similarly by IDEAM, the national agency responsible for GHG monitoring, on behalf of Colombia’s REM program. Conversely, for the case of the Guyana-Norway bilateral, Norway (i.e. the buyer) managed the contracting of the auditor. In both the REM-Colombia and Guyana-Norway cases, the selection of the auditor was agreed by both Parties (buyer and seller). In the case of the FCPF, the World Bank FMT selects and contracts the experts. And in the case of the UNFCCC, the Secretariat selects the technical assessment / analysis teams off a list of country-nominated experts.

Guidance and criteria on measurement and reporting: Evaluations focus on compliance with applicable guidance and criteria for GHG programs. In the different schemes we analyzed, guidance on measurement and reporting, as

well as on the evaluation itself, have been developed at significantly different levels of detail. The context in which such documents have been developed also varies—some were motivated primarily through a scientific lens, while others were developed through a more political process. For example, the IPCC Guidelines for national GHG inventories are developed by a scientific panel, covering hundreds of pages with detailed explanations of methodologies for estimating GHG emissions and removals. UNFCCC guidance for developing REDD+ FREL/FRLs states that countries should use IPCC guidelines, however, it has an import gap: how to project historical emissions into the future as a reference level (including calculation of an ‘adjustment’ of historical estimates, if a country chooses to do so). The absence of guidance and criteria on this crucial point leaves much room for interpretation. Carbon offsetting schemes (e.g. CDM, VCS, CA offset program) have very detailed technical guidance, often over a hundred pages long, which includes great detail on baseline setting.

The FCPF MF is a hybrid, with more detail than the UNFCCC guidance for FREL/FRLs, REDD+ results, and their assessment, but also containing some level of ambiguity common in negotiated texts. Several interviewees suggested that the MF was at times ambiguous or lacked detail; it was even suggested during interviews that the Carbon Fund Participants themselves (who developed the document) do not agree on its interpretation in several places. At the same time, one interviewee also suggested that the MF was also, at times, constraining, i.e. there were points that the expert desired to make that had an impact on the quantified estimates, but the indicators were constructed such that it was difficult to highlight such points in the assessment report. For example, it was noted that there are no criteria and indicators on good practices for setting up a monitoring process, nor any that directly concern the GHG statement and its accuracy, which may cause further constraints during the forthcoming verification.

Evaluation guidance: Among different schemes, there are also varying levels of detail provided to evaluators in how to conduct the evaluation itself. Some provide little to no guidance, while others (in particular, carbon crediting schemes such as CDM, VCS and the CA offset program) provide much detail. For example, the CDM validation and verification standard is more than 70 pages in length, and provide requirements for auditors’ activities. In addition, auditing firms have internal standard operating procedures, often following ISO 14064 that is a prerequisite for obtaining accreditation. Other schemes—particularly those that do not require formal accreditation or that do not use the ISO standard—do not have such formalized, written documentation on conducting the evaluation. For example, the Carbon Fund TAP assessment does not have official FCPF documentation to guide experts, who are left to follow a mixture of ad-hoc instructions from the FMT, miscellaneous clues in reporting and assessment templates and, to the detriment of a consistent approach, their own professional background.

Where there is little guidance, whether on measurement and reporting or on the evaluation approach, those managing the process and conducting quality control may need to be more hands-on to guide evaluators. Precedents also become important when guidance is more generic. For example, some of our interviewees pointed to the influence that the first technical assessment of Brazil’s FREL/FRL and the first technical analysis of its REDD+ results in 2014 and 2015 had as a precedent to evaluations undertaken in the following years, especially because UNFCCC guidance contains little technical detail. The current practice under the UNFCCC of writing REDD+ related technical assessment / analysis reports includes important “artifacts” from that first case; for example, the summary assessment does not mention the ‘accuracy’ of GHG statements, although it is ostensibly an evaluation criterion.

Table: Guidance and criteria for GHG reporting and performance evaluations in analyzed schemes.

	Guidance and criteria for GHG statements	Guidance and criteria for the evaluation process
REDD+ UNFCCC	The guidance provided in a COP decision for submission of a FREL/FRL is just one page long, but in addition countries are expected to use the extensive IPCC Guidelines to estimate emissions and removals; however, there is an important gap with regards to establishing 'adjustments'	Guidance provided for the technical assessment includes objectives, scope, procedures, and timing. It provides the general principles (i.e. accuracy, transparency, completeness and consistency).
Colombia REDD+ Early Movers	Targeted guidance on reporting and the verification process in three outline papers was agreed between the donors (Germany, UK, Norway) and the Government of Colombia.	Targeted guidance on reporting and the verification process in three outline papers was agreed between the donors (Germany, UK, Norway) and the Government of Colombia.
FCPF Carbon Fund TAP assessment	The Carbon Fund developed a MF that contains guidance for carbon accounting, including 20 criteria with 40 indicators.	A short guidance document is provided to the TAP, with little information on how to actually conduct the assessment (e.g., in terms of evaluation methods, the role of technical recommendations and the level of assurance to arrive at). Ambiguities in the MF make assessments challenging.
GCF REDD+ RBP	Use of UNFCCC guidance (as above); more detailed requirements on baseline setting are provided.	There are 24 carbon accounting elements that are scored to generate a GCF volume of emission reductions that may receive payment—however, a lack of detail will likely make reviews difficult.
Norway-Guyana bilateral	The Joint Concept Note agreed by both Parties includes specific indicators for estimating GHGs from reduced deforestation and forest degradation (from different drivers), as well as agreed monitoring methods.	The verification company was asked to propose a verification process when responding to the tender based on general terms of reference.
Amazon Fund	None provided.	Little to no guidance provided on how to conduct the evaluation; evaluation focused on ensuring the calculations to generate a (conservative) value for emission reductions from deforestation were done correctly.
VCS and CDM	Detailed guidance is in the project standard and approved methodologies. Reviews assess compliance with the methodologies and because these are quite detailed, little 'expert judgement' is required.	The VCS has a validation and verification manual, and CDM a standard, that provides entities with detailed guidance and requirements on how to conduct a review.
California offset program	CARB adopts Compliance Offset Protocols (COPs) that define in detail the eligibility of different types of offsets.	In addition to the detailed COPs, CARB issued a document providing detailed guidance for offset verification.
Kyoto Protocol	Countries must use IPCC guidance to estimate emissions and removals (and the targets themselves use a 1990 base year, and are negotiated rather than based on guidance, with exception to LULUCF).	Decision 22/CMP.1 provides fairly detailed "Guidelines for review".

Documentation: Different schemes have different requirements for documentation and public disclosure. For example, the CDM, VCS and the Guyana-Norway review reports are made publicly available and include considerable detail on the process leading up to the concluding assessment. The exchange between the reviewers and the GHG program under assessment is made public, including through a detailed log of information and corrective action requests. All original data collected during the review is also made public. The FCPF also makes a number of reports publicly available, including a draft ERPD (that is assessed by the TAP), the first TAP assessment report as well as comments from Carbon Fund Participants, the final ERPD and final TAP assessment report. In this regard, one can follow the iterative process, including any changes that have been made in the ERPD in response to comments, although there is not a communication or iteration log as such. In the UNFCCC, technical assessments / analyses of FREL/FRLs and REDD+ results include similar exchanges that are documented by the UNFCCC Secretariat, but only the final assessment is made public. The Amazon Fund's evaluation report consists only of a handful of sentences within the minutes of an annual AF Technical Committee meeting, providing no information on the process or considerations that led to the final assessment. The REM was the least transparent scheme we analyzed; public documents are difficult to find and little to no information on the evaluation process or guidance is publicly available.

Selection of evaluators and accreditation: Different schemes have different approaches for choosing evaluators and ensuring their technical and procedural qualification. Expert-led processes (e.g. UNFCCC, FCPF TAP) tend to draw from rosters. In some cases, leading experts are selected to implement the evaluation; in other cases, persons with little technical qualification may become involved. In the case of the UNFCCC, such experts are nominated by their countries to the roster (but represent themselves, not their countries in their work). The FCPF has created its own roster of experts, but has a less clear nomination process. One interviewee pointed out that, in both of these schemes, there is a lack of transparency and selection criteria from within such rosters and that this could create a potential conflict of interest risk. In the Amazon Fund, the composition of the AF Technical Committee is on invitation by the Brazilian Ministry of the Environment and the Brazilian Climate Change Forum.

By contrast, the CDM, VCS and the CA offset program set a more formalized accreditation process for evaluators, modeled along the lines of what is common in other technical standards and their auditing. According to ISO 14065 and ISO 17011, such an accreditation is the "third-party attestation related to a validation or verification body conveying formal demonstration of its competence to carry out specific validation or verification tasks". For example, in the case of the CA offset program, such demonstration of competencies requires extensive documentation on verification team skills, including a list of all verification staff and a description of their duties and qualifications (education, experience, professional licenses, etc.). As per ISO 14065, accreditation also requires the auditing firms to put in place management systems, ranging from handling of liabilities, to procedures for dealing with disputes to internal audits, to human resources. Generally, only firms with solid quality management procedures will be able to obtain accreditation. The CDM Executive Board has a dedicated standard of more than 50 pages regulating accreditation and maintains a designated panel for overseeing its implementation. Auditing firms undergo desk reviews, on-site assessments and unscheduled spot-checks of evaluation performance to obtain and retain accreditation. Because of the stringency of the process, other schemes, such as the VCS, have opted to automatically accredit CDM-approved auditing firms (in addition to other VCS-approved auditors with ISO 14064 accreditation for the scope VCS).

Seeking redress: Where GHG programs disagree with evaluation outcomes, some schemes include a process for raising complaints and appeals. For example, countries can challenge the Kyoto Protocol reviews and any proposed adjustments to country GHG statements before the Compliance Committee, which then acts as an arbiter. For schemes that generate a high volume of emission reduction statements, like the CDM and the VCS, the option to

seek redress is especially important and, as per ISO 14065, auditing companies need to maintain complaints and appeals procedures that can help resolve disagreements. In addition to those, considerations around an appeals procedure against CDM Executive Board decisions on registration of projects and issuance of carbon credits have been a topic of discussions at several COPs, especially because it is not obvious that there could be legal recourse against its decisions under any national legal framework. The VCS includes a well-defined mechanism for raising complaints and for appealing against decisions by the VCS Association (i.e., the body running the VCS).

Key points:

- Evaluations may be structured in different ways, but generally include: validation of GHG program design, in particular the approach to quantify emission reductions, and the verification of stated emission reductions; some processes may include a third step where final decisions are made by another entity, e.g. by donor governments, whose decision-making may require their own evaluation.
- Evaluations of approaches to quantify emission reductions are more complex and cover a range of potentially thorny issues (such as baseline construction), while subsequent verifications are data driven and therefore tend to be more straightforward.
- Generally speaking, evaluations have similar procedural elements: a governing body, a GHG program design, GHG monitoring reports and evaluation reports, and guidance and criteria for GHG quantification, as well as evaluation guidance. In some cases, there is an accreditation procedures and provisions for complaints.
- Guidance on measurement and reporting and especially baseline / reference level setting, as well as on how to conduct the evaluation itself, is provided at significantly different levels of detail and is also developed through different processes, some scientific and others political.
- Secretariats play an important role in some evaluation schemes, especially those that rely on expert panels, through the management of the evaluation process and, at times, by carrying out quality control or by providing informal guidance.
- More formal accreditation processes can provide transparency and sometimes better ensure the quality of the evaluation team.

4 Different evaluation approaches for different contexts

Evaluations of GHG statements are conducted in different contexts, and result in different approaches. Evaluations used in compliance and/or offsetting schemes provide an assurance on whether a GHG statement complies with applicable guidance and criteria and is materially correct. Their approach is designed to generate a minimum level of assurance and draw on specific review methods and approaches, including the concept of materiality. Such evaluations tend to have a strong mandate, if not to pronounce a pass-fail rating, then to propose adjustments to stated GHG emissions. By contrast, some evaluations do not lead up to a certain level of assurance on the correctness of GHG statements and are not meant to demonstrate compliance with targets or be used to issue carbon credits. Such evaluations tend to focus on producing a set of technical recommendations on improving GHG measurement and reporting to improve its quality, in light of applicable guidance and criteria and to *eventually* arrive at correct GHG statements.

Different schemes provide varying guidance on how to approach the evaluation itself and may employ a variety of evaluation methods. Some focus on data, others on methods, and there are also evaluations that assess the monitoring system, aiming to understand processes and institutions for generating information. The choice of methods will be determined by the statements that evaluations are aiming to derive and the assurance that these imply.

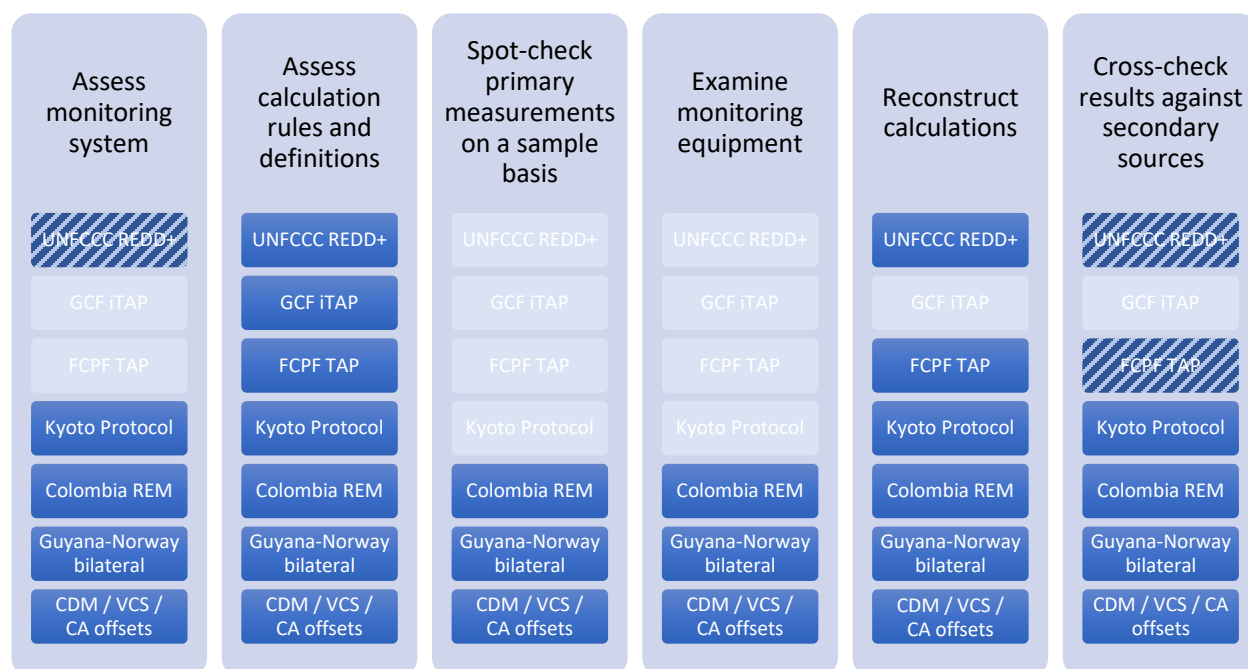


Figure: Methods employed in different evaluations of GHG statements.

Some evaluations are focused on evaluating the data and information provided, while others also focus on assessing the institutions and arrangements, governance and decision-making, and processes for generating data and managing quality control—one interviewee referred to this as the "enabling environment" to generate robust data and information. For example, some reviews (e.g. Guyana-Norway, CDM, KP) aim for understanding the systems used to generate and archive data, provide quality management procedures and sustain such processes. By contrast, the centralized UNFCCC FREL/FRL technical assessment focuses only on the data and information provided and do not include an evaluation of the national forest monitoring system. Evaluation of a country's

monitoring system may occur during the technical analysis of REDD+ results, depending on the evaluator. Finally, the evaluation by the Technical Committee of the Amazon Fund is the 'lightest' process—which some may consider justified since stated emission reductions are based on the PRODES system, which provides considerably transparent information that allows civil society to check its veracity.

There are also differences in whether the evaluations check primary measurements. Auditing approaches (e.g. for CDM, VCS, CA offset program, Guyana-Norway) often include such spot-checking on a sample basis, e.g. through independent re-measurement of land use change or tree measurements. This usually requires an in-country visit, necessary logistical provisions and budget allocation. According to one interview, the impact of individual erroneous measurements on aggregate GHG estimates tends to be most significant where GHG programs rely on fewer measurements.

Cross-checking of data is sometimes also undertaken in the UNFCCC REDD+ technical assessments / analyses, but only against already available data from independent sources and usually only regarding secondary data (e.g. country-level deforestation areas, average biomass densities, etc.). Going back to the primary measurements is hardly feasible, if only because these are centralized efforts at the UNFCCC Secretariat's office in Bonn, Germany, without easy access to original remote sensing imagery or field measurements. The Kyoto Protocol reviews do not include checking of primary data either, one interviewee explained that there is a bona fide assumption that countries' primary measurements are free of willful manipulation—an assumption that those who designed carbon credit verifications (e.g. CDM, VCS, CA offset program) would not have been willing to make. Notwithstanding this assumption, the Kyoto Protocol reviews include the possibility of adjusting estimates in cases where GHG statements appear biased.

While evaluations against detailed methodological guidance and criteria often put data and their uncertainties at the center of attention, other evaluations focus on the methods and definitions applied. For example, UNFCCC REDD+ technical assessments commonly include much discussion around forest definitions and definitions of REDD+ activities. On the other hand, validations for the CDM, VCS, and the CA offset program are able to focus much more on data, since key definitions and methods are already established by the standard itself and its methodologies.

Evaluations also have different ways to represent compliance with criteria. For example, while the FCPF TAP assessment requires a binary yes-no rating for each indicator, the KP reviews and the technical analysis of national GHG inventories in BURs are more detailed, identifying the level (sometimes partial) of compliance with criteria (e.g. not transparent, partially transparent, mostly transparent, fully transparent). In general, 'full' compliance with criteria is hard to achieve. For example, it is not immediately obvious what a 'fully accurate' or a 'fully transparent' GHG measurement could be. In some cases, UNFCCC REDD+ technical assessments / analyses have also begun employing such differentiations, although our interviews suggest that no consistent guidance is imposed.

Given the broad scope of measuring GHG emission reductions (or removal enhancement), exhaustively evaluating all aspects of GHG measurements can be an enormous task. Therefore, evaluators are forced to prioritize the most important aspects of GHG measurements. Because of this, auditors focus on 'material' issues, i.e. those that pose the highest risks to the quantification of the final results. For example, material issues might be those expected to cause at least a 2 to 5% error in the emission reduction estimate. Assessing such risks enables the auditor to prioritize the most critical aspects of GHG quantification for review. Although there is no well-defined concept of materiality in the context of the Kyoto Protocol, the identification of key categories and thresholds for insignificant emissions enables some prioritization of review work. Conversely, the Carbon Fund TAP assessments and UNFCCC

REDD+ technical assessments / analyses use neither key categories nor materiality thresholds. While those conducting the evaluation may focus efforts on areas they believe have the greatest impact on the final GHG statement, there remains a risk that the evaluation results in a lack of prioritizing the most critical data and assumptions, and subsequently that identified areas for improvement do not necessarily relate to the most significant shortcomings in emission reduction terms. The ISO 14064-3 defines an encompassing concept of materiality, relevant to both validations and verifications, shown in the box below.

Box: The concept of materiality (quote from the ISO 14064-3).

"The objective of any validation or verification of GHG information is to enable the validation or verification body to express an opinion on whether the organization's or GHG project's GHG assertion is prepared, in all material respects, in accordance with the intent of its internal GHG programmes or any GHG programme to which they subscribe. The assessment of what is material is a matter of professional judgement. [...]

A discrepancy, or the aggregate of all discrepancies, in a GHG assertion is considered to be material if, in the context of surrounding circumstances, it is probable that the decision of a person who is relying on the GHG assertion, and who has a reasonable knowledge of business and GHG activities (the intended user), would be changed or influenced by such a discrepancy or the aggregate of all discrepancies.

[...] The acceptable materiality is determined by the validator or verifier of the GHG programme, based on the agreed level of assurance — a higher agreed level of assurance generally implies a lower materiality.

In order to ensure consistency and avoid unanticipated discrimination, some GHG programmes or internal programmes assist this decision-making process by including materiality thresholds. This can be defined at the overall level, such as 5 % of an organization's or GHG project's GHG emissions. It can also include varying thresholds depending on the level of disaggregation, such as 5 % at the gross organizational level, 7 % at the facility level, and 10 % at the GHG source level. Furthermore, a series of discrete errors or omissions identified within a particular disaggregation level (individually less than the materiality threshold) can, when taken together, exceed the threshold and can thus be considered material. Identified omissions or errors that represent amounts greater than the stipulated threshold are predetermined as being a "material discrepancy", that is, a nonconformity. [...]"

Auditing companies use a standard risk-based approach while experts serving on assessment panels are often more used to following a list of evaluation criteria or aspects of GHG measurement, providing an assessment of each item. Risk-based auditing approaches in the CDM and the VCS, but also in other evaluations conducted by auditing firms, systematically focus on aspects of a process where risks for errors are greatest. The auditor first analyzes the potential sources of error, then assesses available management controls that enable mitigating these errors. Finally, residual error is assessed drawing on sampling of operational data. The whole process and especially the sampling plan are guided by the level of assurance the auditor is asked obtain and the applicable materiality threshold. Only the Guyana-Norway bilateral and the Colombia REM program used an auditing company for verification without specifying a materiality threshold, implying that all errors needed to be considered material. Not having such thresholds in Kyoto Protocol reviews, UNFCCC REDD+ technical assessments / analyses or in FCPF TAP assessments, in theory, requires evaluating all aspects of GHG measurement and reporting to the same level of detail.

Level of assurance: Depending on their design, evaluations deliver different levels of assurance. The UNFCCC REDD+ technical assessment / analysis reports or the FCPF TAP assessment reports do not include language where evaluators would assure that, after having covered all relevant technical aspects, countries' FREL/FRLs, ERPDs or REDD+ results reports are (materially) correct. Providing a reasonable assurance of material correctness may be translated into everyday language as: "In our opinion this is correct". Auditing companies—whether working for bilateral REDD+ deals or CDM/VCS verifications—issue such declarations as part of the summary statements, duly qualifying them by the applied materiality thresholds. Such terms of art originate from financial auditing and may

therefore be more relevant to market-based mechanisms, where units are treated as tradeable commodities with financial value. For example, under California’s regulations, verification bodies must be able to state with reasonable assurance that reported GHG emission reductions (or removal enhancements) are no more than a 5% overstatement of the ‘true’ value. In contrast, the UNFCCC and the FCPF TAP reports may be understood as providing a more limited assurance, which could possibly be paraphrased as: “We haven’t found big problems.” Understanding the differences in the two statements is critical when interpreting evaluation results.

Corrective actions / opportunities for modifications: In some processes evaluators have the ability to request corrective actions. For example, the KP expert review teams have the mandate to request corrective actions during the reviews. If the country under review does not resolve such issues, they may become either listed as ‘questions of implementation’, if related to a failure in meeting reporting requirements and/or in providing transparent information, or be addressed through a proposed conservative quantitative adjustment to the reported GHG estimates, if related to a failure in the consistency, completeness and/or accuracy of the estimate. Similarly, the CDM and VCS validations and verifications rely on corrective action requests that GHG programs need to resolve before the validation or verification can conclude. In such cases (CDM, VCS), the evaluators have the mandate to provide a ‘binding’ opinion that will largely determine whether projects can register and issue carbon credits. The dynamic is ‘softer’ for the UNFCCC REDD+ technical assessments / analyses that have more facilitative and advisory mandates. In the case of the FCPF Carbon Fund, after the TAP’s first assessment of the advanced draft ERPD, countries often modify certain aspects before submitting a final ERPD and, even then, the Carbon Fund Participants often require further changes before signing an Emission Reduction Payment Agreement. As for the Amazon Fund, because the Technical Committee must attest to the stated emission reductions, any issues raised needs to be resolved; our interviews suggested that at least one time an identified miscalculation led to a correction.

Box: Corrective action requests, clarification requests and forward action requests – quote from the CDM validation and verification standard for project activities.

“The DOE shall raise a corrective action request (CAR) if one of the following situations occurs: (a) The project participants have made mistakes that will influence the ability of the proposed CDM project activity to achieve real, measurable, verifiable and additional GHG emission reductions or net anthropogenic GHG removals; (b) The applicable CDM rules and requirements have not been met; (c) There is a risk that GHG emission reductions or net anthropogenic GHG removals cannot be monitored or calculated.

The DOE shall raise a clarification request (CL) if information is insufficient or not clear enough to determine whether the applicable CDM rules and requirements have been met.

The DOE shall raise a forward action request (FAR) during validation to identify issues related to project implementation that require review during the first verification of the proposed CDM project activity. The DOE shall not raise a FAR that relates to the CDM rules and requirements for registration of the project activity.

The DOE shall resolve or “close out” CARs and CLs only if the project participants modify the project design, rectify the PDD, or provide additional explanations or evidence that satisfy the DOE’s concerns. If this is not done, the DOE shall not submit a request for registration of the proposed CDM project activity.

The DOE shall report on all CARs, CLs and FARs in its validation report. This reporting shall explain the issues raised, the responses provided by the project participants, the means of validation of such responses and references to any resulting changes in the PDD or supporting annexes.”

Summary statement: Often, evaluations are designed to lead up to summary statements that concisely communicate the evaluator’s assessment without the need to digest a much more detailed report. For example, in

case of the UNFCCC REDD+ technical assessments / analyses, the summary statement provides condensed information on whether the submitted FREL/FRL followed the guidelines adopted by the COP. The Amazon Fund's summary statement is reduced to a short attestation of the Brazilian Ministry of the Environment's emission reduction estimates, without referring to evaluation guidance and criteria. Verification statements by auditing companies, e.g., for offsetting mechanisms such as CDM, VCS and CA offset program, are a statement to verify that the estimated amount of emission reductions is free of material misstatements. Examples that illustrate the differences among such statements can be found in the box below.

Table: Examples of summary statements in several evaluation schemes.

Scheme	Typical summary statement
UNFCCC REDD+ technical assessment / analysis	<p>Typical text: "The assessment team notes that the data and information used by [country X] in constructing its Forest Reference (Emission) Level are [mostly/partly] transparent and complete, and are in overall accordance with the guidelines contained in the annex to decision 12/CP.17. [Some important components of the submission are still not fully transparent, accurate, consistent and complete.] [Most of the areas for technical improvement are related to...] [This report contains a few areas identified for future technical improvement...]"</p>
Colombia REDD+ Early Movers verification	<p>Example for the years 2013-2014: "It is the verification team's opinion that the results provided in the Colombia report on emission reductions in the Amazon Biome compared to the FREL registered with the UNFCCC for the years 2013 and 2014 (8,852,106 t CO₂e for 2013 and 10,747,409 t CO₂e for 2014, respectively):</p> <ul style="list-style-type: none"> • have been obtained by applying methodologies in accordance with internationally accepted good practices and defined by the relevant verification criteria; • are free from further omissions and misrepresentations that could lead to considerable errors and / or discrepancies, and can be considered as accurate as possible within the scope of this verification; • are consistent with the methodology established by Colombia in the development of its Reference Level for the Amazonian Biome (FREL); and • can be reconstructed using a transparent and coherent step-by-step process."
FCPF TAP assessment	<p>Example: "Based on the methodological framework (MF), the TAP has rated the ERPD as follows: Final Draft ERPD dated [X]: Of a total of 80 criteria and indicators [X] criteria or indicators are met (YES) and [X] are not met (NO); [X] indicators have been classified as Not Applicable (N.A) to the current assessment.</p> <p>Additional statement: "It is the TAP's assessment that most of the criteria and indicators are met, thereby providing a high degree of confidence in the viability of [X country] proposed ER Program on the basis of established criteria and indicators of the Methodological Framework ... In the TAP's opinion, those indicators currently rated as "NO" (not met) should not prevent the Carbon Fund Participants from further consideration of the Emission Reduction Program as the deficiencies could be addressed through contractual provisions (or conditions) negotiated as part of an Emission Reduction Payment Agreement between the Carbon Fund, the Trustee and the Government of [X]."</p>
Norway-Guyana verification	<p>Example for the year 2015: "It is DNV GL's opinion that the results provided in the Interim Measures Report by Guyana Forestry Commission dated 30 November 2015:</p> <ul style="list-style-type: none"> • have been obtained applying methodologies in accordance with internationally accepted good practices as defined by the verification criteria; • are free from omissions and misrepresentations that could lead to material misstatements. <p>Furthermore, recommendations for improvements in future monitoring periods are summarized as Minor Corrective Action Requests (MINORs) or Observations. [...]</p> <p>DNV GL has verified that the values for the interim indicators in this monitoring period (1 January 2014 to 31 December 2014) are:</p> <ul style="list-style-type: none"> • Indicator 1: Gross Deforestation rate in Year 5 -> 0.065% [further indicators follow]"

Amazon Fund	Example for the year 2016: “The technical note 25/2016/DPCD/SECEX/MMA was approved by the AF Technical Committee. Based on the methods adopted by the Ministry of the Environment for calculating emission reductions from deforestation, considering a mean value of 132.3 tC /ha and a deforestation area of measured by PRODES/INPE of 6207 km2, for the period August 2014 to July 2015, the AF Technical Committee validates the amount of USD 2,505,086,200 as maximum funds capture for the Amazon Fund during the period considered.”
VCS verification	<p>Example: “In our opinion the net anthropogenic GHG removals of the [X project] for the period [X dates] are fairly stated in the monitoring report [ref: date]. The net anthropogenic GHG removals were calculated correctly on the basis of the approved baseline and monitoring Methodology [ref: approved method] ... [The auditor] verified that the net anthropogenic GHG removals from [X project] in the reporting period [X dates] are [X tCO2e, subtracting out e.g. buffer to provide a quantified number of VCUs].”</p> <p>Example: “After review of all project information, procedures, calculations, and supporting documentation, [auditor] confirms that the monitoring conducted by the project proponent, along with the supporting Monitoring Report, are accurate and consistent with all aforementioned VCS criteria, the validated Project Document, and the selected methodology [ref: approved VCS method]. [The auditor] confirms that [ref: project Monitoring Report] has been implemented in accordance with the validated Project Document ... The GHG assertion provided by [ref: project proponent] and verified by [auditor] has resulted in the GHG emissions reduction or removal of [X] tCO2e (baseline minus project minus leakage) by the project during the verification period/reporting period [ref: dates]. This value is gross of the 10% buffer withholding based on the non-permanence risk assessment tool. This results in [X] tCO2e of credits eligible for issuance as VCUs in the above verification period...”</p>
CDM verification	<p>Example: “[Auditor] has performed a verification of the Afforestation/Reforestation CDM [X project] ... The reported GHG removals of the Project on the basis set out within the Monitoring Plan indicated in the registered Project Design Document ... comply with the methodology AR-AM0001/ version 02 ... The verifier assesses that the project is implemented and operated as planned and described in the validated and registered Project Design Document. Established forest being essential for GHG removals is operated reliably and is managed appropriately. The monitoring system is in place and the project is resulting in GHG removals. The verifier assesses that the monitoring was done in accordance the monitoring plan and the GHG removals in the Monitoring Report version [X] are calculated without material misstatements.</p> <p>We pointed out 1 Corrective Action Request and 24 Clarification Requests [...] Based on the information we have seen and evaluated, we confirm the following statement. :</p> <p>Reporting period: From [X to Y date]</p> <p>Verified GHG removals in the above reporting period: ... [X] tCO2 equivalents.”</p>
Kyoto Protocol review	<p>Excerpt from conclusions of KP review of countries’ initial report: “The expert review team concluded that the initial report of [X country] generally covers the elements required by paragraphs 5, 6, 7 and 8 of the annex to decision 13/CMP.1, section I of the annex to decision 15/CMP.1, and relevant decisions of the CMP; that the assigned amount pursuant to Article 3, paragraphs 7 and 8, has been calculated in accordance with the annex to decision 13/CMP.1; that the calculation of the commitment period reserve is in accordance with paragraph 6 of the annex to decision 11/CMP.1; and that the parameters for forest definitions are within the agreed range specified by decision 16/CMP.1. ... Based on [X country] base year emissions – {X} tCO2 eq, including the revised emission estimates provided in the energy, industrial processes, agriculture and waste sectors – and its Kyoto Protocol target of [X] per cent, [X country] calculates its assigned amount to be [X] tCO2 eq and its commitment period reserve to be [X] tCO2 eq. The Expert Review Team agrees with these figures.” [Note: Actual conclusions much longer; also note that the summary from the review of the annual GHG inventory submission is in a chart format that does not lend itself to a be displayed in this tabular format.]</p>

Key points:

- Evaluation approaches need to be chosen commensurate with the objectives of the evaluation, especially regarding the type and level of assurances of the GHG statement that the context requires.
- There are a range of different options when choosing methods for evaluating GHG statements: whether to assess data only or to include systems and processes; whether or not to spot-check primary

measurements; whether to check the correctness of measurements or to focus on definitions, methods and retracing calculations.

- For auditing firms, a risk-based auditing approach is an industry standard that is guided by the concept of materiality and relies on an auditing plan to identify those aspects of the GHG statements with highest risk to the correctness of GHG statements.
- In some schemes, reviewers have the ability to request corrective actions, clarifications (and forward actions) that GHG programs under review need to resolve before evaluations can conclude.
- Based on the evaluation methods and approaches chosen, evaluation summary statements can differ—and, ultimately, provide different outputs from the overall evaluations, including varying levels of assurances on whether the GHG statement is correct.

5 Profile of the evaluators

This section looks at the profile of evaluators, their challenges, and the environment that facilitates their work, including information on resources and logistics.

Auditing companies vs. expert panels: While several schemes (e.g. CDM, VCS, Guyana-Norway, CA offset program, Colombia REM) have drawn on the expertise of auditing firms, UNFCCC REDD+ technical assessments and several results-based payment schemes (e.g. the Amazon Fund) use small teams of technical experts. The FCPF Carbon Fund uses an expert-panel assessment for the validation of the GHG program design, including the GHG quantification approach and reference level construction, but intends to use an auditing approach for the subsequent verification of emission reductions.

Auditing companies approach validations and verifications differently than experts—focusing on ensuring that any errors in the data that are material to the quantification of results are documented and addressed. Such companies develop ‘auditing plans’, which identify critical issues, signed off by a risk manager who also approves the appropriate staffing (expertise) to conduct the review. They see the review statement—which provides a level of assurance on the results—as critical to their reputation, so have a strong interest in only verifying real emission reductions. By contrast, individual experts often have less accountability (not being part of a company with policies and procedures), and less incentive to insist on technical issues that may be contentious, as long as they retain their professional integrity. They may also be more interested in helping to build capacity or improve the methodology, rather than simply ‘auditing’ the numbers. In contrast, some auditing firms take a perspective that providing technical recommendations, during the verification process, to GHG programs would compromise the independence of the evaluation.

Liabilities: Auditing companies typically take on financial liability for any errors in their verification statement, which individual experts on panels are typically unable, unwilling, or not asked to do. For example, under the VCS, an auditor must submit a Deed of Representation, or a legal attestation that the entity has validated the project’s compliance with the VCS standard requirements or verified that the emission reductions (or removals) are in accordance with VCS rules. The liability for auditors is typically around two to three times the contract value. While this sum will not ‘break’ any auditing company, alongside the reputational risks, is a strong motivator to uncover ‘material’ issues. Because of this, auditing companies hold costly professional liability insurance, e.g. in the case of the CA offset program, entities that verify offsets are required to maintain a minimum coverage of USD 4 million of professional liability insurance, including for three years after completing verification services.

Conflicts of interest and undue influence: Several interviewees highlighted that it cannot be universally excluded that evaluators feel pressured to draw conclusions in the interest of the GHG programs being evaluated. While individual experts directly rely on working relations with countries, institutions, and other experts working in the climate change sector, auditing firms may shield their experts from such exposure. Moreover, such firms’ quality management systems safeguard against cases of undue influence, for example, by requiring another expert from within the firm—who was not part of the core verification services—to carry out quality control. To do this, a so-called technical reviewer conducts a second, independent review and concurs (or not) with the verification team findings. Similarly, documentation and public disclosure on communication during the evaluation (see discussion above) can protect against conflicts of interest.

In order to be accredited to perform verifications for carbon crediting schemes (e.g. CA offset program, CDM, VCS) companies show that they have policies and mechanisms in place to manage conflicts of interest. In the case of the

CA offset program, verification entities must identify the industries and locations where services have been provided and also monitor for potential conflicts for one year following provision of the verification services; potential conflicts are evaluated by independent third parties. Similarly, the CDM Executive Board sometimes conducts spot-checks on Designated Operational Entities. Nominations to the expert panels of UNFCCC or the FCPF include a screening for potential conflicts of interest, however this screening is not comparable to the rigorous evaluation that are required by some schemes and there are known cases where individual experts provide consulting services in the same field and to countries they are asked to evaluate. During interviews, some suggested that individual experts engaged in evaluations (e.g. under the UNFCCC and FCPF) are often guided by strong professional ethics. However, it was also noted by several interviewees that there is pressure to write evaluation reports that are ‘encouraging’ to (i.e. not to discourage) the country being evaluated—which results in weaknesses being downplayed in such reports. By contrast, auditing companies do not face such pressures (in fact, quite the opposite as noted above).

Technical skills of the evaluators: Even for world-leading experts, it is difficult to cover the range of skills required for a comprehensive evaluation of a GHG statement. A technical understanding is required on topics as diverse as earth observation, field-based forest inventories, statistics, as well as GHG reporting guidance. Moreover, skills in auditing techniques may be required to engage effectively with country teams. While auditing firms may provide minimum quality standards, the individual experts engaged in FCPF TAP assessments or UNFCCC evaluations are often asked to evaluate data and information that are beyond their immediate comfort zone or they may not complete a full assessment of the GHG quantification due to lack of competencies in particular areas.

Some schemes require proof of technical competencies. For example, the CA offset program requires the verification team to include a CARB-accredited verifier for the specific offset type (e.g. forestry) with at least 2-years of experience and has taken the CARB sector specific training and passed a CARB administered test. Under the UNFCCC, experts provide a résumé for screening by their national focal points, who make nominations to the roster of experts. Once included in the roster, for some evaluation tasks, e.g. regarding the technical analysis of REDD+ results (but not for the technical assessment of the FREL/FRL), experts need to take an official UNFCCC training course, hosted by the GHG management institute, and pass an exam to be eligible. There are no formal requirements for the FCPF TAP.

Table: Profile of evaluators.

	Evaluators	Expertise mobilized	Quality management	Evaluator liability
REDD+ UNFCCC	Team of two technical experts, selected from the UNFCCC roster of (LULUCF) experts	Includes both leading GHG inventory experts as well as nominees with lower levels of expertise	UNFCCC Secretariat is responsible for quality control; for the FREL/FRL assessment, more experienced experts provide advice	No liability
Colombia REDD+ Early Movers verification	Auditing firm	Team of experts: local expertise, technical aspects, environmental auditing, etc.	Auditing firm’s QM system, drawing on senior quality managers	Unknown
FCPF Carbon Fund TAP assessment	Team of 5-6 technical experts covering various topics (carbon quantification, safeguards,	Typically, one GHG measurement and reporting expert	Lead expert carries out quality control, FMT provides advice	No liability

	legal issues, etc.); usually 1-2 carbon accounting specialists			
GCF	iTAP selected from the UNFCCC roster of (LULUCF) experts	As above with UNFCCC REDD+, since drawing from the same roster	TBD	No liability
Norway-Guyana	Auditing firm	Team of experts: local expertise, technical aspects, environmental auditing, etc.	Auditing firm's QM system, drawing on senior quality managers	Related to contract value
Amazon Fund	AF technical committee comprised of national experts	AF technical committee includes representatives of civil society and recognized scientists from government research institutions	Unknown	No liability
VCS	Auditing firms, so-called validation/verification bodies (VVBs), approved by the VCS (and accredited by a different body, e.g., the CDM accreditation panel).	Team of experts: GHG/climate specialist and others specialists (e.g. biodiversity, social, etc.).	Auditing firm's QM system, drawing on senior quality managers, meeting competence requirements set out in ISO 14065	VVBs must provide a legal document (i.e. 'deed of representation') attesting to the review
CDM	Auditing firms, so-called Designated Operational Entities (DOEs), accredited by the CDM Executive Board	Team of experts: local expertise, technical aspects, environmental auditing, etc., including specific experience with the sectoral scope and the methodology	Auditing firm's QM system, drawing on senior quality managers	Related to contract value
Kyoto Protocol	Technical experts, selected from the UNFCCC roster of experts	Uneven, including globally leading experts as well as nominees with lower levels of expertise	UNFCCC Secretariat exerts quality control, more experienced experts provide advice	No liability

Quality management and consistency: Evaluations should be carefully managed to ensure a minimum level of quality and consistency among evaluations. Especially for schemes that create carbon credits or that provide financial reward, it is important to make sure that stated emission reductions are comparable to each other. Because of this, auditing firms that work according to ISO 14065 or follow other accreditation standards, such as in the CDM and the VCS, have involved procedures with lead auditors and technical reviewers that exert a quality assurance and a quality control function, regarding the individual evaluation and over the firms' entire portfolios. These go together with processes for raising complaints and appealing against the auditor's decision. The scheme administrators (the VCS Association and the CDM Executive Board) regularly carry out further quality checking of reviews and at times reject or adjust carbon credit issuance. In the case of the CDM, however, one of our interviewees perceived that such additional checks had grown to a point where it "pretty much became a new assessment, leading to bottlenecks and unpredictability in the system". Sometimes, the CDM Executive Board also conducts spot checks of the auditing firms where accreditations can be suspended in case of irregularities. There is also the possibility of legal recourse in case of contract breaches or negligence by auditing firms. The below chart shows the process of quality management for the case of the VCS.

Similarly, CDM and VCS methodologies include detailed definitions for parameters to be collected and project documents include corresponding tabular lists with guidance on how to present the information required. By contrast, no such tables are required for most REDD+ related evaluations (e.g. FCPF or UNFCCC), hampering basic cross-checking against independent data. Related to this, a lack of clear definitions on data items as basic as ‘deforestation’ or ‘net forest cover loss’, also make their assessment more resource intensive.

Levels of efforts allocated to the reviews diverge. Although the FCPF TAP assessments involve a larger team, there is typically only one GHG measurement and reporting expert, whose costs, including travel, are about USD 20,000 per assessment (USD 60,000-USD 100,000 for the whole team). The UNFCCC pays no fees for expert services and only provide travel costs for developing country experts. In this case, costs are borne by the experts or those paying their wages (e.g. governments who employ many of the experts on the UNFCCC roster)—therefore overall costs are unknown, but tend to be lower than for the FCPF TAP assessments. Validations and verifications under the CDM and the VCS cost from USD 20,000 to 60,000 each. Other audit reviews such as for the Guyana-Norway bilateral might be expected to be more expensive than the somewhat more standardized CDM/VCS reviews, costing up to USD 100,000.

Table: Organization and resources in evaluations.

	Centralized, desk-based or in-country evaluations	Duration	Level of effort and resources	Payments for evaluators
REDD+ UNFCCC	Centralized one-week assessment session with continued follow up over many months	10 or more months	Approximately 15 days each (or more) for two experts	No direct remuneration for technical experts; many are sponsored by their employer
FCPF Carbon Fund TAP	Desk-based assessment, plus one week in-country	About 6 months	Approximately 30 days for carbon accounting assessment	TAP compensated by the FCPF, ~USD 20,000 for the carbon accounting expert
Colombia REM verification	Desk review, plus three days in country	~6 weeks	Unknown	Unknown
GCF	Desk review	TBD	TBD	Cost of iTAP covered by the GCF
Norway-Guyana bilateral	Desk review and site visit	~2 months	50-60 days	Fees of up to USD 100,000, paid by Norway
Amazon Fund	Evaluation during AF Technical Committee’s half-day annual meeting	Unknown	Unknown	No direct remuneration; some may be sponsored by their employer
VCS and CDM	Desk review and site visit	>3 months	35-50 days	Fees ~USD 40,000-60,000 for validation and ~USD 20,000-40,000 for verification, paid by project proponent
Kyoto Protocol	Centralized review, including in-country visits at least twice in an accounting period	About 6 months	Approximately 15 days per expert in straightforward cases	No direct remuneration for reviewers; many are sponsored by their employer

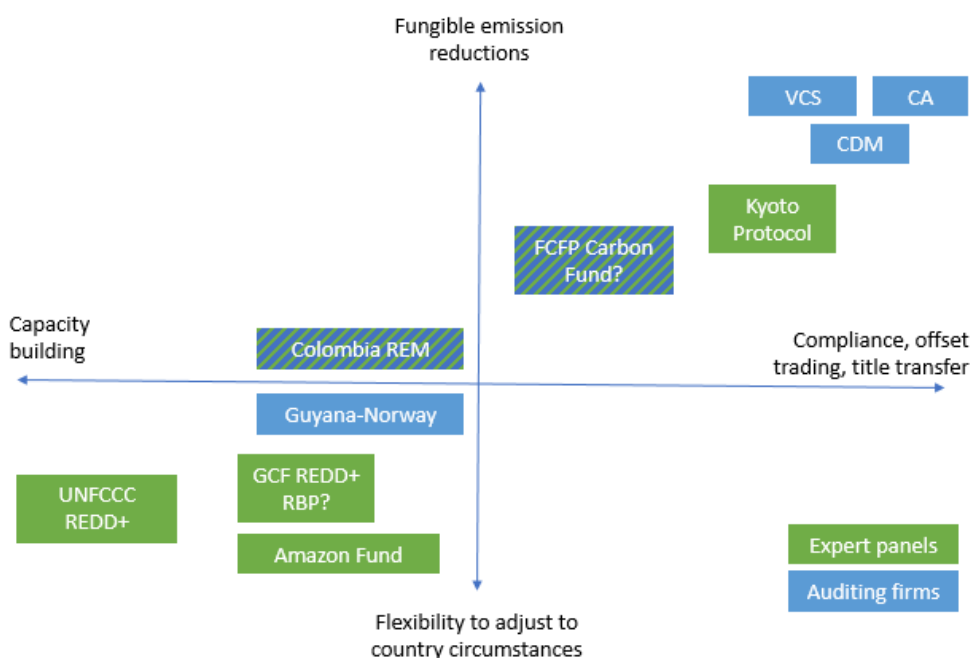
Key points:

- Evaluations conducted by expert panels compared to auditing companies are qualitatively different in their approaches; in addition, auditing firms often take on liabilities for their evaluation report and this may result in different dynamics.
- Avoiding conflicts of interest is an important element when designing evaluations and there are currently varying levels of robustness among schemes when tackling this issue, experts working as part of panels are especially vulnerable.
- There are also challenges in ensuring the right technical skills are available for evaluations. Procedures to ensure the evaluation team contains the right technical skills are mixed among the schemes we analyzed; provision of adequate resources can play a role, especially since it is rare for a single person to have the range of skills necessary to evaluate GHG statements.
- Consistency between evaluations is especially critical for those schemes that aim to create fungible units (e.g. carbon credits); auditing firms have dedicated quality management systems for this, while expert-led evaluations need to make alternative arrangements.
- While in-country evaluations provide opportunities for additional assessment (e.g. of primary data, standard operating procedures, etc.), centralized efforts allow for calibrations among multiple evaluators that may enhance comparability.
- Both the cost and efficiency of evaluations vary; resources allocated to evaluations of GHG statements are negligible compared to potential financial implications of results-based payments in almost all cases.

6 Lessons learned

The context and objective —i.e. why they are done and who decides how they are performed—often dictate how evaluations of GHG statements are undertaken. In particular, our study finds that the most significant variations in the design of GHG evaluation processes are driven by whether the focus is capacity building (often a key objective of evaluations for early stage, large-scale programs) or the desire to create fungible emission reduction units (e.g. to meet compliance obligations or to engage in offsetting).

There are, however, other issues that impact design—such as the entity that designed (and oversees) the evaluation process, which vary from multilateral bodies to donor governments to independent organizations. Sometimes there can be a lack of consensus on, or understanding of, the objective(s) of the evaluations (this is particularly true of the UNFCCC REDD+ technical assessments / analyses) or how they should be conducted (e.g. FCPF TAP assessments). There may also be a weaker mandate for some evaluations, e.g. donor governments often create processes whereby they are the final arbiter. By contrast, market-based and/or high volume systems tend to be highly regulated—both in content and on process—with evaluators clear on their mandate and approach.



Note: The GCF and FCPF have not yet evaluated emission reduction statements; thus, their placement in this diagram is debatable.

Figure: Trade-offs between stringency and flexibility of the evaluation and between the objectives of capacity building and verifying fungible emission reductions.

The differences described above (and illustrated in the Figure) can lead to variations in evaluation methods (from detailed auditing approaches using ISO standards, to high-level guidance only) and choice of evaluators (e.g. accredited auditing companies compared to panels comprised of subject-matter experts, often from rosters developed by organizations managing the process). And ultimately, they lead to quite different outcomes: on one end of the spectrum, expert-led approaches tend to be focused on building capacity, improving methods, and respecting the sovereignty of countries, and on the other, auditing approaches are more concerned with providing assurances of GHG statements.

Some evaluations have teeth, others less so. Auditing-style approaches under the CDM and the VCS include a mandate for validators and verifiers to raise corrective action and clarification requests that must be resolved before evaluations conclude, creating a strong incentive for GHG programs to collaborate. The Kyoto Protocol reviews allow expert review teams to apply adjustments to country estimates where there appear to be overestimates of emission reductions. The technical assessments, whether for the UNFCCC FREL/FRLs or for the Carbon Fund ERPDs, do not have such mandates.

Consideration of priority objectives should be the point of departure for designing any evaluation process, as there are often trade-offs. For example, in early stages when countries are still developing and refining forest monitoring systems, it may be beneficial for evaluations to be more flexible and focus on capacity building. Arguably, this will require sacrificing on the stringency of the evaluation and the comparability, or fungibility, of GHG statements, but can increase ownership and enable capacity building in a stepwise fashion. On the other hand, evaluations that can deliver strong assurances of GHG statements may be needed where they are used to underpin the issuance of fungible, market-based units.

6.1 What type of evaluations lead to stronger assurances of GHG statements?

The design of evaluations can have an impact on the transparency, accuracy, consistency, and comparability of the GHG statement and therefore strengthen the assurance of the statement. (Transparency, accuracy, consistency and comparability are the quality criteria for national GHG inventories as per Decision 24/CP.19. Our analysis below does not include completeness since this criterion has taken different meanings in different contexts.)

Highly transparent evaluation processes enable understanding of results, safeguard against conflicts of interest, maximize learning, and build trust in stated emission reduction.

External evaluations that are transparent and well understood—with clear processes in place and strong mandates—can better build trust on stated emission reduction. Third-party evaluations and their resulting assurances are the basis for trusting statements of achieved emission reductions, similar to their function in the finance sector where audits originated. Thus, it is not surprising that virtually all results-based finance and offsetting schemes we analyzed had an independent evaluation in place. There is large variation in the levels of transparency among the schemes we analyzed—in the case of carbon crediting schemes, much detail is available on both validation / verification results and the process, including communications between auditors and the GHG program, including details on corrective actions. The Guyana-Norway verifications were also notable in their level of transparency. Some schemes lack information on the process itself, and a few provide scant public information.

Transparency on the evaluation process can help safeguard against undue influence. Documenting and making public the communication between evaluators and GHG programs, including details around corrective actions requested and carried out, can also build confidence that the evaluation process was driven by independent judgment. While auditing processes systematically record such interactions, other evaluations do not record such communications and corrective actions, or they may keep such records but not make them public.

Risk-based evaluation approaches that lead up to declarative statements with a level of assurance and applying materiality thresholds provide greater understanding of the accuracy of stated emission reduction.

Evaluation schemes that are not explicit on the levels of assurance they are designed to generate, open the door to misinterpretation. For example, evaluations that are based on desk studies face challenges in determining whether emission reduction estimates are materially correct because their work rarely involves (or is challenged to involve from afar) reviewing primary data; therefore, summary evaluation statements provide much more limited assurance. Reviews that are more resource intensive and include in-country assessments and field visits with checking of primary measurements, on the other hand, are more likely to provide a higher level of assurance. Setting a materiality threshold and requiring an explicit statement on the level of assurance, as typically done by auditing companies in their review statements, may contribute to enhancing the accuracy of stated emission reduction.

Risk-based approaches to evaluation of GHG statements may also better improve accuracy of GHG estimates over broad-based evaluation approaches. Identifying areas of highest risk, and focusing the evaluation on such areas (compared to assessing every facet of GHG estimation) is a more effective tool for ensuring accuracy. Some schemes require evaluation of multiple criteria—some of which may not have a material impact on the estimation of emission reductions—both distracting the evaluator from critical issues, as well as creating an inefficient process.

Careful quality management, beginning with the procedure for selection and/or accreditation of evaluators, can improve the consistency among evaluations and therefore also among stated emission reductions—whether among different GHG programs or successive iterations of the same program.

Where evaluation schemes use auditing firms, they often devise procedures for their accreditation. To ensure quality, firms typically set up required management systems and have a set of processes to manage the quality of their work, including: systems for choosing and training experts, safeguards against conflicts of interest and to ensure impartiality, a set of standard procedures for conducting evaluations, and technical reviews by senior auditors. Expert-led evaluations (e.g. under the UNFCCC and the FCPF TAP) may also include efforts to manage quality, for example, the UNFCCC Secretariat often organizes centralized assessment sessions, allowing for exchange between experts and, in addition, conducts quality control on technical assessment / analysis reports. Although there is no quality control in TAP assessments, the FCPF FMT conducts 'calibration workshops' for TAP members to encourage consistency across assessments. Despite such efforts, shortcomings remain in some expert-led processes, with regard to providing consistent evaluations—both substantively (e.g. providing similar ratings or summary reports), as well as regarding quality management provisions.

Generating emission reduction statements that are fully comparable to each other requires detailed guidance and an auditing approach that dedicated service providers can deliver, but this may reduce the flexibility that diverse country circumstances require.

Detailed guidance can be a basis for developing comparable GHG estimates. Carbon crediting standards have invested in developing detailed carbon quantification methodologies and verification guidance since offsetting requires fungible emission reductions. By contrast, other schemes contain varying levels of ambiguities and gaps on key aspects of carbon accounting, forcing evaluators to make professional judgments with potentially significant implications. On the other end of the spectrum are expert-led technical assessment processes, not usually designed to achieve comparable emission reductions, where guidance and criteria on both the evaluation approaches and for GHG quantification are limited to a set of general principles with ample room for interpretation.

Auditing companies and processes have evolved to deliver standardized reviews, including of emission reduction statements. These companies have access to experts with technical, local and process expertise required and the necessary quality management systems in place to ensure consistent reviews and avoid conflicts of interest. Expert panels do not typically have similar provisions or processes in place, and are therefore less conducive to comparable GHG estimates.

6.2 What type of evaluations lead to stronger capacity building?

Evaluations can help improve the quality of GHG reports over time, especially when they are part of an iterative process with provisions for updating data and methodologies and when the evaluations focus on both systems and data.

Evaluations are an underused opportunity for capacity building on GHG measurement and reporting, both for those who receive feedback and for those who provide it. Receiving feedback enables learning and some GHG reporting and evaluations processes are designed for this. Similarly, many of those engaged in the evaluation process say that they also gain benefits from participating in the process, including the opportunity to interact with experts from other countries as they evaluate each other's GHG inventories. A couple of interviewees suggested that having a clear opportunity to receive finance for results provides important incentives to actually implement the feedback received.

Evaluations are most valuable for capacity building when there are provisions for updating data and methodologies in response to evaluation feedback and when they are iterative over a prolonged period. For example, reviews of UNFCCC GHG inventories have been in place for nearly two decades and such iterative reviews have contributed to the improvement in, and the current high quality of, developed countries' GHG reporting. Those involved in the Guyana verification processes all suggested that there was a strong capacity building impact from the five annual verifications that took place, stating that it was the consecutive nature and the continuity of structure that was important for effective learning. Such continuity of measurement and reporting efforts, and related evaluations, facilitates creation of processes and other institutional structures that are critical for enhanced capacity. By contrast, 'one-shot' evaluations (e.g. FCPF TAP) were deemed less valuable, with a few interviewees wondering if countries were likely to change systems or methods once they entered into the Carbon Fund portfolio. Similar concerns were raised regarding the case of UNFCCC reporting on REDD+ results where, despite theoretically available provisions for updating FREL/FRLs, the need to retain full consistency with the FREL/FRL approach when reporting REDD+ results in a BUR restricts future improvements, effectively 'locking in' countries at the time of reference level setting.

Evaluations that assess not only data and information, but also the systems, processes and institutions that generate the data are the most useful for building capacity. One interviewee suggested that evaluations that focus on "deconstructing and reconstructing data" (such as the UNFCCC REDD+ technical assessments / analyses and FCPF TAP assessments) could result in a positive conclusion, even though the overall process to generate the data was poor—or entirely done by external consultants. Looking into systems and processes not only helps to uncover whether the basis of errors is systemic but also, over time, are more likely to increase country ownership.

Some evaluations have recommendations on improved methods and data as the main result, but little is known yet on how effective they are in building capacity. The technical assessments of FREL/FRLs under the UNFCCC has building capacities as a principal objective and the identification of areas for improvement as its main result, but

there is no requirement to address any of the issues identified, nor does the evaluation lead up to a pass-fail judgement that would force change. And while many countries submit a modified FREL/FRL during the evaluation period and some of our interviewees stated that improvements were real, the extent of progress has not yet been investigated in a systematic manner. Similarly, whether subsequent FREL/FRL submissions will take into account the 'areas for improvement' is unknown since no country has yet submitted a subsequent (i.e. second) FREL/FRL. While we make an initial step towards better understanding how evaluations drive efforts for measurement, reporting and verification of REDD+, it may be useful to analyze empirically observed progress on capacity development and on the strength of assurance of GHG statements.

Finally, evaluations that allow for direct interactions between those doing the evaluation and those responsible for the GHG reporting tend to also be more effective at building capacity. For example, the UNFCCC Secretariat controls and manages all interactions between the country and evaluators. This includes managing calls as well as passing the technical assessment team's questions (after imposing quality control) to the country, and relaying country responses back to the assessors—reducing the amount of direct interactions. While this process protects parties (evaluators and those being evaluated) from over-reach, it also limits capacity building opportunities that could arise from a freer interaction.

6.3 Final considerations for the future...

Evaluations can be designed to maximize learning or to ensure progress towards GHG reduction targets -- over time those objectives should converge in the processes surrounding the Paris Agreement.

In the case of REDD+ emission reduction programs, there may be a need to strike a balance between clarity and detail of guidance with flexibility to accommodate learning in the early stages. Carbon crediting schemes (CDM, VCS, the CA offset program) have detailed methodologies and review processes—complexity is high, often requiring specialized technical skills. By comparison, REDD+ reports under the UNFCCC are only guided by broad principles, and technical assessments allow for flexible decision-making during evaluations. The assessments / analyses routinely justify shortcomings of GHG measurement and reporting in the spirit of 'stepwise improvement'—with an understanding that this allows for gradual building and improving of capacity for GHG measurement.

However, over time, it could be beneficial to increase alignment among certain schemes, including those intending to build capacity and to be flexible in adjusting to country circumstances, and those designed to create fungibility among accounted emission reductions and build trust in GHG statements. Because the Paris Agreement requires most countries to submit and report on quantitative targets, receiving payment for emission reductions or trading mitigation outcomes may require, over time and particularly after 2020, countries to align accounting for REDD+ emission reductions with accounting for their Nationally Determined Contribution (based on the national GHG inventory). Provisions in the Paris Agreement to avoid double counting may require additional consideration—beyond that of this report—on how evaluation processes could help bring these reporting streams in better alignment.

Those countries already engaged in results-based payments, whether as forest countries or as donors, are mostly making an effort to build off the UNFCCC technical assessment / analysis processes. For example, REM-Colombia draws on the UNFCCC FREL/FRL and verifications by an auditing firm assess consistency as most important criterion. The Amazon Fund may be revised to align with the UNFCCC FREL/FRL and REDD+ results reporting. Most recently, the GCF Board decided to build on the UNFCCC REDD+ reporting, adding additional requirements to

enable the allocation of payments. At the same time, some of our interviewees felt that more experiences are still needed, more evaluation required (on how robust emission reductions are that are developed on the basis of UNFCCC REDD+ reporting) and possibly more requirements agreed before a full reliance on UNFCCC processes may occur.

In the meantime, a priority should be to build capacity in developing countries—to measure and report forest-related GHG fluxes, and also to participate in evaluation processes. Although the UNFCCC processes are designed for creating a community of practice, both regarding GHG inventories and REDD+ technical assessments / analyses, participation from suitably qualified, funded and available experts is a current bottleneck. Accredited verification entities (e.g. auditing companies) are also mostly headquartered in developed countries. Some evaluation processes face funding constraints, and it may be useful to consider—given the considerable resources that are being made available for capacity building on measurement and reporting of GHG emissions—whether evaluations should also be prioritized in funding allocations.

Fully appreciating the value of different types of evaluations requires a stronger understanding of their processes than is common among REDD+ measurement and reporting experts. Evaluation approaches for GHG reports are well-developed, particularly at auditing firms, and there is also accumulated experience with UNFCCC evaluation processes (particularly for reviewing developed country GHG inventories, including for the Kyoto Protocol). Fully appreciating the dynamics of evaluations and their results requires an understanding of complex concepts: levels of assurance, the concept of materiality, risks posed by conflicts of interest, the role of liabilities, to name just a few. Many of these concepts are not universally understood among REDD+ experts, except for a few with first-hand experience in conducting audits, reviews, or verifications. This report is an effort to ‘unpack’ some of the concepts and improve understanding—but there remains considerable room for improved knowledge management and analysis of how best to evaluate GHG statements in the future.

7 Sources

The most important sources of information were interviews with experts, including those listed in the acknowledgement section (noting some chose to remain anonymous).

In addition, we consulted the following webpages for information on the evaluation schemes, the standards, guidance documents and samples of GHG reports and evaluation reports:

- UNFCCC REDD+: <http://redd.unfccc.int/>
- Colombia REM: <http://www.minambiente.gov.co/index.php/bosques-biodiversidad-y-servicios-ecosistematicos/vision-amazonia>
- FCPF: <https://www.forestcarbonpartnership.org/>
- VCS: <http://www.v-c-s.org/>
- CDM: <http://cdm.unfccc.int>
- Amazon Fund: http://www.fundoamazonia.gov.br/FundoAmazonia/fam/site_pt
- Guyana-Norway bilateral: <http://www.guyanareddfund.org/>
- CA offset program: <https://www.arb.ca.gov/cc/capandtrade/offsets/offsets.htm>
- GCF: [http://www.greenclimate.fund/documents/20182/820027/GCF_B.18_23 -
Decisions of the Board eighteenth meeting of the Board 30 September 2 October 2017.pdf/b55d8183-005c-4518-91dc-152113506766?](http://www.greenclimate.fund/documents/20182/820027/GCF_B.18_23_-_Decisions_of_the_Board_eighteenth_meeting_of_the_Board_30_September_2_October_2017.pdf/b55d8183-005c-4518-91dc-152113506766?)

Other important sources of information were the following:

- FCPF. 2017. Practical Application of the Forest Carbon Partnership Facility Methodological Framework - Carbon Accounting: Issues and Recommendations.
- UNFCCC. 2016. Technical assessment process for proposed forest reference emission levels and/or forest reference levels submitted by developing country Parties – Synthesis report by the secretariat.
- ISO 14064-3, ISO 14065, ISO 14066 and ISO 17011.