GOLD STANDARD FOUNDATION

Safe Water Supply Grievance

Technical Advisory Committee Grievance Working Group

INVESTIGATION REPORT

Final

27 November 2020

SECTION 1 – EXECUTIVE SUMMARY

GOLD STANDARD FOUNDATION

Safe Water Supply Grievance

Technical Advisory Committee Grievance Working Group

INVESTIGATION REPORT

Final

SECTION 1 –
EXECUTIVE SUMMARY

The Gold Standard Foundation (GS or Gold Standard) Technical Advisory Committee (GS-TAC) Safe Water Supply (SWS) Grievance Work Group (TAC SWSGWG) is pleased to submit to you and the TAC this full report of the investigation. The report sets forth the process followed to investigate the matter, conclusions and recommendations from the Technical Consultant Berkeley Air Monitoring Group (BA), and TAC SWSWG Report Commentary.

The report consists of three sections plus one annex;

Throughout this investigation and in developing the findings summarized below, the TAC-SWSGWG has been guided by the following Gold Standard core principle. *"The Gold Standard rules are interpreted in accordance with the Standard's core principles of fairness, reliability, conservativeness and pragmatism. Where a rule has unintended consequences, the relevant Gold Standard bodies will work with the project to ensure that The Gold Standard's values are upheld and enforced."*

The TAC-SWSGWG appreciates the opportunity to assist The Gold Standard in this important matter.

Scott Harder, SWSGWG Chair
Liza Murphy
Steve Thorne
Jessica Wade-Murphy

Background and Scope of the Investigation

The Grievance allegations are detailed in SECTION I.A. of the BA Report. Grievances are a standard component of the Gold Standard's operating procedures, providing a standardized and transparent process for addressing them. The grievance process may result in both retroactive and forward-looking actions.

The scope of the investigation is governed by Terms of Reference set forth as Section VIII (APPENDIX 1) of the BA Report. This appendix also presents a description of the SWS Grievance. Both the SWS Grievance and Terms of Reference are available at:

(https://www.goldstandard.org/our-work/grievances/potential-over-issuance-risk-safe-water-supply-project-pipeline).

The allegations to be investigated are as follows:

1 – Application of the TPDDTEC Methodology for SWS may lead to over-issuance of VERs; and

2 – The Gold Standard for the Global Goals (GS4GG) scheme may not have, prior to this review, dealt with over-estimation issues at the optimal point in the certification process or in an optimal timeframe.

The scope of the investigation was divided into two components, to correspond to the main allegations described above.

Part A: Analyse the technical aspects of the allegation to assess its veracity. If the investigation finds no material basis in the allegation, the investigation will not proceed further. If the investigation finds basis in the allegation, that is, that material over-estimation has occurred, Part B will be pursued.

Part B: Assess and provide analysis and recommendations on response(s) to the findings of part A.

The BA investigations used independent means and methods as set forth in SECTION III of the BA Report.

During the course of the investigations, it became clear that certain tasks set forth in the TOR were significantly larger and more complex in scope than initially thought. To ensure that conclusions as to Part A matters could be developed expeditiously, the following tasks were, upon consultation and approval of this WG, set aside.

- Systematic review of audit reports.
- Systematic comparison of parameter values across SWS methodologies, schemes, and mechanisms.
- Detailed estimation of the volume of over-issued credits, if overestimation is deemed to have occurred.

The SWSGWG received comments relative to the technical analysis and SWS processes and procedures from SWS stakeholders. These comments provided additional information intending to clarify and elaborate various elements of the SWS scheme. In some cases, the comments raised conflicting points of view with what the WG gathered during the interview process. Many of these comments are addressed in the appropriate report sections. Please also refer to the Stakeholder Comments Annex attached to this report for further information.

Part A - Findings

The key findings of the technical investigation conducted by BA have been framed in reference to a working null hypothesis: that there was no over-estimation of emissions reductions in the GS SWS portfolio. BA then tested that hypothesis by applying its methods to either accept or reject it against available evidence. Part A findings are set forth in SECTION IV of the BA Report. These findings are summarized here.

Project Users. The analysis of reviewed literature values suggest that the number of users reported by project developer values are **high**, as does the comparison with

manufacturer recommended specifications. There is likely overestimation due to this factor, especially considering that this parameter has a strong multiplier effect on the overall project emissions reductions estimates.

Project Usage. The 100% usage percentage applied to most GS SWS projects is **higher** than that supported by the literature.

Project Functionality. GS SWS projects report **significantly higher** functionality than is seen across the literature, although emerging evidence shows the potential for some incentive-based programs to minimize downtime. Given the impact this parameter has on the overall ER estimates, it is very likely responsible for some degree of overestimation.

Baseline Water Boiling Fuel Requirements. For many projects, baseline fuel consumption estimates are **higher** than what would be reasonably expected for project homes to boil water, thus contributing to overestimation of emission reductions.

Borehole Treatment Capacity. Current project estimates appear to be based on a non-conservative best-case scenario, resulting in **significantly higher** capacity estimates--possibly as much as 100% overestimation in some cases.

Degradation of Water Quality between Boreholes and End Users. For community-based systems, such as borehole programs, neither microbiological nor chemical quality are ensured for project users under the current protocol, thus creating a risk that the development benefits of the SWS project portfolio are being overestimated.

Distance to Boreholes. This is a significant parameter affecting estimates of emissions reductions. There is a moderate concern that the lack of measurement guidance for this parameter may allow for overestimation of both emission reduction and other sustainable development goal development benefits.

Safe Water Quantity Estimates Per User. The quantity of safe water per project user is based on a water consumption field test targeting three credited uses: drinking, basic personal hygiene, and food preparation. The definition of the three credited uses is not well defined, however. Many reported projects test results exceed the capped value of 7 liters per person per day. This capped value is reasonable based on the results of the WCFT's reviewed in this investigation.

Allowance	**Average Project Estimate** tCO2e/yr	**Non-Conservative Comparison Estimate** tCO2e/yr	**Regular Comparison Estimate** tCO2e/yr	**Non-Conservative Difference**	**Regular Difference**
No Cap	16,259	6646	2101	145%	647%
Cap	10,000	6646	2101	50%	376%

Rough Estimate of SWS Borehole Over-Issuance. Total emission reductions credits issued for micro-scale borehole projects in the SWS portfolio to date was approximately 3,150,000 tCO2e. A conservative range for the over-issuance of micro-scale borehole projects' emission reductions was estimated by applying the overestimation percentages

for the difference comparison estimates for capped projects shown above, implying that a range of approximately 660,000-2,100,000 tCO2e in reductions would have been more realistic (1,050,000-2,490,000 tCO2e difference).

Materiality of Over-Estimation. The BA analysis concludes that there is a **pervasive, possibly systematic, over-estimation of emission reductions in the SWS borehole portfolio** as described above. The overestimation is judged to be systemic as it is ingrained in multiple components of the methodology and its application. Further, the recent rapid growth in the SWS portfolio has contributed to the materiality of the overestimation, with almost 60% of SWS projects having been added to the portfolio in the last 2-3 years.

Findings – Part B

Part A investigations were the primary focus of the BA Report and the enabling TOR. The TOR also requested investigations into the certification processes that underpin the development of the GS SWS project portfolio. Specifically, these processes include: 1) the SWS methodology itself, 2) the application of the methodology in the design of SWS projects, and 3) project certifications. These processes were examined through document reviews and interviews with key stakeholders to identify potential cause(s) of overestimation and the degree of responsibility that may be attributed to specific stakeholders or parties within the GS ecosystem. BA presented these findings in PART V of the BA Report. It should be noted that these are observations rather than firm conclusions due to significant time constraints on the investigation.

Further, the SWSGWG has generated additional comments on the processes and procedures that have contributed to the Part A findings in a separate section of this report. To Summarize the Conclusions related to Part B, the investigation found weaknesses in the methodology, its application, project aggregation practices, and the Assurance System designed to protect the quality of credit issuances.

SWS Methodological Weaknesses. The methodology was judged to not be prescriptive or restrictive enough for some specific baseline, project boundary, and project impact parameters and not reflecting the most up to date best practices as supported by literature and practices.

Field Measurement Requirements. The methodological basis for requiring field measurement of specific parameters may be based on an outdated perception of a limited evidence base. More than a decade after the creation of the methodology, some parameters have been well researched, and the resulting literature reveals values that fall within a narrower range than may have been expected.

Application of the Methodology by Project Developers. A second type of process flaw appears to occur when project developers apply the methodology. It is the norm to have a chain of parties implementing project monitoring requirements, which leaves multiple connection points where misreporting can occur. Monitoring of these “weak links” may be not as rigorous as needed.

Project Aggregation. There is general agreement that the majority of SWS projects being certified under the micro-scale scheme do not reflect the scheme’s original intent which was to provide pathways to the carbon markets for small projects that would not otherwise be viable. The current system allows project developers to aggregate multiple smaller “cloned projects” into conglomerates with total annual tonnage that would usually

trigger more robust certification requirements, such as those applied to large-scale projects. This circumstance is subverting the risk mitigation measures designed to ensure that large-scale projects do not over-estimate climate or development benefits.

Misalignment of Auditor/Reviewer Skillsets and Validation Demands. There is a potential misalignment of attitude, skills, and training of auditors and reviewers with the demands of providing a substitute for third-party validation of project design and verification of project performance required of larger projects. Auditors and reviewers are aware that the reasonableness of the resulting parameters is also their concern, yet in our limited assessment, there is significant variation in the norms around how to assess reasonableness and what actions to take as a result of this assessment.

Maintenance of the Standard. The processes for updating the standard, and issuing interim guidance while updates are being developed is ineffective and slower than needed. This is further impacted by limitations of retroactive adjustments to estimates made using older requirements thus allowing overestimation to remain in place.

The Assurance Model and Its Intent. As noted earlier the original purpose of the Microscale projects and the Assurance practices to support credible certifications no longer matches in some cases the application of the design.

Application of the Assurance Model by SustainCERT. We discuss several aspects of the approaches SustainCERT has taken in applying the GS Standard and Assurance requirements that we think could be improved.

Oversight by GS. We discuss areas that GS could improve in their oversight of both SustainCERT and of those using the Standards.

Technical Parameters. We note that there may be legitimate variance in chosen values used in the assessment and refer to the need to develop an organizational point of view on choice of these default values.

The subsequent sections of this report provide further details on each of these points and lays out more fully the conclusions reached.

The SWSGWG wishes to thank the team at Berkeley Air for their hard work, their professionalism and their delivery of a report that clearly lays out their investigation and their conclusions. We also wish to thank all those who have participated in any aspect of this investigation, in almost every case information was offered transparently, schedules were accommodated and a commitment to the mission of the Gold Standard was evident. The SWSGWG and BA stand ready to continue to support the TAC and GS in answering any questions they may have about this report.

Working Group Response to TAC Clarification Question

In the course of its review of an internal draft of the Safe Water Supply Grievance Technical Advisory Committee (TAC) Working Group (SWSGWG) Report, the Gold Standard Energy TAC (E-TAC) requested a response to the following clarification question on 25 September 2020.

From the independent report (p.52), BAG sets context of:

*The Gold Standard rules explicitly place the **responsibility of reporting truthful and reasonable parameter values on developers**, and our investigation did not uncover any evidence of **intentional** misreporting. (bolding added)*

The reporting of data is consistently and substantially (647% above reference values, referring to Table 6) above the reference parameter estimates that BAG characterises as 'realistic' based on credible and recent literature.

Recognising this was not the primary purpose of the investigation, is the independent analysis able to ascertain if the parameters reported in Table 6 are likely to be reasonable and true?

If a conclusion cannot be drawn, what further investigation or analysis would be needed to determine this?

The requested response is below.

From: XXXXXXXXXXXXXXX
Subject: SWS Working Group Response to Clarification Question
Date: September 29, 2020 at 1:18 PM

Attached please find the SWS Working Group response to the 25 September Clarification Question from TAC. Please forward the document to the rest of the TAC members.

Best regards,
Scott Harder, Chair
SWS Working Group

We are unable to determine if the reported project estimates of parameters shown in Table 6 are "likely to be reasonable and true". As we note in the report:

- The data ranges that were reviewed were allowable in the Standard, and monitoring practices, as reported, were within the scope of the methodologies and were accepted in the assurance process.
- Many of the parameter values we reviewed vary substantially from the expectations based on literature, available research and interviews with content experts. This degree of variation would be expected of an outlier data point, but instead, values in the outlier range were common in the borehole project portfolio across many different geographies and time periods.
- A significant outlier is the parameter representing the number of days the borehole technology is operational. For the overwhelming majority of monitoring periods we reviewed, most projects report all their boreholes functioning 100% of the time. This result represents a major departure from the bulk of the evidence of borehole durability and maintenance achievements to date.

We recognize that the rules require developers to take responsibility for the reasonableness of their monitoring data and incorporated its intent into our review efforts. We found a range of interpretations by different developers that were accepted by the assurance system. In addition to the matters addressed above, we also note that it is improbable that the highest-end values of each parameter would be consistently

observed across projects, given the impact of underlying operations and field test realities.

In consideration of your request for suggestions for further investigation pathways, we offer the following:

- **Engage an independent investigation and auditing body to conduct a series of on-the-ground and desk based audits** to determine both the replicability and accuracy of the responses submitted and approved, as well as the likelihood of realizing the consistently high results of each parameter identified in Table 5 and typified by the case studies in Table 6.
    - There would be limitations given the time that has passed, thus it could be advantageous to look at both the specific projects that we reviewed and those more recently added to the portfolio.
    - The investigation could include the replication of reported tests, surveys, and operational data collection and any other factors the investigator found useful.
    - A protocol for determining the qualifications of the auditing body should be developed in advance of the engagement. ISO and other accreditation qualifications should be considered.
    - A desk based review of all project documentation to identify if there was any conflicting data reported and whether or how the assurance process engaged with the questions of true and reasonable would complement the on the ground reviews.
- **Conduct a review that compares processes, safeguards and results between Project developers** to understand the differences they may apply to data gathering and reporting in a system of multiple handoffs that could impact the quality of the data gathered and reported.

We close with an additional observation. The issues identified in the report are serious and need attention, we feel that they are addressable with the right effort. The development of MicroScale projects and the mission of the Gold Standard Foundation are immensely valuable and we trust the organization will continue to fill this vital need. With an approach of continuous improvement and transparency in the identification and resolution of challenges, we believe that will happen. We strongly support an additional investigation to confirm the veracity of SWS project submissions coupled with an effort that identifies and adapts internal processes as required. This is a complex matter that has presented an opportunity to review and institute important process improvements regardless of the veracity of the underlying data; but one that concurrently must not avoid a robust exploration of the data on which so much depends and so many rely.

BERKELEY AIR MONITORING GROUP

protecting health and climate

INVESTIGATION OF THE GOLD STANDARD GRIEVANCE,

"Potential over-issuance risk from Safe Water Supply projects"

ORIGINAL: SEPTEMBER 2020

REVISED: NOVEMBER 2020

Berkeley Air Monitoring Group | 1900 Addison St. #350 Berkeley, CA 94704 | Office phone: +1 510.649.9355 | FAX: +1 510.662.1970

Acknowledgements

Berkeley Air Investigation Team

Michael Johnson, Technical Director

Christina Barstow, Consulting Water Expert

Dana Charron, Managing Director

David Pennise, Senior Scientist

Madeleine Rossanese, Project Associate

Nicole Keeney, Intern

We are grateful to the Gold Standard Secretariat and Gold Standard Technical Advisory Committee - Safe Water Supply Grievance Working Group for their on-going support throughout this assignment. We also extend our thanks to experts and stakeholders who made time for interviews and provided written input and documentation (see Appendix 2 for list). Finally, we recognize the stakeholder input invited by the Gold Standard Foundation in response to the initial draft of this report and provided in its entirety as an annex to the final version.

Cover photo credit

DFID - UK Department for International Development
https://www.flickr.com/people/14214150@N02

Acronyms

BA	Berkeley Air Monitoring Group
BWBT	Baseline water boiling test
CFU	Colony forming unit
CO2e	Carbon dioxide equivalent
CoV	Coefficient of variation
DOE	Designated operational entity
ER	Emissions reductions
fNRB	Fraction of nonrenewable biomass
GS	The Gold Standard Foundation
HWT	Household water filtration technology
ISO	International Organization for Standards
PoA	Programme of Activities
SSA	Sub-Saharan Africa
SWS	Safe Water Supply
TAC	Technical Advisory Committee [for the Gold Standard]
TAC-SWSGWG	Technical Advisory Committee - Safe Water Supply Grievance Working Group [for the Gold Standard]
TPDDTEC	Technologies and Practices for the Displacement of Decentralised Thermal Energy Consumption (GS methodology)
TSF	Three-stone fire
VER	Emission reduction credit
VPA	Voluntary project activity
VVB	Validation/verification body
WBT	Water Boiling Test
WCFT	Water consumption field test

Table of Contents

I. Introduction

This report contains the results of the independent investigation by Berkeley Air Monitoring Group (BA) of the grievance, "Potential over-issuance risk from Safe Water Supply projects" filed on July 1, 2020.

A. Grievance

According to The Gold Standard Foundation (GS) website, "[the] certification process is designed with several checks to ensure that projects remain in conformity with Gold Standard rules and requirements for the duration of their project activity. Gold Standard welcomes complaints from any stakeholder."

The current grievance was formally submitted to the GS Secretariat by the GS Technical Advisory Committee-Working Group. An internal review by the Secretariat suggested evidence of overestimation and subsequent over-issuance of emission reductions credits from Safe Water Supply (SWS) projects, forming the basis of the request for an independent investigation. The full text of the grievance is provided below.

> *"As per internal Standard Operating Procedures, Gold Standard Secretariat regularly conducts quality assurance and control (QA/QC). In the spring of 2020, Gold Standard conducted QA/QC for Safe Water Supply projects. The Secretariat identified a risk of potential over-estimation of emission reductions and therefore risk of over-issuance of credits and raised this to a working group of our independent Technical Advisory Committee (TAC) as the risk was material.*
>
> *Following review and deliberation, the TAC working group initiated a grievance, mandating: 1) an independent investigation of the Safe Water Supply methodology and how it is being applied by projects, and 2) a temporary limit on emission reductions credits (VERs) from Safe Water Supply projects until the investigation is complete and recommendations are implemented, to avoid potential over-issuance. The interim limit on emission reductions is based on two parameters:*
>
> 1. *Firewood consumption to boil 1 litre of water for 10 minutes will be capped at 0.400 kg for three stone firewood baseline stove scenarios. For other baseline fuels, projects will be assessed on case-by-case basis. AND*
> 2. *For borehole projects, the number of users per borehole will be capped based on specifications from the borehole technology supplier/manufacturer.*
>
> *The investigation will be overseen by a qualified TAC sub-committee comprised of members having no conflict of interest. This TAC sub-committee has appointed an independent, qualified expert, Berkeley Air Monitoring Group to lead the investigation. The final Terms of Reference for the investigation are posted in the supporting documents below. To ensure independence, Gold Standard Secretariat will not take part in the investigation itself but will support TAC with administrative support.*
>
> *This grievance process intends to prevent over-issuance. It is initiated as part of Gold Standard's commitment to environmental integrity, good governance, and transparency to reach appropriate and fair resolutions. The outcomes of the process will inform measures to resolve and remediate any issues that the investigation confirms."*

Grievances are a standard component of the Gold Standard's operating procedures, and the organization offers a standardized and transparent process for addressing them on its website. One of the primary differences between the grievance process and other types of actions designed to amend the Gold Standard's certification processes is that the grievance process can result in both retroactive and forward-looking action whereas other rule changes can only be put into effect prospectively.

B. Scope of investigation

1. Summary of Terms of Reference

The investigation scope is guided by the terms of reference (see Appendix 1), published as Step 4 of the Gold Standard Grievance Process.[1] The investigation is overseen by the Technical Advisory Committee - Safe Water Supply Grievance Working Group (TAC-SWSGWG).

a) Investigation tasks

The primary tasks required by the TOR are detailed below. These tasks may run concurrently, and not all of them are expected to be addressed in this report.

- Shortlist the key issues which will be further investigated due to potential risk of non-conformity against the relevant GS methodologies or breach of any associated documents, such as the Terms and Conditions.
- Review standards, certification and project documents including internal reviews, validation reports, monitoring reports and verification reports completed by auditors.
- Interview relevant stakeholders as identified by the investigation team, including but not limited to Gold Standard Secretariat, SustainCERT, project developers, auditors, TAC members, experts and other stakeholders as required.
- If necessary, conduct a field study including but not limited to fact gathering, assessing legal findings, interviewing affected stakeholders and others to support the review work. This will be at the discretion of the independent investigator and GS-TAC WG who should assess based on whether a field visit(s) is the only viable option to determine the issues under investigation.
- Within the above analysis, determine to what extent the information is sufficiently complete in order to make recommendation(s).

b) Research questions

The investigation's research questions are divided into two parts: "Part A" outlines the technical analysis of the grievance allegation to assess and characterize its veracity; and "Part B" discusses and provides recommendations on potential responses to the Part A findings. This report aims to answer the research questions #1-3, which constitute Part A,

[1] https://www.goldstandard.org/our-story/grievances-deregistration

and to comment as appropriate on questions #4 & 5, which form the initial section of Part B. All five questions are presented below.

"Part A:

1. *Within the assumptions of suppressed demand within the methodology(ies), is the perceived over-estimation (and thus over-issuance) within the SWS portfolio correctly identified (that is, is the perceived issue 'real' – has over-estimation &/or over-issuance occurred), noting that over-estimation and over-issuance are related but not the same topics (for example the potential resolution for over-issuance may not also resolve over-estimate, which may require other standards, methodological and procedural change)?*
    a. *If not, investigation ends with recommendations on analysis approaches and QA/QC process.*
2. *If so, to what degree? That is, what is the likely amount of over-issuance per credit issued? Ideally this can be quantified (as 'x%'), but may require approximations (eg: quantitative ranges 'up to x%' or 'between y and x%') and/or qualitative assessments (eg: 'insignificantly, somewhat, significant, substantially' etc).*
3. *Does the degree of difference represent a material over-estimation &/or over-issuance?*
    a. *If not, investigation ends with recommendations on analysis approaches and QA/QC process.*

Part B:

4. *If material difference is identified, where in the GS process of standards-setting, methodology development and updating, project documentation establishment, project monitoring and/or validation/verification/certification or any other external, internal or combination of factors do the cause(s) of over-estimation &/or over-issuance lie?*
5. *Where cause(s) can be identified, what degree of responsibility (if any) can be ascribed to the various parties and stakeholders associated with over-estimation &/or over-issuance? Note that any ascribed responsibility does not imply culpability or intent, rather it aims to identify where error(s) may have occurred &/or how they interact &/or how they are propagated. It may also identify where processes could be adjusted to prevent similar results in the future."*

c) Terminology

The TOR refers to both "over-estimation" and "over-issuance," using the terms interchangeably in some sections and specifically in others. "Over-estimation" is accepted to be the broader term most relevant to the implementation of the technical processes used to quantify the tons of carbon offset in each crediting period. "Over-issuance" refers more specifically to the case where over-estimated reductions have already been verified and certified. For the purposes of this report, we have used "overestimation" in most of the narrative, in keeping with the primary focus on the technical aspects of offset

quantification, except for a few instances in Part B, where “over-issuance” is specifically relevant.

d) Investigation timeline

The TOR specified a five-week timeline for the completion of the investigation and submission of a draft report. After further assessment of the complexity of the scope of work, a further three-week extension was provided, for a total timeline of eight weeks.

2. Exclusions

As the investigation progressed, it became clear that certain tasks specified or alluded to in the TOR were significantly larger in their scope than initially perceived. At the same time, all parties recognized the urgency of reaching draft conclusions on the Part A research questions, in order to allow the investigation to reach an expeditious close. Therefore, the Berkeley Air team and the TAC-SWSGWG agreed to postpone the following tasks:

- Systematic review of audit reports;
- Systematic comparison of parameter values across methodologies, schemes, and mechanisms; and
- Detailed estimation of the volume of over-issued credits, if overestimation is deemed to have occurred. Instead, a range of the potential overestimation as a relative percentage was estimated based on average parameter values in the portfolio. Additionally, a rough first-approximation of over-issuance was provided.

The deferral of these tasks did not prevent the Berkeley Air team from completing a full investigation of the research questions in Part A of the TOR or from providing observations in response to the relevant Part B questions.

C. Berkeley Air Monitoring Group as independent investigator

Berkeley Air aims to provide independent, solution-neutral desk, laboratory, and field assessments in the household energy sector and related fields to a range of implementers and funders, working with a global network of research partners. Our broad vision is to enhance global health and climate through the provision of high-quality research that supports data-driven activities and policies. Since our incorporation in 2009, we have completed projects in more than 25 countries, and our staff has contributed to the authorship of over 50 peer-reviewed research publications specifically related to household energy, as well as dozens of government and NGO reports. Prior to operating as an independent consultancy, the group existed within the School of Public Health, University of California, Berkeley, as an applied research center.

Berkeley Air Monitoring Group, Inc., is a social enterprise registered in good standing in the state of California, USA (EIN: 26-3881064). Our company generally employs five or six scientific experts, two management employees, and numerous interns, fellows, and regular consultants in a matrixed organizational structure that facilitates a project-driven workflow. Our clients span government, private sector, and non-governmental entities, and include many major multinational organizations in our sector, such as the World Health Organization, the World Bank, and the UN Foundation’s Clean Cooking Alliance, as well as national development funders, like US National Institutes for Health, UK Department for

International Development, and Deutsche Gesellschaft für Internationale Zusammenarbeit. Our work was recognized by the Partnership for Clean Indoor Air in 2007 and we were longlisted for an Ashden Award in 2018.

Berkeley Air has been associated with and an engaged stakeholder of the Gold Standard since approximately 2007, when we first contributed technical suggestions to the development of a methodology for quantifying and monetizing the climate impacts of improved cookstoves and assisted Climate Care in the development of a project in Uganda. Since then, our staff has regularly participated in unpaid stakeholder consultations and technical advisory task forces to assist the Gold Standard in the development of robust methodologies and protocols. We have also worked with Gold Standard as a paid consultant on two prior reviews, the first of Vestergaard Frandsen's GS water filter project in Kenya in 2013 and the second of usage survey methods used in a CO2Balance's initiative entitled, "Improved Kitchen Regimes Multi-Country Programme of Activities" in 2017. Additional detail on all of these engagements is presented in Appendix 3. Finally, Berkeley Air has also carried out monitoring for projects monetizing climate and development impacts under both the Gold Standard and other mechanisms.

II. Background

A. Characteristics of the SWS project portfolio

At the start of the investigation, Berkeley Air was supplied with a database and summary document containing information publicly available on the Gold Standard Registry pertaining to each project in the Safe Water Supply project portfolio. Of 503 total projects within the SWS scope of the investigation, over 82% of the projects are certified under the micro-scale scheme, subject to the certification process specific to that scale and to the annual emissions reductions cap of 10,000 tons CO2e. Approximately 96% of total projects are grouped under Programmes of Activity (PoA), mechanisms that agglomerate individual projects under a programmatic umbrella based on common characteristics.

While there are many approaches to safe water supply globally, 90% of the projects in the SWS portfolio involve borehole installation or rehabilitation, and the remaining 10% are split between other water treatment technologies, such as tabletop water filters, and community-level clean water kiosks. Almost all of these projects were in rural areas where wood was the baseline fuel, though a small number of baseline scenarios included charcoal. The total portfolio has projects located in 24 different countries, but 95% of all registered and issued projects are in 11 countries in sub-Saharan Africa.

The first Safe Water Supply project entered the pipeline in 2010, and the six years that followed saw steady growth of 5-25 projects per year being registered under this category. In 2017 and again in 2020, there has been a sharp increase in the number of projects registered to the pipeline—about 60% of the total projects in the portfolio were registered in the last two to three years. Of the new projects, almost 90% were registered to the five PoAs that house the most projects, leaving those five PoAs with about 85% of the total project portfolio.

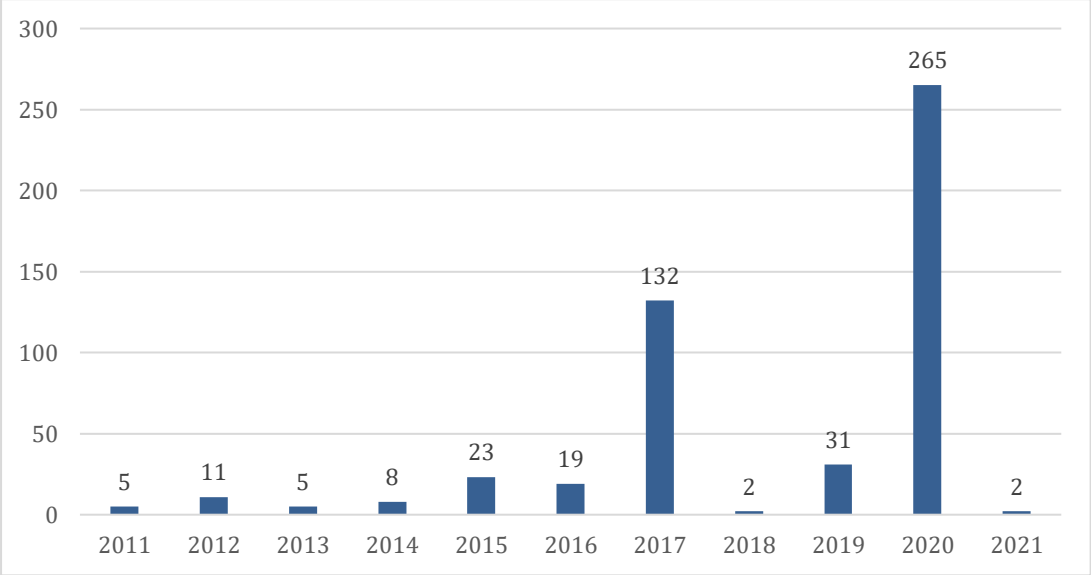


Figure 1. Year-wise distribution of the safe water supply (SWS) project pipeline. ***(Note: year represents the crediting period start year.)*** **Source: GS Secretariat.**

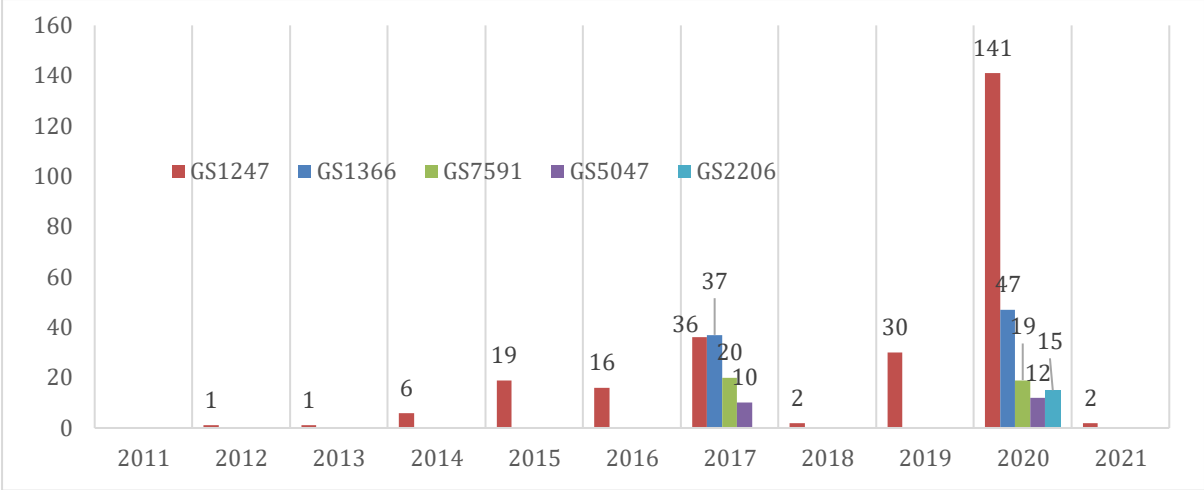


Figure 2. Year-wise distribution of safe water supply projects in the top five PoAs. *(Note: year represents the crediting period start year.)* Source: GS Secretariat.[2]

B. Summary of Gold Standard project certification system

The Gold Standard sets standards that allow the implementors of climate and development projects to quantify and certify the impact of their activities in a recognized, comparable, and transparent manner. Through the certification process, implementors are able to obtain certified emissions reductions credits and/or impact statements that are listed on GS's public registry and can be commercially traded. Impacts statements are commonly issued in reference to progress towards the Global Goals for Sustainable Development.

GS standards are implemented through methodologies, which are developed and revised by subject matter experts based on new sector learnings. Methodologies progress through a consultative process of independent peer review managed by the GS Secretariat and are ultimately approved by the GS-TAC. Safe water supply projects follow the methodology entitled, "Technologies and Practices for the Displacement of Decentralised Thermal Energy Consumption (TPDDTEC)". The TPDDTEC methodology covers a wide range of technologies and practices, from household-level approaches, such as energy efficient appliances, water filtration and treatment, to community-level borehole projects.

The current version of TPDDTEC methodology is 3.1; however many projects are still following earlier versions 1 and 2 (see Table 1 for details). Each Voluntary Project Activity (VPA) continues to follow the methodology that it was originally started under until both its parent PoA begins a new crediting period with an updated methodology, and the VPA itself reaches the start of a new crediting period.

[2] A correction to the year-wise distribution of SWS projects in POA GS5047 was received and is available in the complete record of stakeholder comments included as an annex to this report.

Table 1. Each TPDDTEC Methodology version, the date it was published, and the number of projects that currently use each methodology version in the project portfolio. Source: GS Secretariat & Berkeley Air.

Methodology/Versions	Date published	Number of SWS VPAs
GS TPDDTEC v 1.	November 2011	384
GS TPDDTEC v 2.	April 2015	2
GS TPDDTEC v 3.	July 2015	0
GS TPDDTEC v 3.1	August 2017	104
AMS III.AV	N/A	12

In order to achieve certification, project developers work through a process to demonstrate that they have followed the methodology and met the standard. This process is managed by SustainCERT (SC), an independent impact certification entity that is the exclusive provider of this service to the Gold Standard. Prior to January 1, 2019, the functions currently provided by SC were carried out by a certification unit within GS. In practice, GS and SC work in close partnership, sharing a common vision. Several members of the SC leadership and staff are former GS employees. There are two structural mechanisms designed to provide GS oversight of the SC certification process: a set of key performance indicators are reviewed by the GS Secretariat for a random sample of certified projects prior to issuance; and both the GS TAC and representatives of the NGOs that founded GS are notified by SC and provided a review period for each project, initially after it passes design review and validation, and regularly thereafter when it passes verification and performance review at the end of each monitoring period.

The certification process varies depending on the nature and scale of the activity seeking certification, but all projects displaying impact certifications on the GS registry have progressed through four essential process stages (see Figure 3). Initially, the project undergoes a preliminary design review where eligibility requirements are checked, after which the project makes its debut on the GS registry with a “listed” status. These eligibility requirements include safeguarding principles and active stakeholder engagement, as well as initial estimates of climate and development impacts. Next, the project progresses into a full design review, which involves third-party validation, including a site visit, as well as a desk review of the project’s monitoring plans, and resolution of any issues surfaced by the third-party validation. Upon passing this stage, the project is registered with GS as a having a certified design.

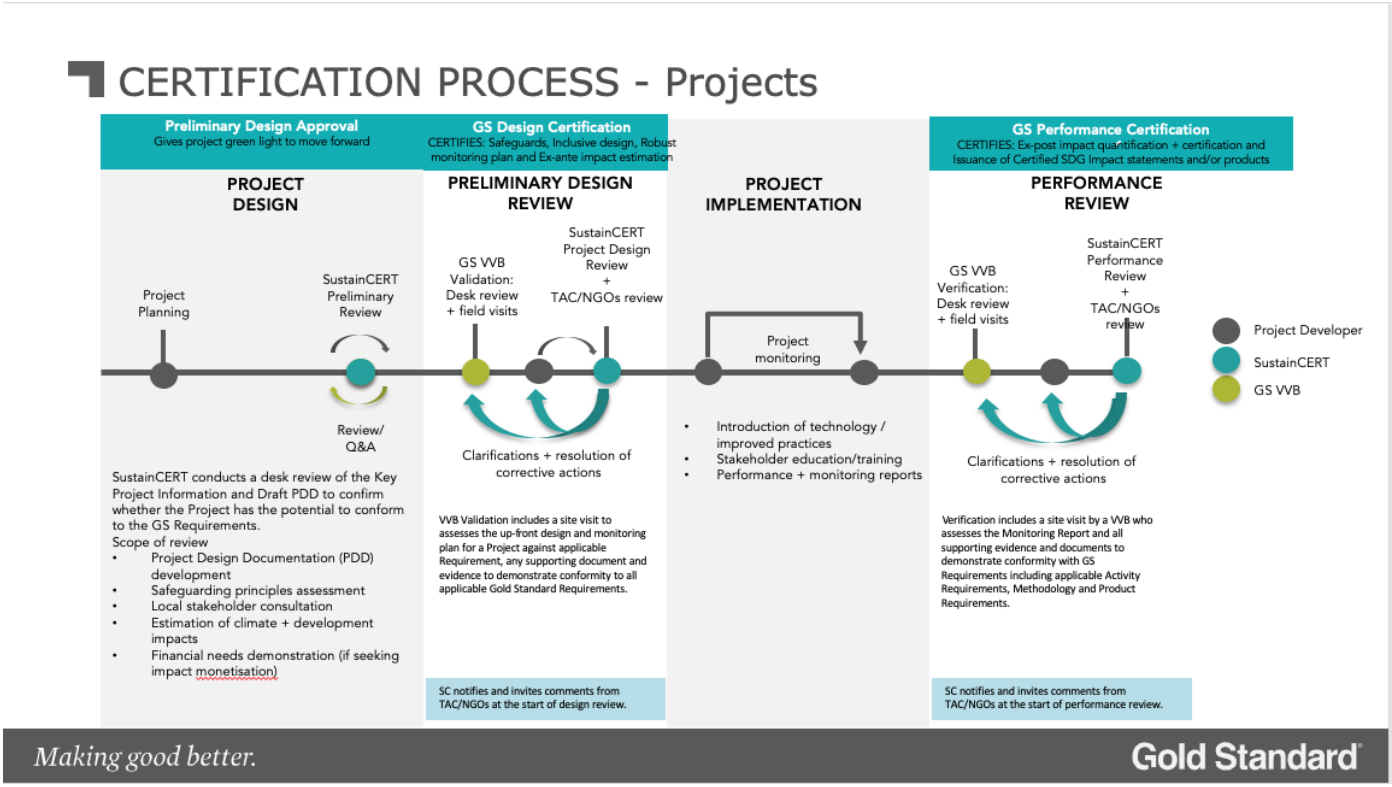


Figure 3. The Gold Standard certification process. Source: Gold Standard Foundation.

In some cases, implementation of the project begins at this stage, but it is also permissible for the project to have been operating for up to two years prior to this certification. Regardless, the project begins its approved impact monitoring in this third phase, collecting data to generate values for the parameters required by the relevant methodology to calculate the emissions reductions (ER) and assess development impacts in each monitoring period. The project progresses through the final certification phase when this data and the resulting parameter values are reviewed, vetted, and ultimately approved, resulting in a performance certification and impact quantification in the form of issued credits or impact statements available on the GS registry.

Projects following the GS standard exist at a variety of scales, and SC adjusts the rigor of the certification requirements for each scale to seek to provide appropriate mitigation of the risk that project impacts maybe be overestimated. In the case of SWS projects, the scale determinations are made on the basis of the expected ER volume: projects with ER up to 10,000 tons of CO2 equivalent per year (CO2e) are eligible and appropriate for the micro-scale scheme, those with ER of 60,000 tons CO2e or less are considered small-scale, and those with ER of more than 60,000 tons CO2e are large-scale. The micro-scale option was introduced in 2011 with the express purpose of increasing access for and inclusivity of the smallest climate and development projects, which could not otherwise afford to participate in the carbon markets due to the transaction costs and in fact might not be viable at all without this access.

In the micro-scale scheme, project developers may opt to have SC (and formerly GS itself) play the role of the accredited third-party Validation and Verification Body (VVB) that is required to perform the third-party reviews under the small- and large-scale project schemes. Once a micro-scale project developer has chosen to use SC for this part of the

certification process, SC then has the choice of placing an internal staff member or external contracted expert in the role of reviewer or contracting with an accredited VVB, as they see appropriate. In the case were SC acts as the VVB, an objective observer is contracted to make a site visit at least once within two years of the start of the crediting period. SC selects the objective observer based on the relevance of their expertise and knowledge of local conditions.[3] The project developer may be asked to submit credentials for three potential objective observers, but SC makes the final selection.

C. Hypothesis generated by Gold Standard Secretariat portfolio review

The hypothesis underlying the grievance put forth by the GS-TAC was that something had gone wrong in the Safe Water Supply project methodology or certification process, leading to a material risk of potential overestimation of emission reductions and possible over-issuance of ER credits. After Berkeley Air was appointed to lead the grievance investigation and supplied by the GS Secretariat with the SWS database and summary document containing information publicly available on the Gold Standard Registry, we formulated our own investigation approach described in Section III. Fundamental to our starting point was to invoke the statistical concept of the null hypothesis, e.g. that there was, in fact, no over-estimation of emission reductions in the GS SWS project portfolio. The following section describes our investigation approach and methods we used to test the hypothesis and attempt to reject the null hypothesis.

[3] Additional information on the selection of Objective Observers is available in the complete record of stakeholder comments included as an annex to this report.

III. Investigation Approach and Methods

A. Overview

Starting with the information available on the Gold Standard Registry compiled by the GS Secretariat, as well as documentation available for the TPDDTEC Methodology (versions 1, 2, and 3.1), Berkeley Air began investigations into the quantitative application of the methodology and the parameters used to calculate resulting emissions reductions.

An initial check of the emissions reduction calculation did not reveal any errors in equations, given the assumptions in the methodological approach. Next, a Monte-Carlo simulation was used to model which parameters had the greatest impact on the outcome of the *per capita* emissions reductions, as described by the TPDDTEC methodology. Those parameters were manipulated in each direction, and some were understood to scale with magnitude, such that any change in the value compounded in the final results. The compounding nature of those changes in key parameters became a central theme in the investigation's understanding of the application of the methodology.

Following the identification of those key parameters, which are detailed later in this document, Berkeley Air investigated their values throughout the entire project portfolio. Berkeley Air consulted experts in the cookstove and safe water fields and drew on in-house expertise of individual methodology components to validate parameter values and their generation. A library of resources was assembled including articles from peer reviewed journals and protocols relevant to key activities in the TPDDTEC monitoring requirements, such as the WBT, to support the team's assessment of these parameters.

Berkeley Air then sought individual PoA and VPA project documentation to better characterize full sets of project parameters and how they were derived (see Section D below for details). From the initial selection of individual project documentation, Berkeley Air selected three case studies, for which several parameters were closely examined in a single case. The selection of those three case studies and the results of their investigation are presented further in this document.

In parallel to the documentational review and calculations analyses, Berkeley Air also conducted interviews with key stakeholders. Interviewees were identified by and contact was sometimes facilitated by Berkeley Air, the TAC-SWSGWG, and/or the Gold Standard Secretariat. Context and expert opinions provided by interviews were integrated into the investigation as it continued. Further detail on those interviews as well as a full table detailing all interviews conducted are available later in this document and in Appendix 2.

Through the duration of the investigation, Berkeley Air attended weekly check-in meetings with the TAC-SWSGWG and Gold Standard Secretariat to share key learnings and concerns and gather essential feedback and resources.

B. Short-listing of key parameters

Key parameters for investigation were determined through a two-stage process. First, a Monte Carlo simulation was undertaken, in which input parameters were tested using distributions reflecting the range of estimates in the project portfolio. The results provided an indication of which parameters were having the greatest impact on the per capita

emissions reductions. This exercise suggested that the fuel consumption per liter water boiled and the amount of safe water used per person were having the largest impact. The water use per person, however, has already been capped at 7 l/person/day[4], and therefore was not a focus of further investigation. Next, the parameters closely related to scale and technology performance were considered as they had a clear impact on the total emission reductions estimates. These parameters included the number of users per technology, project technology days, and usage rates.

The fraction of non-renewable biomass harvesting (fNRB), while also an important input in the emissions reduction calculation, was not considered for further investigation as it extends to other methodologies where biomass use is part of the emissions reduction calculation, and a review of those methodologies is out of scope of this investigation (Bailis et al. 2017). fNRB estimates have also been a subject of debate, as different estimation approaches can lead to large differences in the values used, but again this was deemed out of scope.

C. Literature review

Berkeley Air reviewed several classes of documents throughout the course of the investigation, including a) those related to Gold Standard TPDDTEC methodology and the project certification process, b) those sources available in peer-review and grey literature related to baseline and intervention safe water systems parameters, and c) those related to individual projects in the SWS portfolio (discussed further in Section D below).

Those documents related to the methodology itself and the project crediting process served as reference material and guidance through the investigation from one stage to the next. Individual experts were able to give input on specific components of the methodology, particularly on the guidance to arrive at the parameters used to calculate final emissions reductions. Additionally, the Gold Standard Secretariat was able to offer context on the project pipeline as a whole, including summary documents assessing the character of the relevant projects as a whole and patterns that have emerged along the pipeline in recent years.

To supplement in-house expertise, the Berkeley Air team built a library of literature references. These sources were originally selected for their ability to provide credible context for safe water systems implementation and operation, as well as baseline and intervention scenarios in relevant geographical locations. As the investigation continued, the library expanded to include working papers on funding mechanisms, briefs on the state of safe water interventions in the sector, and manufacturer's specifications for clean water technologies.

Although a significant volume of literature was collected and reviewed, Berkeley Air did not have a mandate to conduct systematic reviews for any of the key parameters. The BA team noted that the volume of published literature varies significantly across topics: some aspects of SWS initiatives are well studied, while others have not yet been investigated as

[4] In earlier versions of the TPDDTEC methodology, this cap was set at 7.5 l/person/day.

thoroughly. In multiple cases, we found that papers reported on more than one research question or study sample, offering some results that were more relevant to the circumstances of SWS projects than others. In the interest of efficiency, the relevant statistics and conclusions from such studies have sometimes been cited in this report without an accompanying explanation that the same reference also includes other statistics that are less or not at all relevant to subject of this report.

D. Project documentation review and case study selection

Following the review of the relevant project portfolio in its entirety, the Berkeley Air team requested individual project documentation. In the first round, documents were chosen to give a broad representation of the full portfolio, and then in subsequent rounds of request, more specifically chosen to offer insights into a variety of monitored project and baseline parameters (see progression in Figure 3). The documents were requested through the TAC-SWSGWG tasked with supervising the investigation and provided by SustainCERT.

The initial request for individual project documents focused primarily on drawing a sample from those PoAs that housed the majority of relevant Safe Water Supply projects. The request was made at the PoA level, and personnel at SustainCERT selected and provided sample projects from each PoA as illustrative examples of the overall PoA. Most of these sample projects utilized borehole technology, but a handful of projects focused on other safe water technologies. Upon further investigation, Berkeley Air narrowed its review to include only borehole initiatives, as these represent 90% of the SWS portfolio, and only those projects with one or more years of complete performance data.

From there, the investigation team also sought representation from clusters of projects under these POAs, characterized by a shared geography, start date, and outlier values for one of three (initial) parameters of interest: fuel per liter of water boiled, usage percentage, and number of users. These projects were selected randomly by Berkeley Air and requested at the project level. Finally, towards the end of the investigation, Berkeley Air requested documentation for three specific additional projects that had been identified by their project developers as having reported less than 100% borehole functionality in a particular monitoring period. In total, documentation for 21 projects was reviewed.

All project documentation was stored securely in a Dropbox folder shared with the GS TAC-SWSGWG. The format of the provided documentation presented certain challenges to the reviewers: the original file date stamps were no longer available, the file folder structure and file naming conventions were not standardized, and many documents appeared more than once throughout the project archives, sometimes in new iterations or versions and sometimes as reference copies. A reasonable effort was made to understand the full scope of information provided and extract data for further analyses as accurately and consistently as possible.

In order to provide an in-depth analysis of the application of the emissions reduction parameters and calculations within the timeframe of the investigation, the Berkeley Air team decided to select three illustrative case studies to provide a focal point. The case study projects selected were broadly representative of three characteristics of the SWS portfolio: they were borehole interventions, were located in three different sub-Saharan African countries, and were developed by two entities with the strongest presence in the SWS portfolio. However, they were not representative of the average values for key parameters

across the portfolio, but expressly selected because they used outlier values that the team felt warranted further investigation.

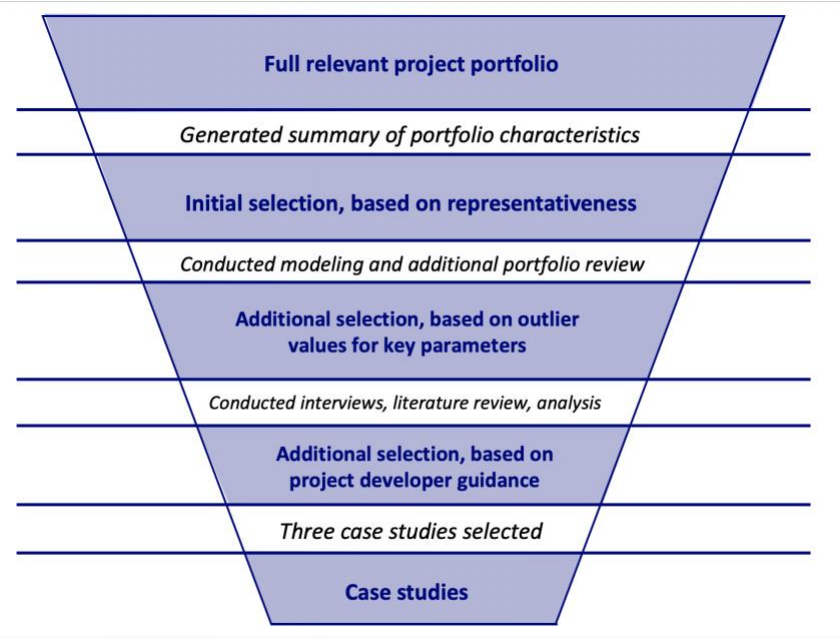


Figure 4. Funnel diagram describing the selection of individual project documents, with simple descriptions of actions taken between selections. Source: Berkeley Air.

Over the course of comparing the methodology, the reported developer and certifier practices, and the literature on and expert opinions concerning key project parameters, the BA team looked at the several groupings of project data. There were several reasons for this evolving data analysis approach. For example, some parameters were much easier to pull out of the project documentation than others, such that a larger data set could be assembled within the investigation timeframe. Further, some parameters were measured in every VPA and monitoring period, such as project technology days, whereas others such as fuel per liter of water boiled, were only measured at baseline for groups of projects. It was also clear that additional data was needed at times after interviews, which highlighted gaps or pointed to the need to check for revised parameter estimates. In the presentation of the findings, the sample underlying each conclusion has been described as fully as possible.

E. Stakeholder interviews

Per the investigation TOR, relevant stakeholders were interviewed during the investigation process to provide additional context and considerations. Interviewees included some of the GS Secretariat staff, leadership and staff at SustainCERT, senior implementers at various project developers, independent auditors and internal and external SC reviewers, and experts in the safe water supply and cookstove fields. A complete list of interviews is available in Appendix 2. In total, Berkeley Air spoke with 22 stakeholders over the course of 15 interviews.

Interviewees were identified by the investigation team as well as through suggestions from the GS TAC-SWSGWG, and in some of those cases, the TAC-SWSGWG facilitated contacts between Berkeley Air and the interviewees. Each interview was conducted by members of

the Berkeley team and recorded with written minutes taken. Notes and recordings were shared with the TAC-SWSGWG.

The focus of each interview was tailored to the interviewee, but questions were broadly aimed to supplement written information and accelerate the progress of the investigation. Experts lent additional perspective on the baseline and intervention scenarios in relevant contexts, shared experience with methodological components, and offered expertise on the performance of safe water technologies. Interviewees affiliated with the Gold Standard and SustainCERT shed additional light on the project pipeline and certification process, supplemented by those auditors and reviewers also interviewed. Interviewees from various project developers offered additional context for their projects and were asked about specific project implementation practices and monitoring protocols. Information gathered through each interview allowed the investigation to consider perspectives along the certification pathway, and to validate analyses and modeling the investigation team had conducted with information from the project portfolio and individual project documents.

IV. Part A Findings

A. Analysis of key parameters contributing to over-estimation

Through the analysis described in Section III, four parameters were identified to be most likely to affect the estimates of emissions reductions for SWS borehole projects. In this section, we compare reported values and monitoring practices for these parameters to the GS standards, methodologies, and guidance, as well as the relevant literature and learnings from key informant interviews, in order to characterize the nature and extent of their potential impact.

1. Number of users

a) Methodological guidance

Section 3.1A of the TPDDTEC methodology provides guidance on the record-keeping required for technology users. "The project proponent must maintain an accurate and complete sales record. The sales record is substituted by a "dissemination record" or "installation record" in projects with non-commercial distribution or dissemination of a practice." The guidance further outlines required data, including the "name and telephone number of all end-users except in cases where this is justified as not feasible. In such cases, the number of names/telephone number/addresses collected must be as many as commensurate with representative sampling." Section 3.1B then describes the "project database," which is "derived from the total sales record (or dissemination record in case of non-commercial distribution) with project technologies differentiated by different project scenarios." No additional guidance is outlined on identification or validation of the records.

b) Reported practice and values

Project developers and pump manufacturers estimate the total number of users of a borehole as not just those who collect the water but as the full number of household members who use the water. Project developers commonly describe collecting household user lists through local channels, including district offices, village leaders, and water committees. One project developer described a process of leaving a document template at the borehole with a local partner who manages a process for collecting household names and the number of household members. Further, in some cases, project developers have capped the number of users based on the total water volume generated from the borehole, calculated from the treatment capacity metric (explained in Section IV.2.b below) and dividing by the water consumed per person per day from the Water Consumption Field Test (WCFT).[5] The three case studies we examined closely reported an average number of users per borehole of 701, 1109, and 1352. The overall portfolio of borehole projects had an average of 544 users with a range of 225 -1704 users.

c) Comparison with evidence base

Determining the number of users in any community level project can be difficult given the complexity of identifying each unique user of the technology. A couple of handpump studies were identified that referenced number of users, including a study in Malawi estimating up to 155 households per source (Holm et al. 2017) and a study in Ghana averaging 113 users

[5] Please see the complete record of stakeholder comments included as an annex to this report for additional information on the certification body's approach to the number of users parameter.

per source. Interestingly, the study in Ghana referenced a maximum borehole user standard of 300 users per source, with 7% of sources in the study exceeding that threshold (Fisher et al. 2015). In addition, as detailed in the treatment capacity in Section IV.B.1 below, the technical specifications for handpumps used in the SWS portfolio suggest they are designed to serve a population of 300 users.

An additional consideration is the change in users year over year. Of the three case studies we examined closely, no change in number of households was reported between monitoring periods. One project developer did mention that they are examining possible changes to their protocol for counting users and working with Gold Standard to formalize those, but they did not believe household numbers were changing substantially, apart from a couple exceptional cases.

d) Conclusion

This limited set of literature values suggest that the number of users reported by project developer values are high, as does the comparison with manufacturer recommended specifications. There is likely overestimation here, especially considering the importance of having a conservative value for this parameter, which has a strong multiplier effect on the overall ER estimates. Additionally, as noted in the case in Ghana, any local standards that exist for this parameter, or any other parameter, should be considered as a limit, as no projects should undermine local guidance or regulation.

2. Usage percentage

a) Methodological guidance

Section 3.1Cb of the TPDDTEC methodology describes the usage survey as a "single usage parameter that is weighted based on drop-off rates that are representative of the age distribution for project technologies in the total sales record. A usage parameter must be established to account for drop-off rates as project technologies age and are replaced." In V2.0 (via a rule update dated 29th April 2015) of the TPDDTEC, Annex 9: *Guidelines for carrying out usage surveys for projects implementing household water filtration (HWT) technologies*, was added. Annex 9 is "applicable for HWT filtration technologies (e.g. sand filters, clay filters, ceramic filters, hollow fibre filters, bio-sand filters, etc.) and not to safe water supply projects, such as chlorine treatment, solar disinfection, bore wells, piped water supply, etc." Annex 9 describes six approaches to assess usage, including reported and observed usage of technologies, with additional "precautions" for enumerators, including not conducting surveys after capacity building activities or wearing the logos of their employers. While this additional guidance is given for HWT projects, there is no additional guidance on survey questions or methodology for other technologies. Interestingly, a footnote mentions the guidelines will be expanded in the future to the other technologies.[6]

b) Reported practices and values

Usage survey questions vary a bit across developers and project vintages. In some instances, households are simply asked to identify which borehole they are using. More extensive

[6] Please see the complete record of stakeholder comments included as an annex to this report for additional information on the status of additional usage survey guidelines.

surveys ask households which source they use to collect water and the purposes for the water collected.[7] Project developers can only receive credit for the volumes of water used for three purposes: drinking, basic personal hygiene, and food preparation; therefore, these uses need to be reported by households in order for the source to be material to the usage estimate. Lastly, a question is sometimes asked to determine if the borehole water is sufficient for most or all of the water needed for these three purposes. If a household answers that they collect borehole water, confirm that it is used for the three purposes credited, and confirm that the quantity collected is sufficient for those three purposes, the household is considered a 100% user. Across the project portfolio, borehole project developers largely report this usage parameter as 100%. A few exceptions report high 90th percentile usage figures, but this is rare among the borehole projects.

c) Comparison with evidence base

A consistent theme that emerged from both the literature and expert interviews was the prevalent household practice of using multiple water sources, across many geographies. In some locations, households may have access to more than one community groundwater source, installed by different entities, in addition to surface water sources such as rivers, streams, ponds, and rainwater harvesting. Further, the literature and experts point to households making rational and convenient choices to use closer, more easily accessible water when seasonally available, usually for a portion of hygiene or cooking needs, in order to reduce the frequency of more challenging trips to community sources in bad weather (Kelly et al. 2018).

Specifically, usage of different water sources due to seasonality has been studied and linked with higher borehole usage in the dry season when surface water and rainwater sources dry up. Sources that may be more commonly used in the wet season are not just those that result from the households directly collecting rainwater, but also sources (rivers, streams, ponds) that may only exist during the wet season or that have a much more limited supply during the dry season. A study in Kenya reported a 34% reduction in groundwater use during the rainy season. In an area with high rainwater usage, 86% of households named boreholes as their primary water source during the dry season, with over half reporting the borehole as their only source. However, during the rainy season, the percentage dropped to just 6% stating their only source of water was a handpump (Thomson et al. 2019). A study in Kenya and Ethiopia, while using electric pumps (not handpumps), similarly found a 23% increase in borehole runtime following weeks with no rainfall (Thomas et al. 2019). Lastly, a qualitative study of rural water systems across Ghana, Kenya, and Zambia, confirmed seasonality as a significant factor, with improved groundwater sources being used more often in the dry season (Kelly et al. 2018).

In most of the project monitoring data we reviewed, households reported that the project borehole was their only water source, inconsistent with literature values and expert interviews. In some cases, surveys did not give households the option to report a source other than a borehole. In cases where multiple sources could be identified, no follow-up questions were included to ensure that water collected from other sources was not used for drinking, personal hygiene, and/or food preparation. Further, none of the project developer

[7] Please see the complete record of stakeholder comments included as an annex to this report for additional information on the certification body's approach to usage surveys.

surveys we reviewed addressed a potential seasonal component to household water usage patterns.

Finally, it is unclear from the GS methodological guidance what specifically falls into the category of basic personal hygiene.[8] One water expert described how widely hygiene behaviors vary across different geographical and cultural contexts. For example, handwashing may be done in the home, while bathing may be performed at the water source.

d) Conclusion

The usage percentage of 100% as applied in most projects is likely too high, based on evidence in the literature of widespread household use of multiple water sources, especially considering seasonal effects and access to other sources. It is difficult to capture and apportion reported water use to the specific water-using activities that fall into the three creditable categories, adding uncertainty to this parameter.

3. Project Technology Days (Number of Person Days)

a) Methodological guidance

Relatively little guidance is provided in the TPDDTEC with regard to the monitoring of Project Technology Days, even though it is a key parameter in the emissions reduction calculation. The methodology simply states in equation (1) of Section 2.7 that the Project Technology Days are the "cumulative number of project technology-days included in the project database for project scenario p against baseline scenario b in year y." While the methodology doesn't state that days should be removed from this parameter when the technology is non-functioning, a second parameter "Number of Person Days" is defined in equation (11) as the "number of person days consuming water supplied by project scenario p through year y." In practice, the Number of Person Days was calculated by project developers as the Project Technology Days, with an adjustment for days when the borehole was non-functioning, multiplied by the number of users.[9]

b) Reported practices and values

A key factor governing the functionality of boreholes is maintenance, as by their nature, handpumps experience wear with use. Maintenance models are not extensively described in project documentation, but some common elements were observed. Generally, project developers describe working with local organizations to train community members on small repairs, with the local organizations responsible for addressing larger repairs when these are reported by community members. Additionally, some project developers describe a regular routine of visiting boreholes to perform preventative maintenance from annually to several times per year. If a larger repair were to occur, the local organization typically documents the problem and the repair, including the number of days the handpump was not functional.

[8] Additional information on the three claimable uses is available in the complete record of stakeholder comments included as an annex to this report.
[9] Please see the complete record of stakeholder comments included as an annex to this report for additional information on the certification body's approach to project technology days.

In our case review for this parameter, across 13 projects with complete data for 195 boreholes over 32 monitoring periods, 186 of those never saw any adjustments in their crediting days. Seven boreholes saw adjustments because they were built or initially rehabilitated some time into the relevant monitoring period, which means that 193 of 195 boreholes (99%) in this sample were never adjusted after becoming operational. One borehole saw an adjustment due to an unsatisfactory water quality test, and during our full review of calendar versus credited days, we found documented evidence of only one case of a borehole seeing a crediting days adjustment due to maintenance or repairs to the borehole.

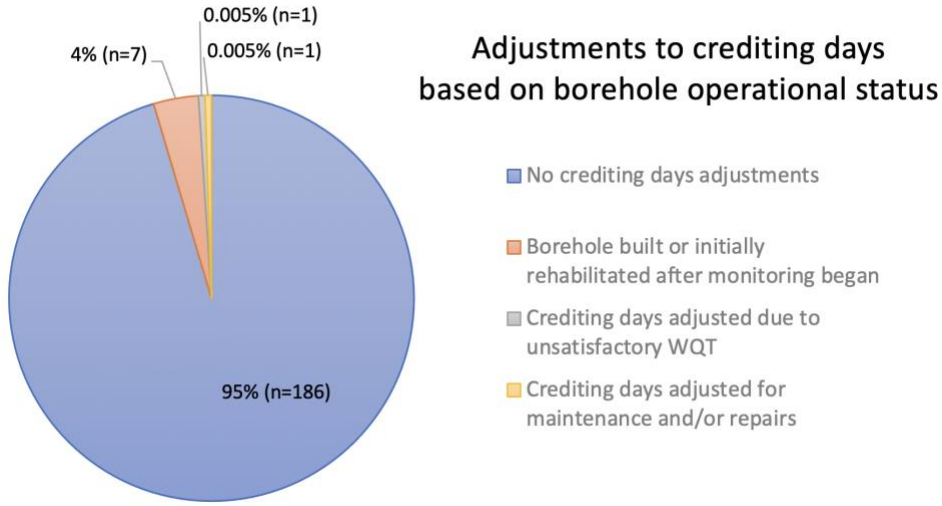


Figure 5. All adjustments made to crediting days for each borehole over all available monitoring periods based on the operational status of the borehole technology. Source: Berkeley Air.

During our interviews with project developers, however, they described instances of boreholes being taken off-line for repairs and adjustments being made to the Project Technology Days value. They subsequently provided three examples, where overall reductions of 2%-5% were made from the total crediting days of all the boreholes under the crediting period.

Project developers explained the effectiveness of their maintenance programs by noting that in the case of carbon projects, it is in the interest of the implementers to ensure that the sources are functional.

c) Comparison with evidence base

Handpump functionality and maintenance is a widely studied topic. A recent comprehensive review of handpump functionality found approximately one in four handpumps in sub-Saharan Africa to be non-functional at any point in time (Foster et al. 2019). Several other studies report functionality rates from 55% to 80% (Fisher et al. 2015; Foster et al. 2018; Owor et al. 2017; Truslove et al. 2019).

Research suggests that the maintenance model plays an important role in handpump functionality rates. In Malawi, a sample of approximately 15,000 boreholes found about a 14% increase in functionality, from 57.6% to 71.4%, when a regular service provider was present (Truslove et al. 2019). In Rwanda, a comparison study of three different

maintenance models showed a decrease in handpump downtime from 152 days under a "nominal maintenance" model to 57 days under a "circuit rider" model and 21 days when real-time sensor monitoring was used to dispatch technicians (Nagel et al. 2015). An additional sensor-based maintenance model called the "Smart Handpump Project" pairs the real-time data from sensors with a performance-based maintenance service model. Under the Smart Handpump project, repair times were reduced from 27 days to three days (SDSN TReNDS 2018).

Emerging evidence in performance-based funding for rural water supply programs suggests higher functionality rates may be possible. In a study of programs across four African countries using performance-based models of both handpump and piped water supplies, an uptime over 94% was recorded with a rapid breakdown response of less than three days (McNicholl et al. 2019). Similarly, a recent study published by one of the implementers in the study just discussed, reported a functionality rate above 97% in rural Uganda (Harvey and Mukanga 2020). While these studies provide encouraging evidence, they additionally describe the use of comprehensive maintenance programs and institutional structures not documented in the GS SWS projects reviewed by the BA team.

d) Conclusion

With respect to Project Technology Days, GS SWS projects report much higher functionality than is seen across the literature. There is evidence that some maintenance models work better than others, and that incentives can play a role in improving performance, but even the most advanced models with sensing and performance-based approaches are not reaching 100% functionality. The emerging evidence of possible higher functionality rates achieved through performance-based models should be treated cautiously until they are more widely replicated across a range of contexts, with additional consultation with researchers recommended. Given the impact this parameter has on the overall ER estimates, it is likely responsible for substantive overestimation.

4. Fuel per liter of water boiled

a) Methodological guidance

Guidance provided by the methodology is short and has not changed in any substantive way between TPDDTEC versions. From Annex 3, pg.50: "The baseline water boiling test (BWBT) is conducted to calculate the quantity of fuel required to purify by boiling one litre of water for 10 minutes using technologies and fuels representative of the baseline scenario ($W_{b,y}$). The BWBT should be conducted using the 90/30 rule for selection of samples, accounting for variability in the types of prevalent baseline technologies."

b) Reported practices and values

All projects reviewed for this report provided instructions on selection of households, pilot testing, and the procedure to follow to conduct the test. The procedures typically involved household participants or technicians using the stove and pots available at the home to bring one liter of water to boil and sustain that boil for 10 minutes, before and after which the fuel was weighed. The difference in fuel is then the output metric for the test (kg/l). This metric is commonly termed "specific consumption." This report focused on woodfuel, as a charcoal baseline was assumed in only a small number of projects. Projects have reported a

large range of specific consumption, summarized in Table 3 below. The range in our sample of the SWS portfolio was 0.31-1.36 kg wood/l, and the portfolio average was 0.83 kg wood/l.

c) Comparison with evidence base

This testing procedure (boiling one liter of water in homes) is fairly idiosyncratic to the methodology, so there are not direct comparisons to other reported field data; however, laboratory testing of traditional open wood stoves has indicated lower specific consumption estimates [e.g. (Jetter et al. 2012): 0.280 kg/l; and (Teune et al. 2020): 0.214 kg/l].[10] These estimates should not be assumed to reflect the conditions under which homes in the project areas boil their water, where similar stoves are often less efficient (Berkeley Air 2012), yet these results still suggest that water can be boiled much more efficiently by traditional wood stoves than reported by the majority of tests in the SWS portfolio.

There are several operational parameters that can affect how efficiently a stove boils water, such as use of a lid, the characteristics of the cooking vessel, the amount of water boiled, the fuel feeding rate, and others. The methodology does not provide prescriptive guidance on these parameters, with the exception of the implication that the test should be conducted with one liter of water, rather than boiling a larger pot of water and normalizing the results to one liter. The WBT protocol 4.2.3[11] (and earlier versions), which has been the sector's standard for laboratory testing of cookstoves and has been applied in many studies, generally does provide more prescriptive guidance, though it is primarily meant to be applied in a controlled laboratory setting (Arora and Jain 2016; Berrueta et al. 2008; Chen et al. 2016; Collivignarelli et al. 2010; Grimsby et al. 2016; Jetter et al. 2012; Jetter and Kariher 2009; Johnson et al. 2008; Just et al. 2013; MacCarty et al. 2010; Quist et al. 2020; Rapp et al. 2016; Shen et al. 2018; Smith et al. 2000; Teune et al. 2020). The WBT consists of three phases as described in the protocol overview here (Section II, pg. 5) with a corresponding diagram in Figure 6.

> "1) For the cold-start high-power phase, the tester begins with the stove at room temperature and uses fuel from a pre-weighed bundle of fuel to boil a measured quantity of water in a standard pot. The tester then replaces the boiled water with a fresh pot of ambient-temperature water to perform the second phase.
>
> 2) The hot-start high-power phase is conducted after the first phase while stove is still hot. Again, the tester uses fuel from a pre-weighed bundle of fuel to boil a measured quantity of water in a standard pot. Repeating the test with a hot stove helps to identify differences in performance between a stove when it is cold and when it is hot. This is particularly important for stoves with high thermal mass, since these stoves may be kept warm in practice.

[10] Laboratory water boiling test results are typically reported as dry wood, and thus an adjustment was made to assume 20% moisture content in alignment with the moisture content for the default energy conversion in TPDDTEC. The WBT version for these tests simmered water for 30 minutes, so the quantity of wood from this phase of the test was divided by three to match the 10 minute boiling phase of the BWBT. Also note that the reported thermal efficiency of TSFs and traditional wood stoves tested in the laboratory have typically ranged from 9%-23% with most around 15-20% (Aprovecho Research Center 2007; Barbieri et al. 2018; Berrueta et al. 2008; Collivignarelli et al. 2010; Jetter et al. 2012; Jetter and Kariher 2009; MacCarty et al. 2008, 2010; Smith et al. 2000; Still et al. 2015; USAID and the Academy for Educational Development 2007), which is generally higher than the 10% assumed in the TPDDTEC methodology.

[11] http://cleancookstoves.org/binary-data/DOCUMENT/file/000/000/399-1.pdf

3) The simmer phase provides the amount of fuel required to simmer a measured amount of water at just below boiling point for 45 minutes. This step simulates the long cooking of legumes or pulses common throughout much of the world."

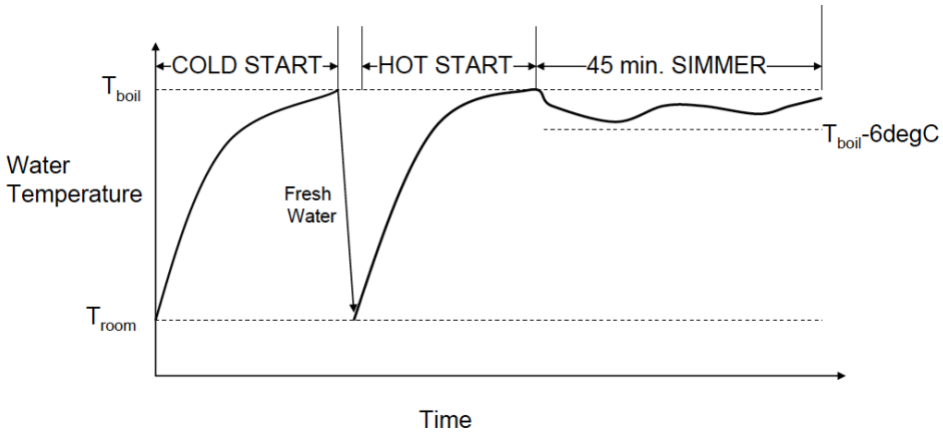


Figure 6. Illustrative diagram of water temperature over the course of a WBT. Source: Nordica McCarty, Oregon State University.

One of the key variances between the BWBT and the standard WBT is the amount of water used for the test. While five liters of water is the recommended amount of water used in the WBT, one liter is implied to be used in the BWBT. It would be reasonable to assume that households boiling water for their daily needs would do so in larger batches than one liter, given that tens of liters of clean water are needed per household per day. Figure 7 shows the impact of water quantity on specific consumption for a typical three stone fire (TSF), illustrating how specific consumption drops as water volume increases, in this case from approximately 0.48 kg wood per liter when 1 l of water is used to 0.26 kg/l for 5 l. The reason that specific consumption drops as larger water volumes are used is that a typical TSF being operated at a constant firepower (see text below for more discussion on firepower)[12] will maintain a boil regardless of whether there is 1 l or 5 l of water in the pot.

[12] Assuming that the stove is operated at a 7 kilowatt (kW) firepower and has 10% thermal efficiency. Note that typical simmering phase firepowers for WBTs with 5 l of water range from 3.2-5.9kW (Aprovecho Research Center 2007; Berrueta et al. 2008; Collivignarelli et al. 2010; Grimsby et al. 2016; Jetter et al. 2012; MacCarty et al. 2010), thus operating a TSF at 7kW implies that a strong boil can be maintained.

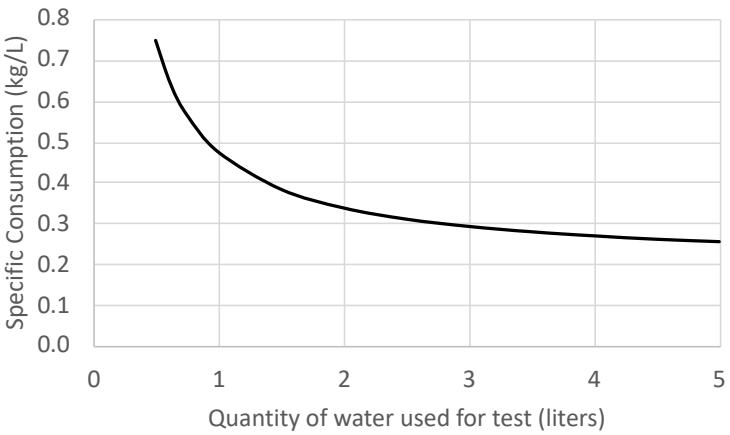


Figure 7. Specific consumption as a function of water quantity used for testing. The relationship assumes a 7kW stove with 10% thermal efficiency, with water being raised from 25-100 °C, and then boiled for an additional 10 minutes. Source: Berkeley Air.

The other test parameter that is most likely to influence the specific consumption is firepower. For this context, firepower is defined as the rate of heat energy released by combustion and is typically reported in kilowatts (kW = kJ/s). Firepower is most closely related to the fuel use rate and provides a diagnostic or sense-check on how realistically representative the tests are of normal operation. Stoves perform differently (e.g. have different fuel efficiencies) at different firepowers (Bilsback et al. 2018; James Robinson et al. 2010; Johnson et al. 2010; Prasad et al. 1985), which is in part why the WBT has high (boiling) and low (simmering) power phases. The new ISO standard for laboratory testing also is based on a test sequence that purposely runs stoves through high, medium, and low firepower phases, due to firepower's impact on stove performance (Bilsback et al. 2018; ISO 2018).[13]

For stick fed stoves like TSFs, firepower is primarily a function of the quantity of wood used and frequency at which it is fed into the stove. Field tests of TSFs (Table 2) in Africa show that mean operating firepower is around 7kW, with the maximum reported for a given event being 13 kW. Importantly, these tests include the range of cooking events from small tasks such as boiling water for tea, to larger tasks such as cooking a batch of matoke.[14] In our sample of BWBT in the SWS portfolio, the mean firepowers ranged from 7.0 to 22.7 kW. The BA team estimated these firepowers by multiplying the wood mass used by the default energy conversion factor for wood in the TPDDTEC methodology (15,400 kJ/kg)[15] and dividing by the time of the respective BWBT(s). Figure 8 shows a comparison of the available field testing firepowers to those from the BWBT sample, illustrating that most BWBT have

[13] The ISO cookstove standards are cited here and throughout this report as a point of comparison only and are not intended to imply that the GSF is or should be a party to these standards.
[14] Matoke is a Ugandan preparation of indigenous bananas, which requires more cooking energy than boiling one liter of water, representing variation in what households may require from their stoves and how they may operate them.
[15] The default energy conversion factor assumes 20% moisture content and was applied to the wet wood used, which was typically less than 20% moisture content for the BWBTs. This assumption was conservative, as lower moisture content wood has higher energy conversion factors, which would have resulted in slightly higher firepowers.

firepowers well above what would be considered normal operation, especially for boiling one liter of water.

Table 2. Field testing estimates of firepower for three stone fires in Africa. Source: Berkeley Air.

Country	Mean (kW)	Min (kW)	Max (kW)	N	Study	Notes
Malawi	~6	~1	~11	13	Wathore et al. 2017	Estimated from box plot
Rwanda	5.9	2.7	9.8	16	Champion et al. 2019	
Kenya	7.1	4.6	10.1	10	Berkeley Air et al. 2016	
Kenya	7.2	3.4	10.1	32	Johnson et al. In review	
Uganda	7.5	2.8	13	23	Johnson et al. 2019	
Ghana	6.8			20	Coffey et al. 2017	Only mean and variance presented (CoV of 26%)
Mozambique	10			29	Robinson et al. 2011	Only mean and variance presented (CoV of 31%)

Table 3. Specific consumption and firepower estimates from sample of VPA BWBTs. Source: Berkeley Air.

Country	POA/VPAs	Specific consumption mean (kg/l)	Firepower mean (kW)	Notes
Zimbabwe	GS1247 GS6518 GS6519 GS6520 GS6521 GS6522 GS6523 GS7485 GS7486 GS7487 GS7488 GS7489 GS7490	1.36	19.0	
Ethiopia	GS1247 GS5322 GS5323 GS5324 GS6037 GS6038 GS6783 GS6784 GS6836 GS7287 GS7288 GS7289 GS7290 GS7291 GS7292	1.27	22.7	
Zambia	GS1247 GS7456 GS7457 GS7458 GS7459 GS7460 GS7461 GS7462 GS7463	1.19	16.1	
Eritrea	GS1247 GS5038 GS5039 GS5040 GS5041 GS5042 GS5043	0.31	7.0	Only 1 or 2 minutes from test start time to boil (time was likely longer, implying lower firepower)
Eritrea	GS1247 GS5951 GS5952 GS5953 GS5954 GS5955 GS6041 GS6042 GS6043 GS6044	0.577	7.6 (10.3 for only TSFs)	Approximately half of the stoves were “improved” and half TSFs

Eritrea	GS1247 GS4422 GS4423 GS4518 GS4519 GS4798 GS4797	0.883	11.6	Very low variability in specific consumption estimate (7.6%)
Eritrea	GS1247 GS5825 GS5826 GS5827 GS7330 GS7331 GS7332 GS7333 GS7334 GS7335 GS7336	0.889	10.0	
Rwanda	GS5047 GS7148 GS6189	0.946	15.8	
Uganda	G1247 GS3433 GS3563	0.679	10.8	Start/stop times not in correct format and some specific consumption estimates were very low (below 0.1)
Kenya	GS1366 GS7128	1.156	16.1	
Rwanda	GS5047 GS7395	0.946	16.2	

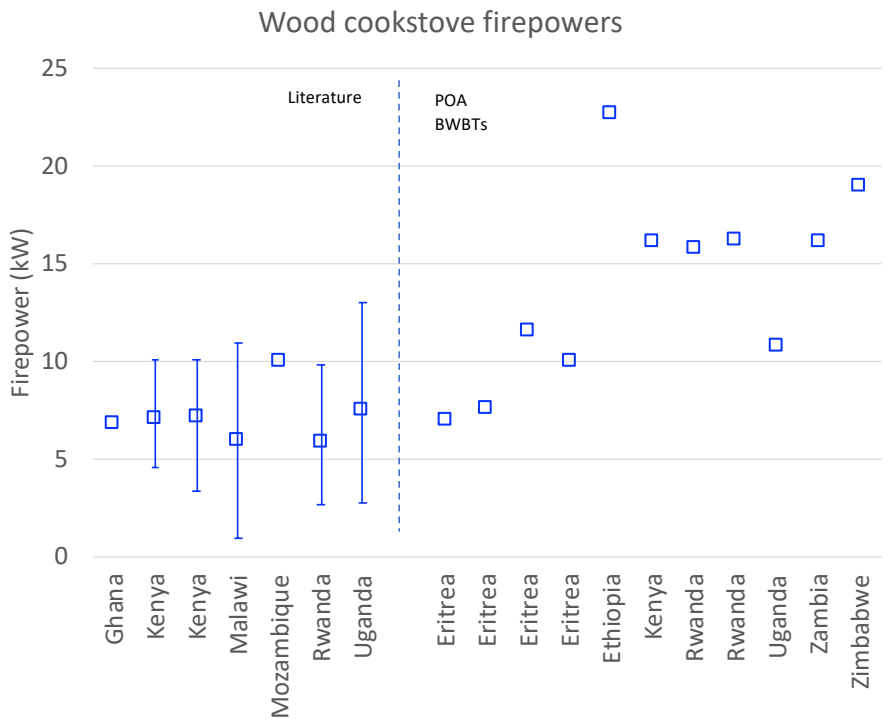


Figure 8. Comparison of firepowers from field tests (literature values) and those reported for BWBTs. Error bars represent the range of firepowers from a given study when available. Also note that the BWBTs represent 71 VPAS, as shown in Table 3. Source: Berkeley Air.

The implication from this comparison is that the stoves in these BWBT tests are being run with excessive amounts of fuel to boil a liter of water, which is resulting in higher specific consumption estimates than would be expected if more typical operation was occurring. Importantly, these higher BWBT firepowers are achievable within the guidance of the

methodology and specific protocols applied. For example, a commonly applied protocol specifies bringing the water to a “roiling” boil and maintaining a “roiling” boil throughout the test. There is no clear upper bound of the firepower to conduct this task, which allows for potentially unrealistic estimates of specific consumption. There is also no guidance provided by the methodology or the project-specific protocols on what constitutes a reasonable upper level of firepower (or fuel feeding), and no requirement to calculate or report firepower as a diagnostic sense-check.

Note that there are other parameters that can affect the efficiency and, in turn, the specific consumption of a BWBT. If wood has a high moisture content, for example, more energy released during combustion will be required to heat up and vaporize that water, resulting in a higher specific consumption estimate. High moisture content did not appear to be a common occurrence in the specific BWBT source data (less than 20%), and the project protocols we reviewed specify using “dry wood”. The size, shape, and type of wood (e.g. bark versus trunk wood, tree species, etc.), as well as fire-starting fuels and techniques and environmental conditions (altitude, wind, etc.) can also impact the efficiency of a stove, but these are likely nominal compared to the impacts that operating at higher firepowers and using small water volumes have on the specific consumption estimates from BWBTs.

d) Conclusion

Overall, the protocols used in these projects, which were in alignment with the methodology and approved by VVB auditors or SC (and previously GS) reviewers, were not prescriptive enough and failed to build in the sense-checks that would have provided more reasonable and conservative estimates for specific consumption. The low volume of water used and high firepowers of the BWBT estimates are likely responsible for specific consumption estimates that are higher than what would be reasonably expected for the project homes to boil water, thus contributing to potential overestimation of emission reductions.

B. Other parameters of concern

Beyond the parameters that are directly implicated in the overestimation of emissions reduction calculations for SWS projects, our team identified four additional parameters that may either have an indirect impact on the ER estimates or may be overstating the overall development impacts of SWS borehole initiatives. In this section, we compare reported values and monitoring practices for these parameters to the GS standards, methodologies, and guidance, as well as the relevant literature and learnings from key interviews, in order to characterize the nature and extent of the impact of these parameters.

1. Treatment capacity

a) Methodological guidance

Versions 1.0 of the TPDDTEC does not require or address treatment capacity. A rule update dated 23rd July 2015 retroactively added treatment capacity as a monitored parameter over the crediting period for project scenarios to projects using Version 2.0 or later. Section A 3.2 states: *"The treatment capacity limits of project technology/source are required to be monitored to ensure that the water consumption level applied for emission reductions must not be greater than the treatment capacity of the technology/sources. In cases where the default value is more than the treatment capacity of the project technology, emission reductions shall be calculated based on the actual monitored values."*

Additional information is provided in A 3.5:

Data / Parameter:	Treatment capacity
Data unit:	Liter per day
Description:	Treatment capacity of the project technology/improved sources
Source of data:	Manufacturer specification/design specification
Monitoring frequency:	Once at the time of registration or at inclusion of new technology
QA/QC procedures:	
Any comment:	The water volume values used in the calculations of emission reduction must be justified in terms of capacity of the project technology/improved sources.

No additional guidance beyond the two sections above is provided in the TPDDTEC.

b) Reported practice and values

In the case of boreholes, project developers use a water discharge rate, outlined in the technical specifications of the handpumps, to determine the liters per minute produced by the handpump. The section of the technical specifications referred to by the developers points to a "discharge test," which is designed to test the performance of the handpump. The guidance states that "40 continuous full strokes of the plunger in approximately one minute…should be generally not less than 16 liters (ERPF 2007)." In an example we reviewed, the project developer thus assumed the capacity of the handpump to be 16 liters per minute, which was then multiplied by an assumed operating time of 12 hours per day, resulting in a treatment capacity of 11,520 liters per day. Additionally, one project developer noted that the approximate discharge was a function of borehole depth, referencing handpump specifications. In our case studies, however, all treatment capacity

estimates except one referenced the 16 liters per minute value, and no other discharge rates were mentioned in the calculations.[16]

Upon further examination, the Rural Water Supply Network (RWSN), a global network of over 10,000 rural water supply professionals that maintains the international standards for handpumps, notes the discharge rates below for the Afridev (Rural Water Supply Network 2007a) and India Mark ii (Rural Water Supply Network 2007b) handpumps.

Table 4. Discharge rates for the Afridev and India Mark ii handpumps used in SWS borehole projects. Source: Rural Water Supply Network.

Approximate discharge (75 watt input)				
	Afridev		India Mark ii	
Head (m)	m^3/hr	l/min	m^3/hr	l/min
10	1.4	23.3	1.8	30.0
15	1.1	18.3	1.3	21.7
20	0.9	15.0	1.0	16.7
25			0.9	15.0
30	0.7	11.7	0.8	13.3

Therefore, any borehole depths greater than approximately 20 meters for the Afridev pumps and 25 meters for the India Mark ii pumps would have flow rates less than the 16 liters per minute estimate from the project developers. Exact borehole depths could not be found in the project documents, which simply state that depths are less than 100 meters, and no further details could be obtained.

The RWSN additionally assume a household size of 3-5 and a population serviced of 300 people at 15-20 liters per person per day (Rural Water Supply Network 2007b, 2007a). Using the estimate of 300 people at 20 liters per day results in a treatment capacity of 6000 liters per day, about half of the value currently assumed by project developers. Consequently, using the RWSN specifications to estimate treatment capacity can yield a large range of values depending on borehole depth and choice of metric.

c) Comparison with evidence base

Reporting of borehole discharge rates is limited in the research. However, a few studies provide some indication of borehole yield. A study in Kenya of 300 handpumps equipped with Waterpoint Data Transmitters, which continually measure pump use, reported an average mean abstraction of 1500 liters per day (Thomson et al. 2019). A study of 200 handpumps in Uganda rated handpumps based on a functionality metric of at least 10 liters per minute borehole yield. In this report, 34% of handpumps met the 10 liters per minute threshold functionality value (Owor et al. 2017). Finally, a quantitative study of the groundwater storage for the continent of Africa estimated African boreholes could support handpump abstraction of 6-18 liters per minute (MacDonald et al. 2012). Given this limited set of data, the project developers assumption of 11,520 liters per day and 16 liters per minute falls on the very upper end of possible values that could be abstracted from the

[16] Additional information on borehole depths is available in the complete record of stakeholder comments included as an annex to this report.

groundwater, with published field-based results as much as 10 times lower than those reported by GS projects.

Finally, the common project assumption of an operating time of 12 hours per day assumes a full 12 hours of operation without any downtime between users. Measurement of a borehole's runtime is difficult to validate and is limited in the literature. One study did measure runtime using sensors in Ethiopia and Kenya, with reported runtimes of 4.09 hours and 5.90 hours respectively, but these boreholes were using electrically powered motorized pumps yielding much higher discharge rates of 300 liters per minute, so they are not comparable to handpumps (Thomas et al. 2019).

d) Conclusion

Using the manufacturer's technical specifications to fulfill the treatment capacity requirement is not as straightforward as it may seem initially and requires more evidence of borehole depth and run times to be vetted effectively. The RWSN specifications give room for multiple interpretations, and current project estimates appear to be based on a non-conservative best-case scenario, resulting in possibly as much as a 100% overestimation.

2. Water quality

a) Methodological guidance

Annex 3 of the TPDDTEC states "safe water is defined as water, which is both clean and consumed hygienically." Section 3.3.E further outlines guidance for water quality:

> **"Water quality testing**: Water quality must be tested every quarter, with the first test within 6 months of the stated project start date. In addition, PPs shall ensure that water quality is tested at least once during seasons where there is a high chance of contamination, for example, the rainy season. Local non-accredited laboratories can do the quarterly water quality testing. However, at least once every two years, accredited laboratories must perform the water quality testing. If accredited laboratory results differ materially from non-accredited laboratory results, testing with the aberrant non-accredited laboratory must be discontinued. If local labs conduct the testing, the testing protocol should be provided to the DOE for validation. Also, in any case where the national laws on water quality testing are more stringent, these national standards apply.
>
> **Water quality standard**: As a first option, projects shall meet host country standards (where available) for treated water quality. Where national standards are not available, projects shall meet WHO standard of less than 1 Colony Forming Unit (CFU) of E.Coli /100 ml. The 90/10 precision rule must be followed in calculating the sample size required for testing water quality. For 'point of use' technologies such as water filters, the quality testing shall be done for samples taken at the water outlet. For boreholes and chlorine dispensers, testing shall be done for samples collected at source as per national/or the above-mentioned criteria. Also, the monitoring of hygienic use of water at the user end shall further complement the testing process."

Project developers are additionally required to carry out Hygiene Campaigns as defined by:

"PPs need to carry out and provide evidence for hygiene campaigns. The following guidelines may be adopted for conducting these campaigns:
• Hygiene refers to access to sanitation amenities, equipment and infrastructure, as well as to the behaviour in respect to regular and correct use of such amenities. It also refers to behaviour that prevents infections from water-related diseases.
• The project proponent shall report the activities conducted each year in the annual monitoring report. Any major changes in the health status of the water users as a result of contaminated water (e.g. an outbreak of water related disease) must be reported and, if relevant, a strategy put in place to address it through the hygiene campaign.
• The detailed method used to assess hygienic handling of clean water must be provided with the PDD and validated by the DOE.
• The details of the method should be adjusted to suit the circumstances of each project and also to suit learning year on year. "

b) Reported practices and values

Project developer documents assert borehole water quality meets the standards outlined in the methodology.

c) Comparison with evidence base

It has been well established, with concerns echoed by the water experts we interviewed, that bacteriological water quality declines from the source to the household (Wright et al. 2004). Thus, by the time a household uses the water extracted from a borehole for drinking, basic personal hygiene, and food preparation, it is less likely to meet the definition of safe water. In requiring water quality measurements only at the source, however, the GS methodology does not protect against this well-known risk.

Implementing a hygiene campaign may result in improved safe water storage and consumption practices in the home. Because hygiene campaigns require a behavior change in the household, however, they are often difficult to implement and require comprehensive and frequent follow-up to be successful.

Lastly, the methodology emphasizes microbiological water quality except in the case where national standards require other testing, even though some chemical elements are also known to cause health risks. It should be noted that some national requirements do include standards for chemical elements, which are tested for by projects in those locations. However, two of the water experts interviewed for this report expressed concern that several chemical elements, including fluoride, arsenic, lead, and nitrate, were generally underassessed in the evaluation of community-level SWS initiatives. While fluoride, arsenic, and nitrate are well known groundwater contaminants, the water experts suggested lead is an emerging issue, as it may sometimes be introduced through the corrosion of plumbing materials.

d) Conclusion

For community-based systems, such as borehole programs, neither microbiological nor chemical quality are ensured for project users at the point of use under the current protocol, thus creating a risk that the development impacts of the SWS portfolio are being overestimated.

3. Distance to boreholes

a) Methodological guidance

V2 of the TPDDTEC added Annex 3 outlining that "water in its improved form should be available within 1km walking/pedaling distance from the households. There is a two-year grace period (from date of registration) for any households falling outside of this distance; however, once this period is over, these households would not be included in the emission reduction calculation." This metric is relevant because distance to an improved water source is a primary component of access, which is central to the Global Goals for Sustainable Development (SDG) target 6.1, "Safe and Affordable Drinking Water."

b) Reported practices and values

Project developers we spoke with currently record the household's distance to the borehole during the water consumption field test (WCFT). Of the WCFT documents we reviewed, there was no specific survey question used to collect this information, only a field to fill in for the distance, so it is not clear how the question is being asked of the household. Of the approximately half dozen projects we reviewed closely, in many cases a large percentage of households were located right at the 1 km distance. We did not find any evidence that these distances are further validated, but a project developer did mention that they only included households as users if they were within the 1 km distance, even if there were households using the borehole beyond the threshold distance.

c) Comparison with evidence base

Studies that compare self-reported distance to actual distance to boreholes show little correlation (Cassivi et al. 2021), with one study mentioning that households would not estimate distance because they did not feel like they could do so accurately (Pearson 2016).

d) Conclusion

The number of households located within the project boundary has an impact on the number of users, which is a potent parameter within the ER calculation, and also indicates contributions to SDG 6 on availability and access to water services. There is moderate concern that the lack of measurement guidance for this parameter may allow for overestimation of emission reduction and development benefits.

4. Water quantity

a) Methodological guidance

The quantity of safe water is determined using the WCFT as outlined in section A3.3E: "The water consumption field test (WCFT) is similar to the FT, except project-supplied clean water consumption volumes and boiling is measured rather than fuel consumption. The WCFT is conducted with end-users representative of the project scenario target population and currently using the project technology." The value is capped at 7.0 liters *per capita* per day in the current methodology version (7.5 liters per capita per day in V1).

b) Reported practices and values

In implementing the WCFT, project developers ask households where they obtain the water they use for the three different credited uses (drinking, basic personal hygiene, and food preparation). The volume of water used by the household is tracked over a three-day period and then averaged to calculate the liters *per capita* per day. All WCFTs we reviewed reported values over the 7.0 liters *per capita* per day capped value.

c) Comparison with evidence base

Overall, it is difficult to estimate a reasonable value for this parameter given that there is little reporting on the combination of these three uses specifically. World Health Organization guidance considers 7.5 liters *per capita* per day to be the minimum water necessary for drinking and cooking (Howard and Bartram 2003)[17], though some water for cooking is likely to be boiled and would thus not be considered as a credited use.

d) Conclusion

As noted in the prior discussion on percent usage, the definition of the three credited uses is not well defined. Nonetheless, the capped water quantity value of 7 liters per person per day is reasonable based on the results of the WCFT's reviewed in this investigation.

C. Magnitude of overestimation / over-issuance

1. Overestimation of emission reductions in SWS portfolio

Potential overestimation in the portfolio was estimated by comparing emission reductions assuming input parameter values based on a combination of literature, expert interviews, first-principles, and sense checks with the averages for those same inputs in the project portfolio. Table 5 below summarizes the inputs, providing a comparison for each parameter as well as the overall emission reductions. Note that average project values are weighted by the number of VPAs using the respective parameter estimates, and these estimates are only based on those from borehole projects, which constitute 90% of the SWS portfolio. Further note that the comparison parameter values were derived solely for illustrative purposes within the context of this report. It should not be assumed that they are proposed here as future reference values for SWS project certification; developing such reference values is outside the scope of this investigation.

[17] Update expected in 2020.

Table 5. Summary of project and comparison parameter estimates, and implications for emission reductions overestimation of borehole SWS projects. Source: Berkeley Air.

Parameter*	Mean project estimate	Non-conservative comparison estimate	Regular comparison estimate	Non-conservative difference§	Regular Difference§
Fuel to boil water	0.83 (kg/L)	0.5 (kg/l)	0.4 (kg/l)	66%	108%
Users per technology (borehole)	630 people	500 people	300 people	26%	110%
Technology days	365 days	347 days (95% of year)	274 days (75% of year)	5%	33%
Usage Rate	100%	90%	75%	11%	33%

Emission reductions	Project	Non-conservative Comparison	Regular Comparison estimate	Non-conservative difference§	Regular Difference§
No cap	16,259 (tCO2e/year)	6646 (tCO2e/year)	2101 (tCO2e/year)	145%	647%
Cap	10,000 (tCO2e/year)	6646 (tCO2e/year)	2101 (tCO2e/year)	50%	376%

§Difference = ([project estimate-comparison estimate]/comparison estimate) *100%

*Assumed other parameter values, which were kept constant for both the average project and comparison estimate: Default IPCC emission factors (112 tCO2/TJ for CO2 and 8.962 tCO2e/TJ for non CO2 species [CH4 and N2O); fraction using safe water in baseline (0); quantity of safe water used per person (7.5 l/person/day); quantity of raw water used per person (0 l/person/day); fNRB (0.90); percentage of suppressed demand users (95%); number of boreholes (7).

Two comparisons are provided. The “non-conservative comparison” set of parameters is meant to represent a possible best-case scenario for the average project, whereas the “regular comparison” estimate is deemed to be a more likely set of input parameters. For the specific consumption estimate, 0.5 kg/l is approximately what would be required to boil a liter of water using a three-stone-fire at 7kW (and 10% thermal efficiency) and have it continue boiling for 10 minutes, and 0.4 kg/l would be required at 5kW. 500 borehole users is a higher-end estimation based on technical specifications of the typical pumps and assuming approximately 12 liters per capita per day of use. For comparison, 300 users is referenced as an upper limit in Ghanaian Water Standards (Fisher et al. 2015), as well as RWSN specifications (Rural Water Supply Network 2007b, 2007b). The regular comparison estimate of pumps being operational for three out of four days is supported by the literature (Foster et al. 2019), and being operational for 95% of days is a reasonable upper limit given that the best example of pump-maintenance in the literature was 95% (McNicholl et al. 2019). For usage rates, the literature suggests that people use several water sources and that these vary over time. A value of 90% was used for the non-conservative comparison estimate, as literature has reported the borehole as the primary source up to an 86% value (Thomson et al. 2019), while 75% was used as the regular comparison estimate given the range in the literature (Foster et al. 2019).

Given these assumptions, the average project emission reductions are overestimated by 145% compared to the non-conservative comparison scenario and by 647% compared to the regular comparison scenario. In practice, any micro-scale project estimating reductions over 10,000 tCO2e/yr has their issuance for that monitoring period capped at 10,000 tCO2e/yr. Under this capped scenario, the corresponding overestimates are 50% and 376% compared to applying the non-conservative and regular comparison parameter inputs,

respectively. Note that the capped scenario is more indicative of the level of over-issuance that may be occurring, while the uncapped scenario provides an indication of the potential overestimation that the data collection protocols and emission reductions calculations can produce.

Importantly, there is no way to precisely and accurately determine what the potential overestimation of emission reductions from the portfolio is, as this would require retrospective independent study of the individual parameters using best-practice methods. At the same time, the comparison inputs are well within the range of what could be expected given their grounding in reported studies, the context provided by experts, and common-sense checks. So, while it is impossible to say there is a specific percent overestimation, the range of 145%-647% (50%-376%, capped) suggests there is substantial emission reductions overestimation in the SWS portfolio.

It is also important to note that the emission reductions overestimation is a result of the overestimation of individual parameters compounding, as they are multiplied together in the emissions reduction calculation. This compounding of overestimation across several parameters means that no single parameter could be singled out as the sole or even primary issue to be resolved.

2. Case study analysis results

Comparisons of parameters and emission reductions between three case study project values and those deemed to be more realistic based on literature (comparison estimates) are presented in Table 6. This case study analysis is presented as a companion to the analysis of the potential for overestimation across the entire SWS portfolio, presented in the preceding section. This supplemental case study approach is presented because it allows for a deeper examination of the impact of outlier parameter values within actual project scenarios. Note that the case study analysis has no quantitative relationship to the portfolio-wide assessments presented in the preceding section or to the following section, addressing the materiality of the portfolio-level assessment to GSF operations.

The overall case study emission reductions were estimated to be 327% to 762% greater than the comparison estimate when the 10,000 tCO2e/year cap was not considered (139%-338% greater with the cap). As in the case of the overall portfolio analysis, the overestimation was compounded through multiple parameters being larger than the comparison estimates. The equation for baseline fuel consumption, for example, is shown below and illustrates how the number of person-days (number of users x operational days), specific consumption, and water consumption multiply together. Usage rate is also a multiplier in the final emission reductions calculation and applies as a simple linear adjustment.

> Bp,y = Number of person-days x baseline fuel used to treat Water (t/l) x total safe water consumed in project scenario (l/p/d). (TPDDTEC, section A3.1, pg. 45)

As these parameters multiply, a 50% overestimation in one and a 50% overestimation in another, for example, would propagate, resulting in an overall overestimation of 125%. Note that the reader is referred to TPDDTED section 3.1 for a full description of the equations used to calculate emission reductions for SWS projects.

Table 6. Summary of case study project and comparison parameter estimates, and implications for emission reductions overestimation. Source: Berkeley Air.

Rwanda (4 borehole pumps, TPDDTEC v 3.1, 1 monitoring period)

Parameter*	Project estimate	Comparison estimate	Difference§
Fuel to boil water	0.946 (kg/l)	0.43 (kg/l)*	120%
Users per technology (borehole)	1351 people	477 people#	183%
Technology days	365 days	274 days (75% of year)+	33%
Usage Rate	100%	90%&	12%

Emission reductions	Project	Comparison estimate	Difference§
No cap	19,679 (tCO2e/year)	2284 (tCO2e/year)	762%
Cap	10,000 (tCO2e/year)	2284 (tCO2e/year)	338%

Uganda (5 borehole pumps, TPDDTEC v 1, 4 monitoring periods)

Parameter*	Project estimate	Comparison estimate	Difference§
Fuel to boil water	0.697 (kg/L)	0.50 (kg/l)*	39%
Users per technology (borehole)	1109 people	510 people#	117%
Technology days	365 days	274 days (75% of year)+	33%
Usage Rate	100%	87.5%&	14%

Emission reductions	Project	Comparison estimate	Difference§
No cap	16,596 (tCO2e/year)	3592 (tCO2e/year)	362%
Cap	10,000 (tCO2e/year)	3592 (tCO2e/year)	178%

Eritrea (6 borehole pumps, TPDDTEC v 1, 4 monitoring periods)

Parameter*	Project estimate	Comparison estimate	Difference§
Fuel to boil water	0.833 (kg/L)	0.50 (kg/l)*	77%
Users per technology (borehole)	701 people	530 people#	32%
Technology days	365 days	274 days (75% of year)+	33%
Usage Rate	100%	87.5%&	14%

Emission reductions	Project	Comparison estimate	Difference§
No cap	17,901 (tCO2e/year)	4189 (tCO2e/year)	327%
Cap	10,000 (tCO2e/year)	4189 (tCO2e/year)	139%

§Difference = ([project estimate-comparison estimate]/comparison estimate) *100%

*Assumed based on firepowers of 5.9kW (Rwanda), 7.5kW (Uganda), and 7.5kW (Eritrea). See Recommendations section for more details on use of firepower to estimate specific consumption.

Based on assumption that pump can produce 6000 l/day and the clean water consumption by users, as provided in project documents (Qp,y).

+ 1 in 4 handpumps in sub-Saharan Africa are non-functioning according to (Foster et al. 2019).

& Adjusted by assuming 75% use (Kelly et al. 2018) of borehole during rainy seasons (Rwanda 5 months; Uganda 6 months, Eritrea 6 months).

Fuel to boil water and the number of users were the two largest contributors to the overestimation, with differences ranging from 39-120% and 32-180%, respectively. All three VPAs measured the fuel to boil water using a similar protocol, in which the stove operator was instructed to bring one liter of water to a “roiling boil” and continue that roiling boil for 10 minutes. The associated firepowers for those tests were 15.8kW, 10.8kW, and 11.3kW for Rwanda, Uganda, and Eritrea, respectively. In the case of Rwanda and Uganda, there are field studies from those countries reporting firepowers of 6.9kW and 7.5kW for TSFs, respectively, which are substantially lower than those from the BWBTs.

Number of users was also determined with similar approaches, relying on local implementation partners and community organizations/leaders to create rosters of users and cross-checking with perceived distance to the borehole. Handpump studies have cited values around 100-300 (Fisher et al. 2015; Holm et al. 2017). Our assumptions based on the capacity of the wells and demand suggest a lower number of users would be likely for these case studies, although in the case of Eritrea, our estimate was only different by 32%.

Project technology days and usage rates had slightly lower impacts, though in all case studies, the usage rates were reported as 100%, and pumps were reported to be operational for every day in the project period. While the projects have an incentive to maintain pumps, and maintenance logs did report minor issues being fixed with no downtime, the literature suggests that borehole handpumps have a more common downtime of 25% (Foster et al. 2019). The usage rates of 100%, which were determined via simple surveys, also seemed unrealistic for these three case studies.

Overall, these case studies suggest that overestimation can arise from several different parameters, though fuel to boil water and number of users had the highest potential overestimates. It is also important to note that there was no evidence the project developers did not follow the methodologies and that the parameter values used in these VPAs were ultimately allowed to be used through multiple layers of oversight in the certification process.

D. Assessment of Materiality

The concept of materiality has several definitions depending on the context. In the investment context, “materiality” refers to any information that if publicly known would change the value of a product, investment, or company. Similarly, an ISO publication comparing its social responsibility guidelines (ISO 26000) with OECD guidelines, echoes this definition with the statement, “Material information can be defined as information whose omission or misstatement could influence the economic decisions taken by users of information.” (ISO Secretariat et al. 2017). We conclude that in reference to either of these definitions, the findings of overestimation from this investigation are material.

The analysis here suggests a systematic overestimation and subsequent over-issuance of emission reductions in the borehole projects in the SWS portfolio. For illustrative purposes, a conservative range for the over-issuance of micro-scale borehole projects’ emission reductions can be estimated by applying the overestimation percentages for the two comparison estimates (non-conservative and regular) for capped projects in Table 5[18]. Total emission reductions credits issued for micro-scale borehole projects in the SWS portfolio to date was approximately 3,150,000 tCO2e, which would imply that a range of approximately 660,000-2,100,000 tCO2e in reductions would have been more realistic (1,050,000-2,490,000 tCO2e difference).

As with the percentage overestimations in Table 5, this range is only meant to be illustrative of the potential over-issuance, as it is not possible to go back and measure the project

[18] The projects in the portfolio include those which have and have not had years capped at 10,000 tCOe/year, and therefore applying the relative percentage difference under the capped scenarios provides a conservative estimate in terms of the magnitude of overestimation.

parameters and calculate the potential differences in emission reductions estimates on a project-by-project basis. Although these over-issuance estimates are not meant to be precise, they still suggest a magnitude of over-crediting that should likely be considered material given the importance of credibility in the offsets, as well as Gold Standard's principle of erring towards conservative estimates. The overestimation is likely to be systemic, as it is ingrained in multiple components of the mechanics of the methodology and its application, rather than being potentially the result of a single calculation error or misrepresentation. Further, the recent rapid growth in the portfolio of SWS activities has contributed to the materiality of the overestimation. Almost 60% of the SWS projects have been added to the portfolio in the last 3 years, so that the overestimation problem is currently being replicated across these many newly registered projects.

In evaluating the materiality of over-issuance, it may also be useful to consider these findings in the context of the suppressed demand scenario, which is applicable to 94% of the SWS projects, according to the GS Secretariat. The suppressed demand approach allows carbon finance to be targeted at the most resource-poor communities, where firewood and/or financial constraints prevent proper water treatment in the baseline. The suppressed demand scenario for SWS projects currently allows 7L of clean water per person per day (down from 7.5L in earlier version of the methodology). A 2014 GS Climate Policy Brief further explained, "… for clean water supply projects in a rural context, where the households are not connected with a piped water supply network, boiling water can be assumed as the baseline technique to treat water but taking into account the penetration rate of non-GHG emitting treatment techniques like chlorine tablets" (Michaelowa et al. 2014). Without implying any judgement on the validity of this approach, we note that the use of a suppressed demand does have implications for the actual climate impact of over-issuing credits from the projects. Therefore, the perceptions of and remedies for an over-issuance of credits representing actual reduced emissions may not be the same as the over-issuance of credits representing averted future emissions from projects with suppressed demand baselines.

V. Part B Findings

While the bulk of emphasis in the TOR for the investigation of the SWS grievance was placed on a technical assessment of whether over-estimation had occurred, and if so to what extent and degree of materiality (Part A), the Berkeley Air team was also asked to report any findings on the research questions in Part B, if applicable. Specifically, input was sought on questions 4 and 5: where in the overall certification process the cause(s) of the over-estimation may originate; and what degree of responsibility for these causes can be attributed to specific stakeholders or parties within the GS ecosystem. The findings related to these two questions are presented in this section, with the caveat that these are observations rather than firm conclusions, due to significant time constraints on the investigation, the focus of which was primarily on Part A of the assignment.

A. Observations on causes of over-estimation within the certification process

The application of the complex TPDDTEC methodology and impact certification process create many opportunities for over-estimation to occur. Three layers of problematic process have been identified: methodological weaknesses, shortcomings in project developers' monitoring activities, and failings in the oversight system.

1. Methodological issues

As described extensively in Section IV, the methodology is not prescriptive or restrictive enough for some specific parameters. The weaknesses of particular parameters have already been explored, but it is worth noting that when viewed as a whole, the project developers' monitoring resources aren't necessarily directed to the parameters that are most likely to vary at the project level. For example, the requirement to conduct water boiling tests for each PoA requires significant project developer resources, yet in 94% of cases, the baseline purification of water by boiling is a representation of a scenario that would only be expected to occur if households had fewer resource constraints, rather than the measurement of an actual baseline practice. By contrast, field measurement of changes in actual project scenario parameters, such as the number of borehole users or the operational status of project boreholes, are subject to less intensive verification requirements.

The methodological basis for requiring field measurement for some parameters may also be based on an outdated perception of a limited evidence base. While it is the case that GS SWS projects already take several forms and exist in multiple geographies, this does not necessarily equate to infinite variation or unpredictability in parameter values. When the TPDDTEC methodology was first developed in 2007-08, the literature on several of the key parameters was much sparser than it is today. A common narrative of the time asserted that climate and development impacts were frequently oversold because the estimates of impact were based on controlled laboratory test that were not good predictors of real-world performance. Consequently, field measurements became a hallmark of robust impact measurement.

More than a decade later, some parameters have been well researched, and the resulting literature reveals values that fall within a narrower range than may have originally been expected (for example see published BWBT values summarized in Figure 8). While some

project developers strongly advocate for monitoring all parameters due to the near infinite variation in context, analysis of the project documentation reveals some very consistent values for parameters such as borehole usage and percent days the safe water source is operational. It is ironic, but not inconceivable, that relying heavily on field measurements, which was originally designed to improve the accuracy of the impact assessment methodology, may instead create the risk of less accurate ER estimates due to the challenges inherent in field monitoring, discussed below.

Another weakness in the GS rules may be the lack of a robust de-bundling requirement for micro-scale projects. The current portfolio contains groups of micro-scale "clone projects" that are so similar in terms of region, implementing partner, intervention technology, population served, etc., that it is hard to discern why they are registered as separate micro-scale projects. The history of these projects regularly generating more emissions reductions than the 10,000-ton ceiling that defines the micro-scale scheme also suggests that these projects may not be appropriately classified.

Finally, although not a methodological weakness, part of the cause of overestimation lies in the compounding of individual inflated parameters. The application of a methodology that was originally designed to assess household-level technologies to community-level borehole projects has exacerbated this problem by requiring monitored inputs that are not necessarily optimally aligned with how community assets are actually used. For example, there are multiple input parameters that connect the quantity of water extracted at the borehole with the ultimate consumption at the household level, all of which have uncertainty and thus give rise to the opportunity for misestimation. As highlighted in Section IV.C. on overestimation, the compounding of overestimation through the emission reduction calculation from several parameters means that no single parameter could be singled out as the one to address.

2. Shortcomings of project developer monitoring

The second component of problematic process occurs when project developers apply the methodology. It is the norm to have a chain of parties implementing project monitoring requirements, which leaves multiple connection points where misreporting can occur. For example, data on the operational status of boreholes might come through a community member or a local technician reporting a malfunction, to a local repair entity reporting it to the local implementing agency, to the local implementer reporting to the project developer. Similarly, for metrics such as the number of beneficiaries, use of other sources, and distance from the borehole, data is self-reported by households or community-level partners, making it susceptible to social desirability bias[19], especially in communities with a long history of receiving development assistance. Ultimately, there are practical and potentially even ethical limits to the extent of ongoing tracking and influencing of communities that are

[19] Social desirability bias is a social science term referring to the tendency of survey respondents to respond to questions in ways that they perceive will please the enumerator or researcher. For example, in the case of SWS projects, respondents may over-report use of the safe water technologies in hopes of receiving further benefits or resources. Additional information on survey respondent bias is available in the complete record of stakeholder comments included as an annex to this report.

geographically remote from the project developers over a span of 5 or more years of project operations.

3. Limitations in the certification process

The third and final component of problematic process potentially contributing to the overestimation is the implementation of the micro-scale oversight system, which has fewer requirements than the schemes for larger projects.[20] Note that, in consultation with the TAC-SWSGWG, and given the time constraints, and unexpected complexities in accessing data and conducting reviews, the Berkeley Air investigation focused primarily on technical considerations presented in Section IV rather than on the investigation of the certification process. Nonetheless, our research did find broad recognition that the majority of SWS projects being certified under the micro-scale scheme do not reflect the scheme's original intent, which was to provide pathways to the carbon markets for small projects that would not otherwise be viable. The current system allows project developers to aggregate multiple smaller "cloned projects" into regional conglomerates (within even larger PoAs) with total annual tonnage that would usually trigger more robust certification requirements, such as those applied to large-scale projects. Intentionally or not, this practice is subverting the risk mitigation measures designed to ensure that large-scale projects do not overestimate climate or development benefits. We also note that the presence of some small- and large-scale SWS projects suggests that the existence of all SWS projects is not dependent on being able to use the micro-scale scheme in this manner.

Another element of this third component is a potential misalignment of reviewer attitudes, skills, and training with the demands of providing a robust substitute for the third-party validation of project design and verification of project performance required of larger projects. Although the process allows SC to hire a VVB to perform these functions for micro-scale projects, to date they have instead exercised the option to use internal staff or external expert contractors for this role[21]. In general, we found that both the SC reviewers and the accredited VVB with whom we spoke approached project review with the priority goal of ensuring the rules of the standard and the methodology requirements had been met.

The SC reviewers and the VVB were aware that the reasonableness of the resulting parameters was also a component of their mandate, yet in our limited assessment, we found significant variation in the norms around how to assess reasonableness and what actions to take as a result of this assessment. There was general consensus among the interviewed parties for the need to reform the BWBT parameter, but this common recognition had not resulted in a consistent response: the auditor reported that auditors are the "eyes and ears" for the standards bodies; the external reviewer had approached SustainCERT with his concerns; and the internal reviewer had applied more intensive scrutiny and suggested more reasonable values to the project developers for this parameter. However none of the interviewed parties were aware of any values that had been adjusted downward as a result of their concerns. This variation in the approaches to vetting parameter values is possibly due in part to the breadth and depth of expertise and

[20] Please see the complete record of stakeholder comments included as an annex to this report for additional information on the effects of the micro-scale scheme.
[21] Please see the complete record of stakeholder comments included as an annex to this report for additional information on the training and accreditation of certification personnel.

training needed to fully analyze the purpose and characteristics of all the parameters used in the calculation of ER for SWS projects.

B. Observations on sources of responsibility for over-estimation

In preparation for examining responsibility, the BA investigation team found it useful to summarize the findings on the perception of the value of Gold Standard certification, as these attitudes may shape the behaviors of various parties. Uniformly, the stakeholders interviewed for this investigation perceive value in the Gold Standard Foundation and the certifications that it confers. All stakeholders agreed that the "rules," i.e. the standards and methodologies, matter because the robustness of these rules is what results in higher value certifications (either monetarily or ecologically/ethically or both). This universal perception of value in the rules also results in shared support for improvements to these rules, as these sustain and enhance the value of the GS certification.

At the same time, there is evidence of significant differences across stakeholders in perceptions of the purpose of the Gold Standard and its schemes, as well as norms around profit-making and the degree of ethical responsibility born by project developers for accurate estimations of credits and development benefits. The Gold Standard rules explicitly place the responsibility of reporting truthful and reasonable parameter values on developers, and our investigation did not uncover any evidence of intentional misreporting. To the extent discernable, the findings show developers conforming to the rules and methodology requirements and the certification process. Nonetheless, there is evidence that project developers interpret the responsibility for reasonableness in various ways, and not all prioritize conservativeness. In one case, a developer pro-actively sought and received permission to reconduct their BWBT using a protocol that they felt was more representative of local practice and that also resulted in a less conservative parameter value. In another case, the project developer described vetting their BWBT results with estimates from the literature to confirm reasonableness. The range in the degrees of responsibility taken by developers for the reasonableness of their monitoring results suggests an over-reliance by Gold Standard on assumed norms, which are less effective than explicit prescriptive guidance, in a system where stakeholders do not share the same frames of reference or incentive structures.

Another finding relevant to the apportionment of responsibility for over-estimation was a shared sense across stakeholder groups that the current certification system is not supporting GS as well as it could. Project developers report frustration with the inconsistent application of the standards, methodology, and certification processes. Different reviewers assigned to the same project at various stages or to similar "clone projects," will reportedly return strikingly different requests for clarification or corrective actions. The investigation team's own limited review of the project documentation files resulted in the observation that documentation could be more standardized and streamlined, supporting the finding of the need for improvements in SustainCERT's certification operations. Furthermore, the recent and significant increase in the growth of *de facto* large-scale projects being routed through the micro-scale system may also be straining the SustainCERT's operations, which were arguably designed to fulfill a different purpose. SustainCERT may choose to employ accredited VVBs rather than reviewers and objective observers, which might alter the

consistency and rigor of the oversight of micro-scale projects, but they have so far not chosen to do so.

Finally, many stakeholders expressed support – some strong and some measured – for the idea that a project developer that has followed the certification process through to the end and ultimately achieved project certification and credit issuances should not be punished if weaknesses are subsequently discovered that led to overestimation. As noted earlier, there is consensus that forward-looking changes are appropriate and welcomed, but there is much less agreement on how to handle retrospective action and who should take responsibilities for its consequences. If the responsibility for retrospective action is placed exclusively or even significantly on project developers, the GS should expect that such action would likely bring legal challenges.

VI. Recommendations

A. Methodological

Revisions to the current methodology would likely address some, though not all, of the factors that contributed to the overestimation of SWS impacts. Many of these recommendations point to the need to evaluate community-level projects differently than household technologies and approaches, suggesting that perhaps the current TPDDTEC methodology should be split into two separate standards that would allow a more streamlined fit-for-purpose approach to each category.

1. Water Quantity
    - For point source projects such as borehole or kiosks, it could be more straightforward to place the focus of monitoring on directly estimating the amount of water provided. There are sensor-based approaches or simple observation sampling (e.g. counting the number of pump strokes), which, despite having their own set of challenges, may be preferable to the current approach. These approaches would mitigate the need to estimate the number of users and to conduct the water consumption field test, although testing of water quality and technology days would still be needed, if not captured by the sensor or sampling approach. Another benefit of more direct quantification of the water supplied would be that it could be easily combined with the first principles approach[22] to estimating the fuel use required to heat the quantity of dispensed water (see item 7 below).
    - Alternatively, a more detailed/prescriptive methodology for the WCFT would be useful, which could help quantify appropriate uses, as well as provide guidance on measurement of multiple sources.
2. Number of users
    - A more rigorous field process for determining users, with checks at the source and household level, could be used, including oversight verifying results with random samples. The standard could provide more guidance on best practices for monitoring this parameter.
    - Alternatively, a reasonable cap on the number of users may be applied.
    - User lists should be revised and verified year over year, as families go through transitions and experience changes in their water demands.
    - Enhanced surveying methods, ideally based on or validated by GPS/GIS, would help to ensure that users are within 1km (walking distance) of the safe water source.

[22] In this context, a first principles approach is one that starts directly from established laws of science, does not make any assumptions, and cannot be simplified any further.

3. Usage percentage

- Enhanced surveying is needed to account for seasonality, and to determine when and how much other sources are being used for the creditable uses.
- The community-based projects should have access to the same level of guidance as the household treatment projects receive through the HWT appendix.
- Additional clarity on what it means to be a "100% user" would help to improve the accuracy of usage rate measurements. A more specific or more operationalized definition of basic hygiene tasks would aid in including/excluding water uses that may contribute towards the definition of 100% use.

4. Project technology days

- More rigorous monitoring and documentation of borehole downtime is necessary.
- Monitoring borehole usage with sensors, as mentioned in the section on measuring water quantity, could likely provide estimates for this parameter and/or simply be subsumed by the sensor-based estimate of water quantity provision. GSF could play an important role in educating project developers on sensor options, costs, and benefits, and supporting their transition to real-time monitoring.
- The 100% or near 100% operational days for borehole hand pumps are a significant departure from what is described in the literature, and from the impressions of the relevant experts interviewed for this report. However, a handful of new publications have documented innovative approaches achieving high operational levels for boreholes in demonstration projects. It would be highly beneficial to replicate these results across multiple contexts and geographies to understand best practices and key success factors.
- A default value, based on the literature, could be considered. Robust field-based data could then be provided to prove higher functionality and incentivize project developers to maintain high performance.

5. Treatment capacity

More clarity and guidance are needed on how to calculate treatment capacity. Relying solely on the manufacturers technical specifications introduces ambiguity with regard to which metric/value to use.

6. Water quality

- Sampling at the household (at closest storage point to use) should be added for borehole projects, if the intention is to provide safe and clean water for consumption.
- Chemical water quality parameters should be added to borehole-level testing, particularly to ensure lead is not being added to the water.

7. Fuel to boil water

The current guidance for measuring this parameter is minimal and allows for a large range of outcomes, depending on the protocol used and/or how the project-specific protocol is applied. While being able to capture the variability in the parameter across projects in theory and provide for more representative estimates, in practice the flexibility in estimation of this parameter has resulted in estimates that are unrealistic, in comparison to comparable field-based studies. There are several potential options which could provide more accurate or conservative estimates.

- Use a first-principles approach to determine how much energy would be required to boil (and potentially simmer) the total clean water estimated to be provided by the project. This approach would remove the need to do any direct measurements. An example calculation is provided below (Table 7), showing an estimate of wood required to boil one liter of water, which could be scaled to any amount that the project is demonstrated to have provided. Note that the estimate for one liter is 0.39 kg of wood is at the lower end of what has been used in the VPAs.

Table 7: Example calculations of how fuel use could be estimated from first principles of heating water. Source: Berkeley Air.

First principles approach		
Water end (clean for use)	1	kg
Stove efficiency	10%	%
Water loss (%)	10%	%
Heat of Vaporization	2260	kj/kg
Specific heat of water	4.186	kJ/kg per deg C
Water start	1.1	kg
Temp difference	75	C
Energy to boil	349	KJ
Energy to vaporize	251	KJ
Energy for task	600	KJ
Energy required	5999	KJ
Wood energy per kg	15400	KJ/kg
Wood per liter	**0.39**	**kg**

Assume the firepower and efficiency of a given stove based on literature or field data, and apply those values to determine how much fuel would be required to boil water for 10 minutes. This approach, presented in Table 8, could be based on values from available literature for the respective regions. Note that 1 kg of water is used in the calculation, as 1 kg of water is equal to 1 liter of water at standard temperature and pressure, and specific heat of water and vaporization are in units of energy required (kJ) per unit mass (kg).

Table 8: Example calculation of how field-based firepower of three-stone-fires could be used to estimate specific consumption for boiling water. Source: Berkeley Air.

Firepower approach		
Water start	1	kg
Firepower	7	kW
Stove efficiency	10%	%
Boiling time	10	minutes
Time to boil	7.48	minutes
Specific heat of water	4.186	kJ/kg per deg C
Water loss (%)	10%	%
Temp difference	75	C
Energy delivered (bringing to boil)	314	KJ
Energy used (bringing to boil)	3140	KJ
Wood used (bringing to boil)	0.204	kg
Energy (continue boil)	4200	KJ
Wood used (continue boil)	0.273	kg
Wood per liter (no water loss)	**0.48**	**kg/L**

- Place a cap on the parameter estimate, such as has been done for the period of this grievance (0.4 kg/l). If a cap is introduced, it should be based on first principles or grounded in the best available evidence of real-world stove performance.
- Provide more prescriptive guidance on how the BWBT is conducted, such as a more realistic amount of water to boil (e.g. minimum 5 l), and how to gauge or use firepower as a diagnostic measure (e.g. exclude tests with firepowers out of a sensible range).

B. Certification Process

At this stage in the investigation, there are three primary recommendations that our team can offer concerning the certification process. The first concerns the use of the micro-scale scheme to register clusters of projects equivalent in scope and operation to large-scale projects. Although some stakeholders view this practice as an innovative way for project developers to prosper while also delivering development benefits and income generation to the most resource-poor and remote communities, it is also clearly a subversion of the GS risk-management measures. The Gold Standard urgently needs to address the enhanced risk that derives from allowing large-scale projects to proceed without commensurately stringent third-party reviews. A first step in this process could be to institute a de-bundling test and transition the existing clusters of cloned projects to small- or large-scale projects, as appropriate.

Second, given the complexity of the current SWS methodology, it would also be helpful if the project monitoring report template incorporated a cover sheet summarizing the project's monitored values to help focus the reviewers' attention on the most critical

parameters. To further support reviewers and auditors, we recommend that a manual be provided and regularly updated, offering reasonable reference values and key considerations for important parameters. Such a manual might also offer guidance on conducting sense-checks, where those are available and relevant.

The final recommendation concerns the certification system more broadly. In general, the checks and balances put in place by GS to ensure a robust and accurate certification system proved slow to catch the problem of overestimation in the SWS portfolio, and therefore these should be comprehensively reviewed and improved as quickly as possible. In particular, it seems that there could be a more explicit and transparent mechanism for identifying and addressing unreasonable parameter values.

C. Areas for further investigation

1. Supplemental fieldwork

The current investigation did not include any field visits, which could add significant depth and nuance to the understanding of SWS project implementation and monitoring. It would be valuable, for example, to repeat some of the water boiling tests that resulted in high outlier values for the baseline fuel value to better understand what local circumstances and/or practices may have caused these variances. Water quality tests could also be performed to check for deterioration in water quality between the community source and the point of use as well as to check for chemical contamination. Independent checks on borehole functionality and surveys for a number of users could also be conducted to help benchmark the Project Technology Day parameter and supplement the sparse literature on borehole operating hours. Finally, they could allow the investigation to incorporate the perspectives of stakeholders in the project implementation chain as well as objective observers, community members, and user households.

2. Comparison to other schemes

Another investigative avenue that could be pursued would be to compare the GS SWS ER parameter values with those of similar projects registered under other climate impact schemes, such as VCS/Verra and Clean Development Mechanism. This additional step would expand the universe of possible comparisons, especially in areas where the literature is less complete.

3. Investigation of the additionality of community boreholes

The rehabilitation of community boreholes is at the core of many SWS projects. Project design documents frequently suggest that under-resourced communities lack funding to support the upkeep of these facilities, leading to their dilapidated state. Water experts consulted for this project noted, however, that there is value in conducting more nuanced investigations of why community boreholes are abandoned or allowed to deteriorate. In some cases, these boreholes may not be the sole community safe water supply source. This investigation did not have the bandwidth to delve into the GS due diligence process for accessing the additionality of the boreholes in the SWS portfolio, suggesting additional review of this topic could be valuable.

4. Systematic review of oversight processes

Within the scope of this review, we also did not carry out a systematic assessment of the certification and audit processes and compliance documentation. We were only able to speak to a handful of the personnel involved in reviewing or auditing SWS projects, leaving the potential for further interviews to uncover other insights and recommendations for improving the auditing system. Further study and recommendations from an expert in process improvement could also be very useful to improving the efficiency and consistency of the system.

VII. References

Aprovecho Research Center. 2007. Laboratory Testing of Rocket Stoves of Various Capacities As Compared to the Three Stone Fire.

Arora P, Jain S. 2016. A review of chronological development in cookstove assessment methods: Challenges and way forward. Renewable and Sustainable Energy Reviews 55:203–220; doi:10.1016/j.rser.2015.10.142.

Bailis R, Wang Y, Drigo R, Ghilardi A, Masera O. 2017. Getting the numbers right: revisiting woodfuel sustainability in the developing world. Environ Res Lett 12:115002; doi:10.1088/1748-9326/aa83ed.

Barbieri J, Parigi F, Riva F, Colombo E. 2018. Laboratory Testing of the Innovative Low-Cost Mewar Angithi Insert for Improving Energy Efficiency of Cooking Tasks on Three-Stone Fires in Critical Contexts. Energies 11:3463; doi:10.3390/en11123463.

Berkeley Air. 2012. Stove Performance Inventory Report.

Berrueta VM, Edwards RD, Masera OR. 2008. Energy performance of wood-burning cookstoves in Michoacan, Mexico. Renewable Energy 33:859–870; doi:10.1016/j.renene.2007.04.016.

Bilsback KR, Eilenberg SR, Good N, Heck L, Johnson M, Kodros JK, et al. 2018. The Firepower Sweep Test: A novel approach to cookstove laboratory testing. Indoor Air 28:936–949; doi:10.1111/ina.12497.

Cassivi A, Tilley E, Waygood EOD, Dorea C. 2021. Evaluating self-reported measures and alternatives to monitor access to drinking water: A case study in Malawi. Science of The Total Environment 750:141516; doi:10.1016/j.scitotenv.2020.141516.

Chen Y, Shen G, Su S, Du W, Huangfu Y, Liu G, et al. 2016. Efficiencies and pollutant emissions from forced-draft biomass-pellet semi-gasifier stoves: Comparison of International and Chinese water boiling test protocols. Energy for Sustainable Development 32:22–30; doi:10.1016/j.esd.2016.02.008.

Collivignarelli C, Mazzu A, Vaccari M, Vitali F. 2010. Comparison of biomass stoves appropriate technologies for household energy in developing countries. Third International Symposium on Energy from Biomass and Waste.

ERPF K. 2007. Installation and Maintenance Manual for the Afridev Handpump.

Fisher MB, Shields KF, Chan TU, Christenson E, Cronk RD, Leker H, et al. 2015. Understanding handpump sustainability: Determinants of rural water source functionality in the Greater Afram Plains region of Ghana. Water Resources Research 51:8431–8449; doi:10.1002/2014WR016770.

Foster T, Furey S, Banks B, Willetts J. 2019. Functionality of handpump water supplies: a review of data from sub-Saharan Africa and the Asia-Pacific region. International

Journal of Water Resources Development 0:1–15; doi:10.1080/07900627.2018.1543117.

Foster T, Willetts J, Lane M, Thomson P, Katuva J, Hope R. 2018. Risk factors associated with rural water supply failure: A 30-year retrospective study of handpumps on the south coast of Kenya. Science of The Total Environment 626:156–164; doi:10.1016/j.scitotenv.2017.12.302.

Grimsby LK, Rajabu HM, Treiber MU. 2016. Multiple biomass fuels and improved cook stoves from Tanzania assessed with the Water Boiling Test. Sustainable Energy Technologies and Assessments 14:63–73; doi:10.1016/j.seta.2016.01.004.

Harvey A, Mukanga J. 2020. Rural water service delivery: addressing the hidden development crisis. Waterlines 39:180–218; doi:10.3362/1756-3488.19-00004.

Holm R, Kamangira A, Kasulo V, Kaponda P, Hara E, Carney-Filmore C, et al. 2017. The handpump choice is yours: a pilot study in Rumphi District, Malawi. Waterlines 36:358–366; doi:10.3362/1756-3488.17-00006.

Howard G, Bartram J. 2003. Domestic Water Quantity, Service Level and Health. 39.

ISO. 2018. Standard 19867-1: Clean cookstoves and clean cooking solutions -- Harmonized laboratory test protocols -- Part 1: Standard test sequence for emissions and performance, safety and durability.

ISO Secretariat, Swedish Standards Institute, Associação Brasileira de Normas Técnicas. 2017. ISO 26000 and OECD Guidelines: Practical overview of the linkages.

James Robinson, Crispin Pemberton-Pigott, Tafadzwa Makonese, Harold Annegarn. 2010. Heterogenous Stove Testing Protocols for Emission and Thermal Performance.

Jetter J, Zhao Y, Smith KR, Khan B, Yelverton T, DeCarlo P, et al. 2012. Pollutant Emissions and Energy Efficiency under Controlled Conditions for Household Biomass Cookstoves and Implications for Metrics Useful in Setting International Test Standards. Environ Sci Technol 46:10827–10834; doi:10.1021/es301693f.

Jetter JJ, Kariher P. 2009. Solid-fuel household cook stoves: Characterization of performance and emissions. Biomass and Bioenergy 33:294–305; doi:10.1016/j.biombioe.2008.05.014.

Johnson M, Edwards R, Alatorre Frenk C, Masera O. 2008. In-field greenhouse gas emissions from cookstoves in rural Mexican households. Atmospheric Environment 42: 1206–1222.

Johnson M, Edwards R, Berrueta V, Masera O. 2010. New Approaches to Performance Testing of Improved Cookstoves. Environmental Science & Technology 44:368–374; doi:10.1021/es9013294.

Just B, Rogak S, Kandlikar M. 2013. Characterization of Ultrafine Particulate Matter from Traditional and Improved Biomass Cookstoves. Environ Sci Technol 47:3506–3512; doi:10.1021/es304351p.

Kelly E, Shields KF, Cronk R, Lee K, Behnke N, Klug T, et al. 2018. Seasonality, water use and community management of water systems in rural settings: Qualitative evidence from Ghana, Kenya, and Zambia. Sci Total Environ 628–629:715–721; doi:10.1016/j.scitotenv.2018.02.045.

MacCarty N, Ogle D, Still D, Bond T, Roden C. 2008. A laboratory comparison of the global warming impact of five major types of biomass cooking stoves. Energy for Sustainable Development 12: 56–65.

MacCarty N, Still D, Ogle D. 2010. Fuel use and emissions performance of fifty cooking stoves in the laboratory and related benchmarks of performance. Energy for Sustainable Development 14: 161–171.

MacDonald AM, Bonsor HC, Dochartaigh BÉÓ, Taylor RG. 2012. Quantitative maps of groundwater resources in Africa. Environ Res Lett 7:024009; doi:10.1088/1748-9326/7/2/024009.

McNicholl D, Hope R, Money A, Lane A, Armstrong A, van der Wilk N, et al. 2019. Performance-Based Funding for Reliable Rural Water Services in Africa.

Michaelowa A, Petersen T, Feige S, Galante A, Thorne S, Pedersen M. 2014. Suppressed Demand Climate Policy Brief.

Nagel C, Beach J, Iribagiza C, Thomas EA. 2015. Evaluating Cellular Instrumentation on Rural Handpumps to Improve Service Delivery—A Longitudinal Study in Rural Rwanda. Environ Sci Technol 49:14292–14300; doi:10.1021/acs.est.5b04077.

Owor M, MacDonald AM, Bonsor HC, Okullo J, Katusiime F, Alupo G, et al. 2017. UPGro Hidden Crisis Research Consortium. Survey 1 Country Report, Uganda. Available: http://nora.nerc.ac.uk/id/eprint/518403/ [accessed 2 August 2020].

Pearson AL. 2016. Comparison of methods to estimate water access: a pilot study of a GPS-based approach in low resource settings. Int J Health Geogr 15:33; doi:10.1186/s12942-016-0062-8.

Prasad KK, Sangen E, Visser P. 1985. Woodburning Cookstoves. In: *Advances in Heat Transfer* (J.P. Hartnett and Jr. Thomas F. Irvine, eds). Academic Press:Orlando, FL. 159–317.

Quist CM, Jones MR, Lewis RS. 2020. Influence of variability in testing parameters on cookstove performance metrics based on the water boiling test. Energy for Sustainable Development 58:112–118; doi:10.1016/j.esd.2020.07.006.

Rapp VH, Caubel JJ, Wilson DL, Gadgil AJ. 2016. Reducing Ultrafine Particle Emissions Using Air Injection in Wood-Burning Cookstoves. Environ Sci Technol 50:8368–8374; doi:10.1021/acs.est.6b01333.

Rural Water Supply Network. 2007a. Afridev.

Rural Water Supply Network. 2007b. India Mark II.

SDSN TReNDS. 2018. Handpump Data Improves Water Access.

Shen G, Hays MD, Smith KR, Williams C, Faircloth JW, Jetter JJ. 2018. Evaluating the Performance of Household Liquefied Petroleum Gas Cookstoves. Environ Sci Technol 52:904–915; doi:10.1021/acs.est.7b05155.

Smith KR, Uma R, Kishore VVN, Lata K, Joshi V, Zhang J, et al. 2000. Greenhouse gases from small-scale combustion devices in developing countries.

Still D, Bentson S, Li H. 2015. Results of Laboratory Testing of 15 Cookstove Designs in Accordance with the ISO/IWA Tiers of Performance. EcoHealth 12:12–24; doi:10.1007/s10393-014-0955-6.

Teune B, Ha HT, Salinas D, McLean K, Bailis R. 2020. Low-cost interventions to reduce emissions and fuel consumption in open wood fires in rural communities. Energy for Sustainable Development 58:119–128; doi:10.1016/j.esd.2020.07.007.

Thomas EA, Needoba J, Kaberia D, Butterworth J, Adams EC, Oduor P, et al. 2019. Quantifying increased groundwater demand from prolonged drought in the East African Rift Valley. Science of The Total Environment 666:1265–1272; doi:10.1016/j.scitotenv.2019.02.206.

Thomson P, Bradley D, Katilu A, Katuva J, Lanzoni M, Koehler J, et al. 2019. Rainfall and groundwater use in rural Kenya. Science of The Total Environment 649:722–730; doi:10.1016/j.scitotenv.2018.08.330.

Truslove JP, V. M. Miller A, Mannix N, Nhlema M, Rivett MO, Coulson AB, et al. 2019. Understanding the Functionality and Burden on Decentralised Rural Water Supply: Influence of Millennium Development Goal 7c Coverage Targets. Water 11:494; doi:10.3390/w11030494.

USAID and the Academy for Educational Development. 2007. Fuel Efficient Stove Programs in IDP settings – Summary Evaluation Report, Uganda.

Wright J, Gundry S, Conroy R. 2004. Household drinking water in developing countries: a systematic review of microbiological contamination between source and point-of-use. Tropical Medicine & International Health 9:106–117; doi:10.1046/j.1365-3156.2003.01160.x.

VIII. Appendix 1: Terms of Reference

Appended here is version 4.1 of the terms of reference for the independent investigation of the grievance, "Potential over-issuance risk from Safe Water Supply projects". The document can be accessed publicly at this site.

1.0 Introduction & Purpose

Gold Standard responded to a request for an investigation into observations[23] (table below) of potential erroneous estimation and issuances in Safe Water Supply projects (i.e. projects applying the 'Technologies and Practices for the Displacement of Decentralised Thermal Energy Consumption – TPDDTEC – Methodology). The concerns were raised to the Gold Standard Technical Advisory Committee (GS-TAC) Working Group (WG) by the Gold Standard Secretariat following analysis of the project pipeline and portfolio. The research and due diligence formed part of regular quality assurance and review conducted by Gold Standard, towards improving the scheme and informing planned updates of the TPDDTEC methodology. The research was not initiated due to any specific issue or concern.

A formal grievance was then submitted by the GS-TAC WG It is noted that:

- *All grievances are treated seriously and as potentially significant, but veracity, accuracy and responsibility are not presumed.*
- *Grievances may be subject to independent investigation where required.*
- *The issuance appears, in the initial assessment of GS Secretariat and GS-TAC, to be potentially significantly higher than would be typically expected. This, however, is not presumed and must be fairly reviewed and ascertained.*
- *The cause(s) of over-estimation (if proven correct) are not presumed, and the grievance does not represent an allegation of non-conformity against any particular project or project-developer or any stakeholder involved in the processes of standards-setting or certification. In the event a non-conformity is observed, action may be taken under the remit of this investigation.*
- *The grievance should be managed independently of Gold Standard and SustainCERT to ensure impartiality of findings and fairness of any action taken.*

This document represents the output of Step 4 of the published Gold Standard Grievance Process[24], namely the TORs for the investigation. The TORs represent a framework through which Gold Standard will commission an independent investigation, overseen by a GS-TAC WG for grievances, and published to the Gold Standard grievance page. In this case the TOR was first presented to the GS-TAC to review and revise as they, independent of the GS Secretariat, see fit. The TORs may further alter at the review of independent investigators later involved, to avoid leading or unduly influencing the investigation.

2.0 Summary of the Allegations

[23] For details of the grievance, visit GS website: https://www.goldstandard.org/our-story/grievances-deregistration

[24] https://www.goldstandard.org/our-story/grievances-deregistration

The allegations have been summarised in Table 1. It is reiterated that the Gold Standard nor the GS-TAC or GS-TAC WG have reached a position on either veracity or accuracy of these allegations, pending the independent investigation.

Table 1 – Summary of allegations

Allegations	**Summary**
1 – Application of the TPDDTEC Methodology for SWS may lead to over-issuance of VERs	A review of the portfolio of projects applying the TPDDTEC (V1.0 to 3.1) SWS approach was produced as part of ongoing QA/QC and for background research to support planned methodological updates. In doing so, Gold Standard Secretariat considered that the projected issuances may exceed what may typically be expected by experts/literature and hence it was self-reported to GS-TAC. GS-TAC considered this concern and issued a grievance allegation to Gold Standard Secretariat. It was noted that the grievance does not assume fault nor responsibility of any particular party and should consider both the potential over-issuance (including verifying that indeed the projections represent a valid concern) as well as its underlying causes and potential rectification.
2 – The GS4GG scheme (and previous versions) may not have, prior to this review, dealt with over-estimation issues at the optimal point in the certification process or in an optimal timeframe.	While it is encouraging that the QA/QC process run by Gold Standard on project design/performance certifications detected a potential issue, it is noted that such matters should ideally be picked up earlier in the certification and regular review processes to ensure fairness, predictability and stability for all parties. Hence it is alleged that within the GS scheme there is a failure to prevent potential over-issuance. It is not pre-identified where this may be absent or how it is best resolved for future projects.

3.0 Grievance process

Step 1 (COMPLETED): Receive and review the grievance: Conduct an initial review assessment of whether the issues raised may lead to over-issuance, noting that amount, cause and rectification are not presumed.

OUTPUT: An initial review conducted by GS-TAC lead to the conclusion that:

1. *An initial freezing of issuances followed by a period of partial issuance (decided by and determined in discussion with GS-TAC) should be initiated until an independent investigation is conducted and recommendations implemented.*
2. *An independent investigation should be conducted and recommendations implemented.*

Steps 2-4:

Step 2 – Notify parties, publish grievance on website and flag in the registry – Grievance is published on Gold Standard website. This investigation was not raised against any specific project and hence a flag will be added to all SWS projects linking to the

grievance page if feasible, or alternatively if this is not possible then no registry account will be flagged. Instead the investigation TORs and briefing will be published to the Gold Standard grievances page and issuance will be limited to a partial amount (determined in discussion with GS-TAC) until the investigation is complete.

Step 3 (appoint grievance team) – this TOR will provide the briefing to appoint an independent investigation team.

Step 4 (draft TORs for the investigation) – COMPLETED with this document.

OUTPUT: An appointed grievance team and completed TOR.

Step 5 (Solicit additional feedback and desk review):

These steps will run concurrently, allowing the investigation team to review and conclude any issues that are dealt with in certified documentation and focusing the gathering of further evidence on remaining open items. This will therefore involve as follows and will inform both Parts A and B of the investigation:

Shortlisting the key issues which will be further investigated due to potential risk of non-conformity against the Standard or breach of any associated documents such as the Terms and Conditions.

Reviewing standards, certification and project documents including internal reviews, validation reports, monitoring reports and verification reports completed by auditors. GS-TAC WG Members may elect to join any such interviews at their discretion.

Interviewing relevant stakeholders as identified by the investigation team, including but not limited to Gold Standard Secretariat, SustainCERT, project developers, auditors, TAC members, experts and other stakeholders as required. In addition, an open opportunity to comment/make representation is to be added to the grievance page.

***If required**, a field study including but not limited to fact gathering, assessing legal findings, interviewing affected stakeholders and others to support the review work. This will be at the discretion of the independent investigator and GS-TAC WG who should assess based on whether a field visit(s) is the only viable option to determine the issues under investigation. It is acknowledged that potential for field visits may be impacted by COVID19 and hence other options should be explored as priority. This decision should be taken as early as possible within the investigation to mitigate delay.*

Within the above analysis, to what extent the information is sufficiently complete in order to make recommendation(s). Previous steps may be revisited as needed. If appropriate, any findings regarding over-estimation &/or over-issuance may be implemented sooner than wider changes to the scheme to allow issuances to proceed unimpeded.

OUTPUT: A Draft Preliminary Report providing an analysis of desk-based information and fact gathered from field study (if required), making recommendations. To expedite the review findings, the first Draft Preliminary Report should respond to investigation Part A and so far as feasible Part B 4 and 5 (see below).

From the Preliminary Report, steps 6-8 will be progressed[25], including documenting findings of fact, recommendations and the reasoning underlying those recommendations in a report for review by the Technical Advisory Committee, External Expert, Project Owners/Developers (where relevant), SustainCERT and an executive committee of the Gold Standard Foundation Board of Directors.
The GS-TAC WG (or a nominated member) will support the executive committee if and as required by attending discussions.

4.0 Investigation overview, Scope, TORs and workplan:

Overview:

The investigation will be briefed by the GS-TAC WG(through these TORs, initially developed by the GS Secretariat and reviewed, amended and approved by GS-TAC and investigators) and then conducted by an independent investigator. The process of investigation and decision making will be overseen by the GS-TAC WG but the investigation will be fully independent. This is to avoid any conflict of interest (perceived or real) as the investigation allegation includes the Gold Standard.

Scope:

The investigation will seek to answer the following questions, based upon the principles of prevention of over-issuance, fair rectification and prevention of future occurrence.

The questions are in two parts:

Part A:
Analyse the technical aspects of the allegation to assess it's veracity.
If the investigation finds no material basis in the allegation, the investigation will not proceed further.

If the investigation finds basis in the allegation, that is, that material[26] over-estimation has occurred, Part B will be pursued.

Part B:
Assess and provide analysis and recommendations on response(s) to the findings of part A.

Part A:

6. *Within the assumptions of suppressed demand within the methodology(ies), is the perceived over-estimation (and thus over-issuance) within the SWS portfolio correctly identified (that is, is the perceived issue 'real' – has over-estimation &/or over-issuance occurred), noting that over-estimation and over-issuance are related but not the same topics (for example the potential resolution for over-issuance may not also*

[25] See: https://www.goldstandard.org/our-story/grievances-deregistration
[26] Material defined as: sufficiently different from the stated value that it would likely effect the decision making of the information-user. It is suggested that the independent investigator recommend a fairly stated value with rationale for context/basis of decision making.

resolve over-estimate, which may require other standards, methodological and procedural change)?

   a. *If not, investigation ends with recommendations on analysis approaches and QA/QC process.*

7. *If so, to what degree? That is, what is the likely amount of over-issuance per credit issued? Ideally this can be quantified (as 'x%'), but may require approximations (eg: quantitative ranges 'up to x%' or 'between y and x%') and/or qualitative assessments (eg: 'insignificantly, somewhat, significant, substantially' etc).*
8. *Does the degree of difference represent a material over-estimation &/or over-issuance?*
   a. *If not, investigation ends with recommendations on analysis approaches and QA/QC process.*

Part B:

9. *If material difference is identified, where in the GS process of standards-setting, methodology development and updating, project documentation establishment, project monitoring and/or validation/verification/certification or any other external, internal or combination of factors do the cause(s) of over-estimation &/or over-issuance lie?*
10. *Where cause(s) can be identified, what degree of responsibility (if any) can be ascribed to the various parties and stakeholders associated with over-estimation &/or over-issuance? Note that any ascribed responsibility does not imply culpability or intent, rather it aims to identify where error(s) may have occurred &/or how they interact &/or how they are propagated. It may also identify where processes could be adjusted to prevent similar results in the future.*
11. *In light of the above, how should this be remedied, for example should:*
    a. *any action be taken on credits already issued (including those transferred, sold or retired)?*
    b. *issuances from projects already in pipeline (Listed, Under Validation, Under Design Certification, Design Certified (registered) or Issued) be limited or amended in some way to ensure credibility?*
    c. *Alternative suggested remedies as put forward by the independent investigator/TAC?*

    If so, what actions, degree of limitation or any other proposed interventions would ensure credible issuances?
12. *How, where, and in what ways can Gold Standard improve it's processes and products to prevent recurrence of such issues? Is this part of BAG ToR?*
    a. *What are high priority actions (that is, those that are necessary and near-term) to ensure credibility?*
    b. *What are other actions that will contribute to ongoing improvements?*

Neither the veracity of the claim, nor identification of possible areas of responsibility for the potential over-issuance has not been determined. To ensure expedience of questions 1-6, question 7 which relates to the second allegation, may take place over a longer timeframe, though aspects of Part B may be required and naturally be investigated in parallel to Part A.

It is intended that, in order to expedite the review to minimize delay for projects awaiting issuance, that the independent investigator should focus on Part A and make initial recommendations for Part B4/5 as part of Draft Preliminary Report (see timeline below)

Terms of Reference:

The investigation will focus on the grievances raised and assessed against the relevant and applicable standards requirements for the project. The applicable requirements are as follows:

*Standard version: Dependent on project Registration/Transition dates - Gold Standard for the Global Goals and previous versions of the standard**

Methodology/version: Technologies and Practices for the Displacement of Decentralised Thermal Energy Consumption (TPDDTEC) Version 1, 2, 3 and 3.1

Terms and Conditions and Terms of Use

Gold Standard core principles of fairness, reliability, conservativeness and pragmatism

Gold Standard for the Global Goals Principles & requirements

- *GS4GG GHG emissions reductions & sequestration product requirements*
- Gold Standard Principles (available here)
- *The Gold Standard Requirements*
- *Requirements as per relevant project documents including but not limited to project design documents, monitoring reports, validation & verification reports, internal reviews and analysis, recognizing that a representative sample can be selected to inform analysis.*

**Note that previous versions of Standard is now superseded by Gold Standard for the Global Goals (GS4GG).*

The GS-TAC WG will provide oversight and peer review to the investigation and infer and recommend decisions for the correct courses of action where required.

The GS-Secretariat will independently administer the process, overseen by the GS-TAC WG. GS-Secretariat will remain neutral in its administration role, limiting any views or considerations to where formally requested by either GS-TAC WG and/or the independent investigator to make formal representation. The GS-Secretariat will pro-actively support both GS-TAC WG and independent investigator in collating information as required.

The investigator(s) will first conduct a desk-based review of the summary produced by Gold Standard (noting that this should disaggregate where necessary by version of methodology applied as these may have different relevant clauses). The investigator will further review all relevant documentation, including but not limited to the Gold Standard Requirements and Toolkits, overview of project portfolio and issuance projections, project design documents, passports, monitoring reports, validation and verification reports, along with interviews with stakeholders. This will be supplemented, if necessary, by further field study to complete the investigation. This will include any agreed changes, exceptions, deviations or approaches approved by Gold Standard during certification

The proposed investigation plan will be subject to change as investigation progresses and new information emerges. Changes to the plan presented above will be documented and public updates provided.

The investigator(s) will review the following project documents, or a sufficiently representative sample, (not exhaustive):

- *Standards and methodological requirements*

- *Registry information*
- *Certification processes*
- *Project documentation*
- *Validation/verification and certification documentation*
- *Relevant literature related to impact of SWS projects*
- *Field reports*

Based on the above desk-based assessment the investigation team will further:

- *Interview the Gold Standard, SustainCERT, Validating and Verifying DOEs, Project Developers, stakeholders and experts (as well as any other pertinent stakeholders) to establish any discrepancies/gaps and any conflicts of interest that may have arisen within any aspect of the project monitoring and reporting, standard or methodology, assurance process and certification review and decision making.*
- *GS-TAC WG Members may be requested or themselves request to be involved in any interview.*
- *Consider comparative literature, studies and experiences relating to the reasonableness of estimations.*
- *Conduct an on-site assessment via a third-party expert(s), if/as required.*
- *Review findings and decisions.*

The investigation will follow the approximate timeline shown below. This timeline is subject to availability of independent investigators and the findings of the investigation and may change. Changes will be documented and made publicly available. Urgent elements of investigation (i.e. those that affect the potential issuance of existing projects) are anticipated to be completed within 8 weeks from commencement.

All dates below are target/indicative and may change during the course of the investigation.

Activity	Timeline – to be completed with independent expert.
Initial Review/non-Conformity Assessment	Completed
Grievance Investigation Terms of Reference (TORs) (This document)	14th July 2020
Appointment of investigation team	20th July 2020
Appointment of independent local expert(s)*	W/c 20th July 2020
Evidence collection, desk review and stakeholder interviews (records of any sampled documents, interviews etc should be declared)	by 14th August 2020
Draft report by independent investigator	By 21st August 2020
Preliminary Review/ Peer Review (TAC, External expert)	w/c 24th August 2020
Review by project developer(s) (an opportunity for comment only, it is at Gold Standard TAC's discretion to adjust the report)	w/c 31st August 2020
Review by GS Board	w/c 31st August 2002
Final report and publication on GS website	w/c 07th September

* Local expertise may be required, potentially provided by more than one individual.

5.0 Transparency and declaration of conflict of interest

Gold Standard, as part of its commitment to transparency and good governance will publish all documentation to the Gold Standard website. In addition, Gold Standard will specifically notify these key stakeholders and refer them to the publication page and investigation team as necessary:

- *ISEAL*
- *WWF (Switzerland)*
- *ICROA*
- *Carbon Market Watch*
- *All affected project developers*

In addition, TAC will review/monitor any Conflicts of Interest arising amongst GS Secretariat, SustainCERT and investigation team.
The GS-TAC WG was recommended for appointment by Scott Harder, an independent advisor to the Gold Standard Board (who has no connection or conflict to/with the TPDDTEC methodology, SWS projects or the GS-TAC). The members of the group are below, recommended based on whether they have relevant expertise and do not have any conflicts of interest. These members were informed to all GS-TAC members (Energy Sub-Committee) to raise any concerns prior to confirming, none was raised. The GS-TAC group will subsequently be chaired by Scott Harder.

The Members appointed are:

Scott Harder (Chair)
Steve Thorne
Jessica Wade-Murphy
Liza Murphy

6.0 Investigation Team

As noted a bove, the investigation will be overseen by the GS-TAC WG who are:
Expert in the TPPDTEC methodology and SWS projects
Have no conflict of interest

The investigation will be carried out by an independent expert appointed directly by Gold Standard via a GS-TAC WG approved scope and workplan based on this TOR (which the investigator will also confirm). The contract between GS and the investigator will be purely for the carrying out the investigation under the TOR and the contracting shall be observed by GS-TAC WG Members to ensure impartiality of terms.

The expert(s) shall:
Be fully independent of GS, SC and all project developers currently applying the TPPDTEC methodology and shall have no other conflict of interest
Be preferably either an independent operator or employed at a public or non-profit institution. Private company employment will be considered only if necessary.
Be an expert in the application of SWS projects and the calculation of emissions reductions from them, including the individual parameters applied
Have a minimum of 7 years senior experience at a public institution or similar
Have relevant post-graduate qualifications.

IX. Appendix 2: Complete list of interviews

Name	Affiliation	Role	Date
Evan Thomas	University of Colorado, Boulder	Expert in safe water solutions	08/14/2020
Jim Jetter Christian L'Orange	US EPA Colorado State University	Experts in testing improved household energy technologies	08/14/2020
Lisa Rosen	SustainCERT	CEO	08/14/2020
Richard Iliffe	SustainCERT	Compliance Director	08/19/2020
Lucas Emmerson Huw Jones James Walker Paul Chiplen	CO2Balance	Head of Project Developer Financial Director Regional Manager—E. Africa Sales & Marketing Director	08/20/2020
Frederico Gallo	BelieveGreen	Founder and Managing Director	08/20/2020
Owen Hewlett Vikash Talyan Abhishek Goyal	Gold Standard	Members of Gold Standard Secretariat	08/20/2020
Sean Furey	Rural Water Supply Network	Secretariat Director	08/24/2020
Matt Spannagle	Independent	Expert in climate change solutions Chair of GS TAC, chair of GS Energy TAC subcommittee, and member of Board of Directors of GS Foundation	08/27/2020
Huw Jones James Walker Emma Donnachie	CO2Balance	Financial Director Regional Manager—E. Africa Regional Manager—S. Africa	08/28/2020
Sanjay Kandari	KBS Certification Services	Head of Climate Change Services	08/30/2020
Michael Novoszad Manfred Stockmayer	Likano	Executive Director Carbon Asset Manager	09/03/2020
Ha Thanh Hoang	SustainCERT	External reviewer	09/04/2020
Jamie Bertram	University of North Carolina, University of Leeds	Expert in safe water solutions	09/08/2020
Anurag Juyal	SustainCERT	Senior Certification Officer	09/08/2020

X. Appendix 3: Chronology of Berkeley Air Monitoring Group's involvement with The Gold Standard

2007-08

Evan Haigler and David Pennise at The Center for Entrepreneurship in International Health and Development (CEIHD) in the University of California-Berkeley School of Public Health (the predecessor of Berkeley Air) contributed as pro-bono technical advisors to the monitoring methodology for the first "improved cookstoves" methodology ("Methodology for Improved Cook-stoves and Kitchen Regimes V.01," dated 30-May-2008). In parallel CEIHD was funded by the Partnership for Clean Indoor Air to provide technical consulting to Urban Community Development Association of Kampala, Uganda, which was an initial project developed by Climate Care under this methodology.

2010-11

Michel Johnson, David Pennise, and Dana Charron (Berkeley Air) and Seth Shonkoff (University of California-Berkeley and Berkeley Air) contributed as pro-bono technical advisors to the next version of the "cookstoves" methodology, particularly regarding the cookstove monitoring methodology, sample sizing, and statistical analysis ("Technologies and Practices to Displace Decentralized Thermal Energy Consumption V1.0," dated 11-April-2011)

2013

Berkeley Air was contracted by GSF to perform a review of the usage methods of Vestergaard Frandsen's GS water filter project in Kenya. We produced a report titled, "Assessment of Usage Methods of GS886 Sustainable Deployment of the LifeStraw Family in Rural Kenya," dated Nov 2013 and authored by Jay Graham of George Washington University and Maneet Kaur and David Pennise of Berkeley Air.

2015-17

David Pennise (Berkeley Air) contributed as a pro-bono technical advisor to the development of the GS ADALY methodology, titled "Methodology to Estimate and Verify Averted Mortality and Disability Adjusted Life Years (ADALYs) from Cleaner Household Air, Version 1.0," dated 17-Jan-2017.

Berkeley Air was contracted by the GS TAC to review the stove use survey assessment methods and data analysis of the project "Improved Kitchen Regimes Multi-Country PoA (Project ID GS1247)" by CO2balance and to provide recommendations to improve the monitoring of stove usage in GS cookstove projects. We produced two reports titled "Evaluation and Development of Usage Survey Guidelines for Improved Cook Stove Activities, Final Report 1: Project Evaluation," dated Feb 2017, and "Final Report 2: Cookstove Usage Survey Guidelines," dated Mar 2017.

Michael Johnson (Berkeley Air) contributed as a pro-bono technical advisor to the development of the GS black carbon methodology, titled “Quantification of climate related emission reductions of Black Carbon and Co-emitted Species due to the replacement of less efficient cookstoves with improved efficiency cookstoves, Version 1.1,” dated Aug 2017.

I. Process and Procedure

A. Scope and Rationale

The SWS Grievance WG (SWSGWG) has chosen to write this section in addition to the Berkeley Air (BA) report to present its perspective on the root causes of the Grievance, the processes that affected the performance of GS, the overall application of the Assurance Systems, and the application of those processes and systems by SustainCERT (SC).

In consultation and agreement with the SWSGWG, BA focused its efforts substantially on Part A technical investigations. The complexity of the matter, large volume of datafiles, and tight time constraints led to this decision. The GS Secretariat was apprised of this decision. In keeping with the TOR, Berkeley Air has also presented observations on the causes, and responsibility for those causes, based on evidence collected while primarily investigating the research questions in Part A.

The SWSGWG Report on Process and Procedure is separate from the BA work and intended to be complementary and broader in scope as to its conclusions and observations. The TOR for the SWSGWG in section B, points 5 and 7 calls for a discussion of causes and recommendations for improvements and this section focuses on those elements. Together SWSGWG members represent over 30 years of engagement with GS and as well as with other standard-setting systems.

We also wish to note that in some cases we have extrapolated our conclusions from reviews, interviews and research conducted. The SWSGWG received comments relative to SWS processes and procedures from SWS stakeholders. These comments provided additional information intending to clarify and elaborate various process and procedural elements of the SWS scheme. In some cases, the comments raised conflicting points of view with what the WG gathered during the interview process. Please refer to the Stakeholder Comments Annex attached to this report for further information.

B. Root Cause Analysis

1. Introduction

As discussed in the BA report and in the Executive Summary the investigation of the grievance was a significantly more complex matter than originally anticipated. To meet the requests of an expedited investigation that would allow the full TAC, GS Board, and the GS management team to develop a robust response to bring the matter to a resolution, we focused the majority of effort on the Technical Review at the root of the matter. During that work it became apparent that the causes of the issue were not limited to flaws in the application of the methodology, but were driven also by inconsistencies in the application of the Assurance Systems and related flaws or breakdowns in the processes that were intended to maximize consistent and credible certifications.

2. Focus Areas

We have grouped our discussion points into the following areas of focus in response to the results of the reviews.

- The Standard and how it is maintained

- The Assurance Model and its intent
- Application of the Assurance System by SC
- Oversight by the GS Secretariat
    - Of SustainCERT
    - Of Project Developers
- Technical parameters used in emissions reduction estimates

3. The Standard and How it is Maintained

It is vital to keep the methodologies and all other components of the Standard fully and robustly current. One issue identified in the TPDDTEC methodology was its lack of alignment with the currently available research and literature. While the TPDDTEC methodology has been updated since it was first developed, it appears that those updates may not have been as thorough as was optimal.

We would suggest that updates to methodologies, triggered for any reason, always include an assessment of emerging best practices for determination of baseline scenarios and parameters, as well as project scenarios and parameters. For example, a check of the recent literature by the GS Secretariat could contribute to such an assessment. There have also been recent additions to ISO Standards that address cookstoves and solid fuels that should be considered as well.We would recommend a regular scheduling of reviews of all methodologies using original authors as well as those with new perspectives. Complementary would be the development in advance of a checklist of items to consider and document for any update being reviewed.

Various methodologies will likely be expected to need review on different timescales as technology progresses. Off-cycle reviews would still be appropriate as issues arise, as under the existing system.

Another issue identified is how updated versions of methodologies are applied to projects and Programmes of Activities (PoAs) with approved Design Certification. As per current GS rules, once a PoA design certification is approved, there is no requirement for a retroactive application of a methodology update, even for VPAs not yet included, until the renewal of the crediting period of the PoA. We suggest reconsidering this requirement. Typically a Standard Setting System develops a protocol that indicates the first date a change may be adopted (and verified) by existing participants and a second date of when the change must be adopted by all participants as well as a date that all new certifications must follow the new requirements. This is especially relevant for our work in the context of large PoAs.

Additionally, considering the often complex nature of the methodology updates and the long time frames for review, development and approval, we support development of a policy of issuance of formal guidance notes that can direct interim implementation by developers and certifiers. These guidance notes can be used as bridging strategies for complex updates as well as the primary method for minor adjustments that are off-cycle to updates to the Standard. Any Guidance Note issued should have a discussion of requirements for timing of uptake.

A further issue identified in the TPDDTEC methodology is the lack of detail in the requirements for some parameters. Clearer requirements and default or reference values

on the baselines parameters would improve its application. This could also have a follow on result of minimizing requirements for verification when the default or reference number was used. The number would be determined by using scientific literature, publications and other externally validated sources to avoid an appearance of conflict. Project Developers could still present their case for any proposed variance of a parameter, which would be able to be vetted in a more consistent way in the framework of the latest developments. A review of the SW quantities requires consideration of a more conservative application; perhaps reducing the amounts of water for hygiene and food preparation.

Another example would be that of an update that includes the necessary incentivising of the repair of project equipment. There could be application of baselines that default to an application of a low, default value of technology days if the operation of the equipment is not directly monitored, while permitting higher values if the operational days are measured directly.

As voiced by several stakeholders during the investigation, GS project development standards and procedures are generally complex, multi-layered, and often difficult to understand. This complexity does not aid the oversight and supervision of the application of the standards and procedures. Ensuring as much clarity in language used as is possible should be a priority in all publications, including the use of Technical Writers.

4. The Assurance Model and Its Intent

Overall GS has designed its Assurance model to have two foci, Small/Large Scale and Microscale projects. This has been a deliberate choice and one that supports the overall goals of the organization. As the organization has grown, uptake of the various types of projects has changed to reflect both the growth of the organization as well as the changes and evolution of the carbon markets. This should be seen as a marker of success. However, there have emerged in the last years some unanticipated applications of microscale projects that have had a role in the issues under investigation.

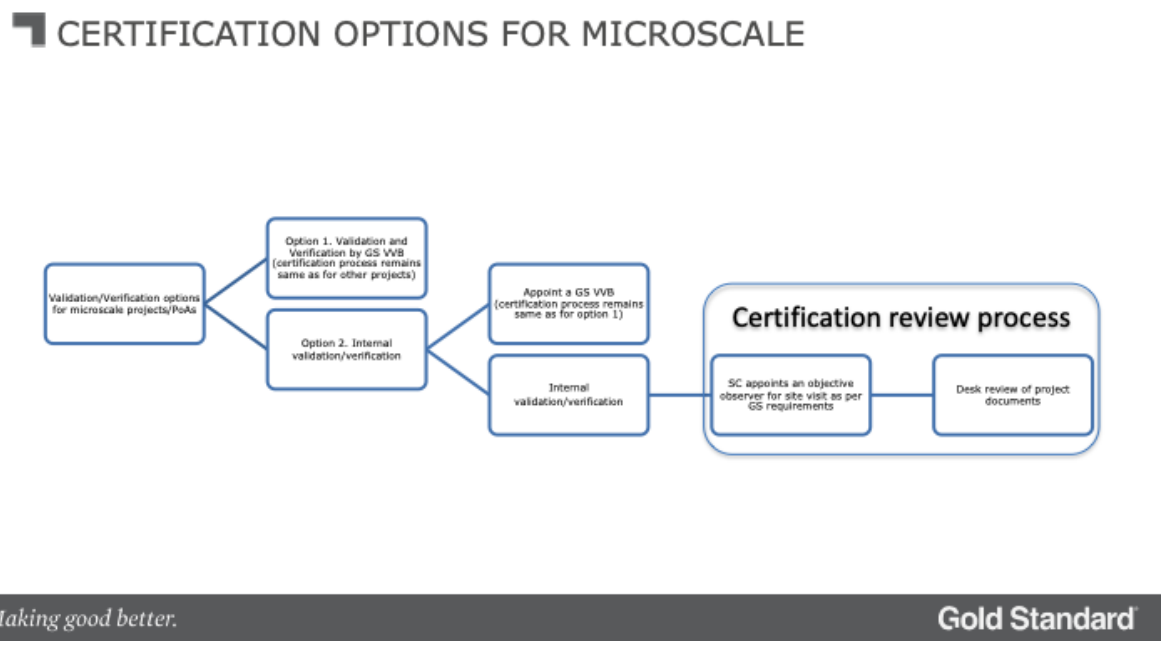
Non-Microscale This has not been a focus of this investigation and so we will not address our primary comments to these categories of projects and how they are assured. We will make the important note that the mandatory use of Assessors (VVBs/DOEs) who have been accredited by an outside body and trained to the GS Standard adds a layer of important risk management and institutionalized quality checks, and helps manage the inherent conflicts present in the use of SC, a wholly owned subsidiary of GS, as the single certifier for the system.

Microscale The Assurance model for the microscale approach recognizes that a barrier to entry for many small communities, who paradoxically would benefit the most from participation (at least on an incremental basis), has been the lack of resources to make the initial expenditures for project implementation and certification, thus freezing them out of benefits from emissions reductions and sales of resulting credits. In some cases small projects and communities may also have issues with access to the skill sets needed to design, manage and monitor programs that comply with the rules base of Emission Reduction programs.

To facilitate market access, a microscale model was developed that allows very small projects, with limited capacities, to be certified via a dedicated process. This process uses a

combination of Objective Observers, often local individuals who gather data about the application of the project and report it to SC, and detailed desk reviews by the certifier for certification decision making.

FIG. WG-1 SustainCERT "Internal" certification process flow for Microscale projects



Another important development in the application of the Microscale approach was the establishment of SC to, in part, more effectively address the Conflict of Interest inherent in the comingling of Standards Development, Accreditation and Assurance within the Standard Setting organization. This followed the requirements of ISEAL membership, an important credibility indicator in the world of Sustainability Standards.

The use of Objective Observers is one of the areas that Microscale projects differ in the Assurance process. Rather than using an externally accredited assessment providers (VVB/DOE), the gathering of data used for verification is managed by the Certifier (SC), incorporating - sometimes - an independent individual, an Objective Observer, who has no requirement for external accreditation. Earlier in the history of microscale projects, the Objective Observers tended to be individuals well known to GS and with deep familiarity with the GS processes of certification. As the GS succeeded and grew and markets evolved, several changes were integrated, either by design or via a gradual evolution of the Microscale approach. Objective Observers now are frequently suggested by the project developer and it is unclear how any potential Conflicts of Interest are identified or managed. Minimal training is provided to the Objective Observers, who appear over time to have become less familiar to those reliant on their inputs: SC and GS.

Given the expected users of the microscale approach, there were no limits put on the number of projects that could be bundled together and still be considered microscale, likely because it was not anticipated to have high levels of uptake. It is also important to note that at the time of design of the Microscale assurance model, and still today, it is welcome news

that many communities and small projects are receiving SDG focused benefits and a share of the revenues generated by sales of emission reduction credits.

Today's result of this design is that Project Developers were in some cases given what amounted to global approval at the project design stage. The individual microscale sub-projects (VPAs) embedded in the PoAs were subject to caps on credits issued, however the larger PoAs are not limited in size nor subject to caps on cumulative issuances. The Project Developers also had the ability to have a VPA included, and then clone it. So a bundled set of projects, administered by subcontractors, could result in a total issuance of very large numbers of credits that were issued without the control, or risk management, that a full verification process conducted by an accredited third-party verifier (VVB/DOE) would offer.

Some Project Developers seem to have developed an understanding of the possibilities of economy of scale this bundled and cloned approach could offer them, as well as an appreciation that the lack of upper issuance limits could likely present significant earnings opportunities for themselves (and to a lesser degree their subcontractors who were working regionally, and likely to an even lesser degree the communities participating in VPAs). In sum, the Project Developer appears to reap the largest benefit from the economies of scale in this approach and Gold Standard has heightened risk due to the application of the Assurance processes.

Another factor in this matter and its relationship to the Assurance Model is the ongoing reliance on a voluntary, unorganized (except loosely and unofficially) review of pending certification decisions by the various GS Advisory Groups and Stakeholders (E and LUF TACs and/or a finite list of NGO supporters). While a smaller organization, GS TAC members played a direct, required and partially compensated role in taking certification decisions. As the organization grew, SC was established and this mandatory role evolved. SC reports that they rarely if ever receive input on certifications presented for review from NGO Supporters, and that the input from TAC members constitutes a random review driven by the voluntary capacity at any given time by individual members and the interest they might have in a particular topic or country. As a result, this step in the Assurance Process is not being as robustly engaged as the model would imply and thus its effectiveness is greatly diminished.

Engagement by these entities (TACs and NGOs) does not seem to be monitored or reported by either SC or GS, and what is not measured often does not count. The result of the low participation is to obviate the intended effectiveness of this screening step. The role that the TAC has played in Certification has changed substantially, and the Assurance Model should be modified to address that. It seems unlikely that the QA check role that was once played by them is possible to maintain in this world of increased volumes and evolution of the role of the TAC in the GS design.

The outputs of the microscale approach as it stands today seem to have moved away from the original intent of providing access to small communities and projects that did not have representation to facilitate access to these markets. It is impossible at this point to gauge its full effectiveness, both on a relative and an absolute scale. This grievance investigation has called into question the delivery of both SDG benefits via accurate assessments of the delivery of, in this case, Safe Water, and the generation of credible emission reductions. To understand its effectiveness, there would need to be a full review of the benefits & revenues generated to each of those involved in the Microscale activities (Communities, Local Service Providers, Project Developers, SC and GS).

In light of the institutional arrangements between GS and SC, as well as the rapid growth and uptake of the microscale certification models, it would be urgent that the GS review its Assurance Model and its application to close the gaps that have emerged and are now identified.

Further, an assessment of the intended and actual impacts of the microscale approach is appropriate at this time. Development of a tighter definition of microscale would be an important starting point. We wish to stress that we strongly believe that rules designed specifically to support access to the positive impacts of carbon finance by those otherwise unable to access them is vital to the mission of the organization. We also note that this is an opportune time to review the intended impacts of microscale and the effectiveness in reaching them in light of the considerations of credible and viable outcomes.

5. Application of the Assurance System by SustainCERT

SC has been shown to have a number of individually significant and cumulatively large problems with the implementation of the Assurance System that GS has asked them to follow in their verification and certification activities.

Training For Reviewers - for those reviewers who are either staff or consultants, there does not seem to be evidence of formal training of either the contents of the Standard or the practices of SC. Training keeps staff up to date, it helps ensure consistent application of the standard and is a vital part of ongoing operations of a Certification Body.

For Objective Observers - from a discussion with SC staff, it appears that there is no formal training for these important roles. Rather, they are provided with a manual and their work is reviewed. The detail regarding their selection, monitoring or training to give consistent information is unclear. It is also unclear how any possible Conflicts of Interest are managed, a concern given the indication that Objective Observers are proposed by Project Developers. This is likely an aspect that needs some timely attention to gather more details, diagnose and improve. While we did not assess the training question regarding VVB/DOE focus we would also suggest this topic be considered as well.

We strongly recommend that SC develop a training program that can be documented and that is delivered in a consistent and way to Reviewers who are either staff or consultants, VVBs, DOEs and Objective Observers.

The Curriculum should include a full review of the Standard(s) they are responsible for validating, verifying and/or certifying, as well as internal SC procedures and protocols.

Internal Quality Review Processes As a Certifier, SC has a responsibility to maintain consistent outputs in their reviews and certification decisions. Further, they must ensure that as the Standard and its components evolve, those changes are applied appropriately. And, they must be proactively assessing Risk in their work and implementing actions to mitigate, track and manage it. It appears that the reviews that would support actively managing these goals may not be done in a regular way, if at all. Some examples we identified are:

Microscale projects may opt at the Design Certification stage to use either a DOE or the SC staffed “Internal” review process (See ***FIG WG-1***). If the SC “Internal” path is selected (as it is

in virtually all Microscale projects) SC may, at its own initiative, choose to use Objective Observers to gather data on the ground, or to engage an Accredited VVB/DOE to do so. Despite the rapid growth of Microscale projects, and the bundling effect of cloning, SC informed us that they have never used a VVB/DOE to verify any of these projects. In discussions with SC the use of VVB/DOEs was deemed inappropriate in Microscale projects although it is presented as an option in the assurance process. By not ever using these external verification options, this alternative is ineffective in acting as a "check" on the less independent SC "Internal" process. Further, the Quality Control implicit in this potential external review is not realized and its inclusion as an alternative can be seen as misleading.

Upon inquiry, SC was unable to identify if they were doing any regular reviews of Certification Data to identify outliers or inconsistencies that might indicate areas of their work that were underperforming for any measure (timeliness, completeness, accuracy, etc.) or reporting inconsistent results across Reviewers that might indicate the need for calibration and/or training.

It was discussed in a number of interviews that there were infrequent changes in the Estimated Issuance of Credits when conducting the processes supporting issuance of Verified Credits in the SWS portfolio (and potentially others). We recognize that changes were made at various points however we suggest a further exploration of this issue based on document reviews to quantify this area of concern and determine and track the frequency that the approved Estimated Issuance is changed, by how much, the reasons for any change and understand the rationale for relying primarily on the Design Stage Estimate.

It appears that in more than one instance in the application of the problematic TPDDTEC methodology, overestimation was not adjusted after a verification review was conducted, despite the accuracy of the estimate being questioned at various points. Also, in discussions with SC there was a clear view that the Design Review and Certification stage was seen as more rigorous than the Performance Review including Verification and, therefore, no revisions to parameters were needed despite verification findings. And also in discussion with SC, it was identified that no Issuance had ever been rejected, as they recalled, although downward adjustments were made.

This reliance on the Design Certification stage may have resulted in a reluctance to revise estimates downward based on the verification audit/process/review. This has the potential to allow errors or needed adjustments identified late in the certification process (at the performance review stage) to remain unresolved.

It also appeared that, in practice, Project Developers often forward-sell estimated credits through committing to future delivery. It appears that some Project Developers did not allow for the possibility of certifying fewer emission reductions than their Design Certification estimates and communicated strongly the negative impacts downward adjustments would have. SC staff verbalised sympathy to this potential difficulty for Project Developers.

Risk Management: Quantification of Risk SC was unable to discuss any systematic quantification of risk in their systems and procedures to mitigate and manage it.

Risk Management: Coordination with GS SC raised a concern about the SWS methodology with GS as early as August 2019, but did not take action to address the issue internally. While there were discussions between SC and GS that took place sporadically over several

months, even this did not cause SC to conduct any additional reviews that their processes allowed for. Nor does it appear that there was additional rigor applied to the Certification Review processes and there do not appear to be instances of lowering of emission reductions volumes certified for issuance, although there was a strong concern that overestimation was occurring. It was not until the Grievance was filed that there was a temporary, reduction enforced.

6. Oversight by the GS Secretariat

Of SustainCERT It seems reasonable that key performance indicators (KPIs) for timeliness, quality measures and other important indicators would have been developed as part of the oversight of SC by its owner and client GS. We have been told that this is currently under discussion but has not been finalized. GS has a responsibility as the Standard Owner to require certain levels of performance, to identify and track results and any risks they represent and to develop processes to address any shortfalls.

Of Project Developers It would benefit GS to be aware of the various Project Developers and formally track their relative size in the full portfolio of projects to help monitor concentrations or other matters that could result in risk. There are reputational risks that can become associated with PDs that may have differing objectives for participation than those originally envisioned by GS. While this is of course normal, it is prudent to track and be aware of any differences so risk can be well managed.

7. Technical Parameters Used in Emissions Reductions Estimates

We would like to note that we fully support the results of the BA report. We concurrently note to the reader that, as we have discussed with BA, some of the parameters are estimates of expected rational values and others are based on literature. As such there is room for different values to be considered. BA has noted these cases in their report transparently and discussed the reasons for the choices made in the values used in their analysis of overestimation and its magnitude.

C. Conclusions

The TAC SWSGWG concludes that the problems uncovered by this grievance and their root causes are complex and the result of a number of magnifying effects.

With the exploitation of the shortcomings in both the methodology and the methodological guidance, as well as combination with a series of process and procedural shortfalls, this matter has been magnified and resulted in both overestimation of credits and over-issuance. In a rising market, this has reached a significant scale.

Equally as troubling are the failures in the process of checks and balances that have been designed into the Assurance Systems. Options available for verification were un- or under-utilized, notification from reviewers of problems in the methodology were not pursued as priorities, estimates were left to stand as final numbers despite problems having been identified.

We believe that in addition to addressing the implications of the problems identified so clearly in the BA report, GS must urgently undertake a series of actions to address the

shortfalls in the application of the Assurance Systems. This includes, as a priority, revising methodologies, implementing training, active management of the quality of SC activities, and construction of strong oversight of the ability to deliver consistent and replicable outcomes no matter who conducts the review. The most urgent need is revisiting the boundaries of credible use of the microscale project rules, including definition, debundling, requirements for external reviews, PoA size, etc.

We would close our conclusions with the note that we are not aware of any individual or single act that caused this matter to arise and grow. Rather, as the organization has grown in the last years, and has experienced a rapid series of changes, the management and design of the systems have not kept pace. While the decision the institution and its leaders may take to remediate the results of this specific matter may be difficult, we want to stress that the review of processes and procedures is of equal, if not higher, importance to the long term success in meeting the mission of the organization.

Thank you for the opportunity to be of assistance in this matter. We appreciate the trust put in us and hope that you find it was well placed.

ANNEX – STAKEHOLDER COMMENTS

GOLD STANDARD FOUNDATION

Safe Water Supply Grievance

Technical Advisory Committee Grievance Working Group

INVESTIGATION REPORT

Final

CO2BALANCE'S RESPONSE TO TECHNICAL ADVISORY COMMITTEE GRIEVANCE WORKING GROUP: INVESTIGATION REPORT

TABLE OF CONTENTS

DETAILS

REPORT DATE	PURPOSE
15/10/2020	This report sets out CO2balances responses to the grievance investigation into the Gold Standard Safe Water Supply methodology and its application to project developers – full Investigation Report made available on Monday 5th October 2020.

FOREWARD

CO2balance asserts that the SWS borehole projects have been developed in accordance with the methodology and with the best of intentions. It has been open about all practices related to the measurement of the parameters mentioned in the Report, which have been validated and verified by SustainCERT and the Gold Standard for over 7 years.

A significant concern regarding the BAMG Report is that it relies heavily on general literature related to borehole functionality, use and maintenance. Few studies have actually been conducted on performance-based models such as the SWS projects, which specifically focus on the long-term sustainability and maintenance of boreholes, so it is arbitrary to draw comparisons with business as usual scenarios. The primary objective of the SWS projects is to address the problems related to borehole failure in Africa by developing specialised maintenance and community engagement programmes that are managed by local partners, who understand the local context and are in regular contact with the communities.

CO2balance is not of the opinion that the report provides a fair or accurate representation of the SWS projects. Many of the sources which form the basis of the criticisms in the BAMG Report are unique to individual studies and journals. For example, the source which supports the prevalence of rainwater harvesting is based on an area with high rainfall and extensive RWH infrastructure, and is therefore not comparable to the project areas of the SWS projects. Some sources even state that comparisons should not be drawn with other projects due to their unique nature (Thompson, et al 2019, pg. 727).

Furthermore, CO2balance strongly refutes any suggestion that the company has intentionally sought beneficial results for economic gain- this is simply untrue. The income generated from the projects has been reinvested in new programmes which aim to repair over 2000 boreholes by the end of 2021 and without carbon finance there is a risk that these programmes may no longer be able to continue or that they will need to be significantly reduced, depriving some of the poorest communities of safe water. Moreover, CO2balance has conservatively built in a buffer of credits generated by each VPA above the microscale 10,000 ER/year cap (an average of 38.5%). This is a measure that has been self-imposed by CO2balance to mitigate for the range of factors that are difficult to account for, and are not required to be accounted for, under the TPDDTEC methodology.

CO2balances recognises that there are challenges related to measuring certain parameters and would welcome the opportunity to work with SustainCERT and the Gold Standard to enhance the integrity of the methodology.

IV. PART A – FINDINGS

A. 1. USER NUMBERS

CO2balance asserts that the statement that the technical specifications from the RWSN "*suggest that…[the pumps] are designed to serve a population of 300 users*" is wrong. The figure of 300 is a conservative figure based on best case scenario figures including per capita daily consumption of 20 litres. In reality, these pumps serve many more users than this. The figure of 300 is generally a target, such as in the Ghanaian National Standard referenced in the Report (Fisher et al, 2015). The journal referenced in the Report regarding the study in Ghana itself states that borehole user numbers can exceed 300. Fisher et al (2015) states that the median number of water sources per community in their study was three and goes on to suggest that the actual number of sources may be higher. Therefore, the user numbers per borehole with hand pumps are arguably lower due to more sources being available in the communities. CO2balance assesses safe water access in communities and only includes boreholes in the SWS Portfolio if there are no other sources of safe water in the community, which explains the higher user numbers. Cha et al (2018) study of boreholes in the Volta Region of Ghana found that average user numbers per borehole were 499 and recognises that 300 users is a recommendation rather

than an absolute figure. Similarly, a study by the International Red Cross (2013) recognises that pumps in Ghana theoretically should serve 300 people (in line with national targets), but data shows that the average number is actually 450.

The Report also references "*a study in Malawi estimating up to 155 households per source*". The average household size in Malawi is 4.6 people (Rep of Malawi, 2012), meaning that the handpumps in the study were supplying water for as many as 713 users. This is significantly higher than the average user numbers of **544** across the SWS Portfolio (which is within the range supported by literature). Furthermore, Martínez-Santos et al (2020) state that a hand pump can meet the daily needs of 500 people. As noted in the Report the case studies chosen for analysis by BAMG are *"not representative of the average values for key parameters across the portfolio, but expressly selected because they used outlier values"* (p13). CO2balance wants to highlight that these case study project figures are extreme outliers, as evidenced by the much lower overall average figures.

The RWSN assumptions for calculating user numbers are based on a conservative yield of roughly 10 litres per minute for Afridev or India pumps. It is worth noting that this varies based on pump head, with a pump delivering more when the water table is shallow, and less when it is deep. Assumptions also conservatively include a 12-hour operating period. A study from Kenya using Waterpoint Data Transmitters found that pumps could be used up to 24 hours per day (Thomson *et al*, 2019) , and a study conducted by the Centre for Water Systems (Ingram and Memon, 2020) suggests a more conservative average of 12 - 14 hour use per day. Furthermore, the assumptions are based on an average consumption of 20 litres per person per day, whereas the actual average water consumption in rural areas is more likely to be around 13 litres per person per day (Uganda National Water Development Report, 2005).

This evidence shows that boreholes can and do support many more than 300 users. CO2balance submit that BAMG have misinterpreted the RWSN recommendation and policy targets of countries of 300 users per borehole operated by handpump as an absolute maximum number of users. BAMG appear to have selected examples from literature to support this view, while ignoring numbers which do not support their position. CO2balance collect and verify user lists for each borehole included in the portfolio, cross checking annually against randomly selected households for monitoring.

A. 2. USAGE PERCENTAGE

CO2balance disputes the figures from the referenced Thompson et al (2019) to evidence a reported 80% reduction in boreholes as a primary water source across seasons (86% in dry season to 6% in rainy season). The findings within the Thompson et al (2019) study actually demonstrate a 34% reduction in groundwater use during the wet season compared to the dry season. The 80% figure referenced in the Report is based on data from "*the area where the reduction in pump use was most marked*" (pg. 722) which is due to rainwater harvesting being widespread within this area (75% using rainwater harvesting), with many of the households having "*extensive*" rainwater harvesting infrastructure such as metal roof, guttering and dedicated harvesting tank with a mean capacity of 183l (pg. 726). Such infrastructure is not common in the Project Areas of the SWS Projects and is not targeted by CO2balance (see Picture 1). Furthermore, studies have shown that domestic rainwater harvesting systems are over 90% unreliable (Taffere et al, 2016) and are not suitable on roundhouses or grass roofs (Thomas, 2006). Typical project areas targeted by CO2balance are remote, rural areas where communities are reliant on unsafe water sources such as open wells, rivers, or streams year-round before project implementation. The communities largely do not have the resources or money to fund the maintenance or repair of their handpump and can be considered as living below the international poverty line on less than $1.25 a day.

Picture 1: Images showing typical household structures in project communities

Furthermore, the study focuses on the impact of "extreme rainfall events" and as such is not an accurate representation to which the SWS Projects should be compared. In fact, the Thompson et al (2019, pg. 727) paper explicitly states that *"the patterns of behaviour described and analysed in this paper are for this area only, and generalisations from this study should be made with caution"*.

CO2balance disputes the statement that the usage percentages of the SWS Projects are too high "*based on evidence in the literature*" as it has been shown above how the circumstances of the studies in the referenced literature are not comparable or applicable in the SWS project areas. The literature referenced shows communities with access to both boreholes and other improved sources, such as rainwater harvesting via extensive equipment. In those cases, it is reasonable that users may use the borehole less when safe water is available at their homestead. However, as stated before in the interviews, the communities in the SWS projects do not have access to other sources of safe water. Therefore, the usage percentage will not be affected by this.

CO2balance would like to point out in relation to the statement in section 2.c) *"In most of the project monitoring data we reviewed, households reported that the project borehole was their only water source, inconsistent with literature values and expert interviews"* that, in line with the Methodology, project monitoring data is only collected in relation to water use for the 3 specific claimable purposes.

CO2balance disputes the statement that *"no follow-up questions were included to ensure that water collected from other sources was not used for drinking, personal hygiene, and/or food preparation".* In order to get quantifiable data on this, the Water Consumption Field test is designed to measure all water used for drinking, basic personal hygiene, and food preparation and distinguishes between

measurements of water for these purposes sourced from the borehole and measurements of water for these purposes from other sources.

Furthermore, while CO2balance agrees with the statement that *"it is unclear from GS methodological guidance what specifically falls into the category of basic personal hygiene"*, the following definitions of the 3 claimable uses (agreed with GS) are supplied to the field teams for reference:

	Eligible uses	Non-Eligible uses
Basic personal hygiene	-Hand washing -Face-washing -Gender related hygiene	-Bathing water
Drinking	-Drinking water for human consumption	-Drinking water for livestock
Food Preparation	-Water used for washing foods (e.g. vegetables, fruits)	-Water used for washing cutlery and all other cooking utensils
Other		-Water used for watering crops -Water used for laundry purposes

A. 3. PROJECT TECHNOLOGY DAYS

CO2balance disputes the statement in section 3.a) that *"The methodology simply states in equation (1) of Section 2.7 that the Project Technology Days should reflect the crediting period with days removed when the technology is non-functioning".* CO2balance asserts that no mention is made to non-functioning days for the monitoring of Project Technology Days throughout the methodology.

In regard to the Report's observations on recorded downtime across the SWS portfolio, Kelly et al (2018, pg. 2) states that "*most system breakdowns occur three to five years after construction*". As stated in the Report, the majority of projects have been implemented within the last 2 years, supporting that significant breakdowns are not likely to occur in this period. CO2balance have in-country partners in place who carry out preventative and reactive maintenance, further protecting boreholes from suffering breakdowns, as supported by Rural Water Supply in Africa (Harvey and Reed, 2004, pg. 167) *"If preventative measures are carried out at appropriate intervals … the handpump should continue operating without breakdown. This requires regular preventive maintenance visits or services to detect minor faults and pre-empt problems that may result in future breakdowns".* Furthermore, there are funds available in-country that can be drawn down immediately if and when maintenance is required and therefore, significant breakdowns are kept to an absolute minimum if indeed any actually occur.

Furthermore, while the report suggests that functionality rates of boreholes range from 55% to 80%, these values are based on studies of boreholes that are not included in carbon projects or performance-based funding mechanisms. One of the primary challenges for the majority of safe water projects is securing funding for the long-term maintenance and monitoring of the safe water technology, particularly for projects implemented by NGOs who are reliant on donor or grant funds. For example, a study by the University of Oxford (2014) found that the main cause for handpump failures and delays in repairs were that repairs were too expensive.

Conversely, safe water carbon projects are unique to the majority of safe water projects as they depend on maintaining the functionality of the boreholes and therefore have a vested interest to do so, using the income from carbon credits to fund repairs when needed. Studies based on projects with maintenance service providers in place (a comparable model to that implemented by CO2balance) demonstrate that boreholes have an uptime of greater than 94% (McNicholl et al, 2019). The same study finds that for projects with providers that commit to rapid breakdown response, over 90% of breakdowns are repaired within 3 days. Longer breakdowns are observed in the 'preventive only' maintenance model that performs periodic handpump servicing but is not designed to respond immediately to breakdowns.

In addition, the SDSN (2018, pg.9) paper quoted in the Report recognises that the handpumps become non-functional due to maintenance issues and the burden placed on communities to maintain and repair the pumps. In the projects implemented by CO2balance, while there is basic maintenance training provided to the communities, the projects are premised on using carbon finance income to ensure that maintenance and repairs are carried out by the project. The SDSN report (2018, pg. 11) shows that "*98% of*

participating handpumps were operating at any given time" when the maintenance programme was implemented, Fisher et al (2015, pg. 8431) states 97% functionality is possible "*with optimal management*", and Komives et al (2008) states that sources are more likely to be functional if it is the only improved source in the community. This supports the project data that the CO2balance SWS projects have high functionality, particularly when taking into account that no other improved sources are available in the communities, such as extensive rainwater harvesting infrastructure. The SDSN paper also states that after an initial period of data transmitters playing a dominant role in reporting issues, as of 2018 the majority of issues are reported through the community hotline, which is a similar approach to the CO2balance model.

The literature that has been stated in the report actually **supports** the need for carbon finance in order to maintain the boreholes to a supply of safe water to communities; we are surprised, and somewhat concerned, that this was not acknowledged within the report.

A. 4. FUEL PER LITRE OF WATER BOILED

CO2balance disputes that projects included in the review *"generally provided instructions"* on the WBT procedure. CO2balance provides household selection and in depth guidance on carrying out the WBT in all projects.

CO2balance would like to point out that the BWBT process is designed to reflect real-world practices of fuel and technology use for the purposes of boiling water. While water may be able to be boiled more efficiently in theory (or based on laboratory based testing), based on the data collected through the BWBT process by CO2balance this is not reflective of real-world processes in the areas in the SWS project portfolio. During the BWBT process implemented in CO2balance SWS projects local households are asked to build and light a fire/stove in line with their normal practices and data is therefore gathered taking local conditions, practices, and stove efficiencies into account.

The WBT protocol 4.2.3 laid out in the Report was developed *"to assess stove performance in a controlled manner"* (pg. 2) and was therefore adapted to meet the needs of the methodological requirement to determine *"the quantity of fuel required to purify by boiling one litre of water for 10 minutes using technologies and fuels representative of the baseline scenario"* (TPDDTEC v.3.1, pg. 50). It is highly unlikely that a household in the rural communities in which CO2balance works would carry out a 3 step boiling process in order to purify water for drinking, basic personal hygiene, and food preparation purposes in line with the full WBT protocol.

Furthermore, CO2balance would like to highlight that, as noted by the Report, the BWBT process is fairly idiosyncratic to the methodology, and therefore the firepower values that the project values are being compared against are not based on the same process or cooking/boiling events. In addition, the number of tests that the values are being compared to (averages based on a total of 143 field tests – Table 2 of the Report) are less than 15% of the total tests conducted by CO2balance across the SWS project portfolio (a total of 969) and therefore consideration has to be made for the relatively minimal sample size referenced by the Report, and the large reported ranges (average of 7.9kW) in each referenced Country data set (and CoV for Ghana and Mozambique). It should furthermore be noted that the WBT Protocol itself it states that *"If the COV of the benchmark values for fuel use and energy use is greater than 25% among 3 tests, the tester should perform an additional test to increase confidence in comparisons"* (WBT Protocol v.4.2.3, 2014, p58), as such CO2balance asserts that there is not adequate confidence in the reported firepower results to draw comparisons where the CoV is above 25%.

It is furthermore clear from the Report that given the ranges of results based on field test values there is not an accepted TSF firepower range and results vary based on the area in which the tests are conducted, supporting the collection of field data in the SWS project portfolio. CO2balance reiterates that all the BWBT tests have been carried out in line with the methodology and have been approved by SustainCERT and GS multiple times.

B. 1. TREATMENT CAPACITY

CO2balance disputes the statement that *"all treatment capacity estimates referenced the 16 litres per minute value, and no other discharge rates were mentioned in the calculations"*. CO2balance provided examples (e.g. in ER Calcs of GS7135) whereby borehole depth was taken into account to estimate discharge based on the Handpump Technical Specifications. In addition, if an average of

l/min across the borehole depths is taken, it is 17.1 l/min for Afridev and 19.3 l/min for India Mark II, indicating that a yield of 16l/min as conservative.

Furthermore, the assertion that based on the technical specifications boreholes have a treatment capacity of 6000 litres per day would assume that boreholes are only used for 6.25 hours per day (based on 16 l/min – a conservative average based on the same technical specifications). As evidenced by a study conducted by Centre for Water Systems (Ingram and Memon, 2020, pg. 8) *"there appears to be no period in daylight when collection is not done"* and the study demonstrates an average of 12 - 14 hour use per day, while also highlighting the variability demonstrated by each different community setting even within the same country. This also calls into question the reference in the Report for runtimes of 4.09 and 5.90 hours per day. Furthermore, this reference relates to an electric pump that delivers 300 litres of water per minute (generating between 73,620 – 91,620 litres per day based on hours per day runtimes),which far exceeds the amounts claimed for in the SWS projects and is therefore not a valid basis for comparison.

B. 2. WATER QUALITY

The statement *"the methodology only addresses microbiological quality" is incorrect.* As per the methodology; *"As a first option, projects shall meet host country standards (where available) for treated water quality"* (TPDDTEC, v.3.1, p52). CO2balance conducts microbiological and chemical water quality testing for all projects, even in countries where national standards are not available, going beyond methodological requirements. Common chemical parameters tested include but are not limited to; Nitrate (NO_3), Nitrite(NO_2), Arsenic (As), Flouride (F), Ammonia (NH_3), Iron (Fe), Chlorine (Cl), Manganese (Mn), Iron (Fe), Phosphates (PO_4^{3-}), Sulphates (SO_4), PH, Total Dissolved Solids and Magnesium (Mg+).

Furthermore, the statement that Lead (Pb) is *'potentially introduced into the water by safe water intervention, thus violating the principle of "do no harm"* is unsubstantiated. There is little evidence to suggest that lead contamination is prevalent in handpump systems in Africa. As per the WHO Guidelines for drinking-water quality (WHO 2006 p.393), *"Lead is exceptional in that most lead in drinking-water arises from plumbing in buildings and the remedy consists principally of removing plumbing and fittings containing lead",* which is not applicable to the borehole SWS projects. All pipes used in boreholes are either made of steel or plastic and one can therefore deduce that the incidence of lead contamination is very minimal or more likely non-existent.

B. 3. DISTANCE TO BOREHOLES

CO2balance wishes to clarify that the distance to the borehole for each household is collected during the Household List data collection process. This data is cross-checked against values reported in the biennial WCFT to ensure consistency.

C. MAGNITUDE OF OVERESTIMATION/ OVER-ISSUANCE

In order to ensure conservativeness and account for any uncertainty arising from the methodology, CO2balance has inbuilt a self-imposed 'buffer' of credits for each VPA above the microscale 10,000 ER/year cap (a decision made by CO2balance to ensure conservativeness). In line with Table 5 in the Report this buffer is an average of 38.5% (6,529 tCO2e/yr). As, in line with GS Requirements, only the capped values have ever been issued or claimed. CO2balance asserts that the comparison should be restricted to the capped (and claimed/issued) values, to do otherwise is a misrepresentation of the figures.

In relation to the "Reference estimate" for each Parameter, CO2balance comments are as follows:

- Fuel to boil water, calculated based upon a 5kW TSF firepower: CO2balance asserts that a 5kW firepower value is not only far below the values that CO2balance has recorded from a significant volume of field tests (based on the cooking/boiling event being claimed for) but is also below the values suggested by literature (which as indicated in Section IV.A.4.c is 7kW) and is therefore not appropriate to use as a reference estimate
- Users per technology (borehole): CO2balance asserts that based both on reported user numbers in literature and technology treatment capacity that 300 people per borehole is an unrealistically conservative figure.
- Technology days, based on 75% functionality: CO2balance asserts that the references used in the Report do not take into account the maintenance and repair structures put in place by the safe water projects implemented by carbon project

developers. Comparable studies based on performance-based funding mechanisms demonstrate functionality rates of between 94% to 98%.

- Usage Rate: CO2balance asserts that the values from literature in the Report are non-comparable as they are based on communities that also have access to other safe water sources (such as rainwater harvesting infrastructure) and therefore shouldn't be used as reference values

V. PART B – FINDINGS

A. 2. SHORTCOMINGS OF PROJECT DEVELOPER MONITORING

In relation to the statement that *"there are practical and potentially even ethical limits to the extent of ongoing tracking and influencing of communities that are geographically remote from the project developers over a span of 5 or more years of project operations",* CO2balance wishes to emphasise that it works with local field teams for the implementation of the projects with community engagement being a key aspect of all projects. Without the project's existence and ongoing maintenance the communities would likely revert back to not having a SWS which itself is a significant ethical concern. Ongoing engagement, monitoring, and maintenance is a fundamental aspect of safe water carbon projects and is vital to ensure long-term use and accurate calculation of emission reductions generated by the project. As such, this statement is inaccurate and needs removing from the report.

A. 3. LIMITATIONS IN THE CERTIFICATION PROCESS

CO2balance contends that the reason for developing micro scale projects was not to subvert any risk mitigation measures, which is supported by the fact that the company already has a large scale project and has developed a new small scale PoA for SWS projects. The reasons for choosing to develop micro-scale projects are three-fold. Firstly, many buyers prefer to have exclusive rights to an entire project. Secondly, the micro-scale programme offers a quicker route to issuance. Thirdly, it provides a more cost-effective way to road-test small projects in new countries that would otherwise be unfeasible.

In relation to the statement that *"The reviewers and the VVB were aware that the reasonableness of the resulting parameters was also a component of their mandate, yet in our limited assessment, we found significant variation in the norms around how to assess reasonableness and what actions to take as a result of this assessment."* CO2balance would like to emphasise that the feedback received from SustainCERT varies from project to project and it is sometimes apparent that the reviewers are not well versed with the projects or methodology. The lack of clarity in the methodology combined with the inconsistent auditing processes represent a major challenge for project developers.

B. OBSERVATIONS ON SOURCES OF RESPONSIBILITY FOR OVER-ESTIMATION

The Report states that *"In one case, a developer pro-actively sought and received permission to reconduct their BWBT using a protocol that was reportedly more representative of local practice, and consistent with their economic self-interest, also more likely to result in a higher parameter estimate."* CO2balance asserts that the test was not reconducted due to economic self-interest but to more accurately reflect the real boiling practice in the project area. CO2balance would not benefit from any increased issuance resulting from the test, as the credits are owned by a partner NGO.

After the original BWBT was conducted in 2017, it was discovered that kerosene was used by field staff to light the fire for the test which is not common practice among the communities. The results varied considerably from other tests that had been carried out in similar project areas in Eritrea, so there was reason to believe that they were inaccurate. Whilst progressing through the registration review for the project, CO2balance explained the situation to the Gold Standard and requested to reconduct the BWBT, which was approved on 23/05/2018.

The results of the new BWBT were audited and approved through the Design Change procedure by SustainCert. At no point, did SustainCert raise any concerns in regard to reconducting the test. Furthermore, it is important to note that the test results are in line with the field testing range estimates provided by BAMG on p.24 of the report.

With the explanation clearly provided, it is requested that this section be removed from the report as the statement is inaccurate, misleading and libelous.

We are appreciative to see that on Page 36 of the BAMG Report the comment that following a review of case studies (which we consider are extremes anyway) "it is important to note that there was no evidence the project developers did not follow the methodologies and that the parameter values used in these VPAs were ultimately allowed to be used through multiple layers of oversight in the certification process"

We are also pleased to see that the BAMG Report on Page 41 says *"our investigation did not uncover any evidence of intentional misreporting. To the extent discernable, the findings show developers conforming to the rules and methodology requirements and the certification process."* Co2balances applied the rules and methodology correctly both in carrying out the projects and having them certified by both Gold Standard and SustainCERT over many years. The BAMG Report also concluded in this section *"If the responsibility for retrospective action is placed exclusively or even significantly on project developers, the GS should expect that such action would likely bring legal challenges"*. As CO2balance have done nothing wrong, we would endorse this view.

VI. RECOMMENDATIONS

A.1 WATER QUANTITY

CO2balance would like to add that by moving to a sensor or observation sampling approach it would not be possible to differentiate between water used for each of the claimable purposes in the project. If applying a first principles approach to this volume it would still be necessary to ensure that the amount being claimed for is not water that will still be boiled (e.g. for cooking, bathing/washing), and is being used for claimable purposes under the methodology.

The nature of the 7 litres/person/day cap accounted for in the methodology is that any WCFT value below this would apply the WCFT survey value, this account for SWS projects that experience less water consumption per capita (e.g. water filter projects where the water is predominantly only used for drinking water). Therefore it would be incredibly difficult to apply a sensor or observation sampling approach to calculate water quantity claimable by the project; and to combine this with first principles calculations to accurately estimate fuel to boil water provided by the project for which emission reductions are claimable.

As such, CO2balance would welcome additional, more prescriptive guidance on the WCFT process in the methodology or alternatively applying a default value of 7L for the 3 claimable purposes.

A.2 NUMBER OF USERS

CO2balance would welcome additional methodological guidance on putting in place a more rigorous field process for determining user numbers. It is important to note that the current user number figures collected by CO2balance (an average of 544) are in line with the range of values supported by the literature. While some policy targets exist (such as the one in Ghana referenced by the Report), CO2balance maintains that it is vital to collect accurate user data that reflects the on the ground practices of the communities that CO2balance work in. Applying a standardised cap across project areas does not reflect community variation or specific pump capacity. Therefore, CO2balance maintain that calculations should be based on data collected from the field and supports the development more rigorous processes to ensure that these figures can be verified.

It would be necessary to ensure practical and logistical constraints are taken into consideration in regards to GPS/GIS mapping of every household at each SWS within the rural communities that project developers work in.

A.3 USAGE PERCENTAGE

In regards to ensuring enhanced surveying accounts for seasonality, consideration needs to be had for how this would interact with the methodological requirement that usage surveys are conducted within the last 6 months of the monitoring period (TPDDTEC v3.1 p31). This requirement means that project developers are currently unable to carry out seasonal surveys under the methodology and a change to allow this would be welcome.

A.4 PROJECT TECHNOLOGY DAYS

Based on the literature from comparable project maintenance and repair models, CO2balance would assert that a default value of between 94% - 98% be applied with the option to demonstrate higher functionality rates based on maintenance and downtime reporting protocols.

Additional methodological guidance on documentation for reporting maintenance and downtime would be welcomed.

A.6 WATER QUALITY

CO2balance supports WQTs at the household to account for possible contamination of water during transportation and storage. This would be required to be carried out on a sample of households to be logistically and financially feasible.

Chemical WQTs are already carried out at borehole level in line with National and WHO Standards.

A.7 FUEL TO BOIL WATER

First Principles Approach – based on energy requirements (Table 7):

CO2balance asserts that the assumed stove efficiency and water loss (%) included in the calculation would be highly susceptible to local fuel and stove practices, as well as environmental conditions (including humidity, wind, pot size, use of lids, etc.) and therefore applying default values across all project areas without conducting field tests would be wrong.

Furthermore, the example calculations provided in the Report do not transparently take into account the energy loss of the system due to heat exchange with the environment (based on Newton's Law of Cooling) for the duration of the simmering time, and instead make an assumption of water loss during the test. This could be accounted for if the BWBT process measured the Water Loss % which would accurately represent the system's total energy requirements for both evaporation to the environment, and evaporation due to heat exchanges with the environment based on applying heat to the system for a set period of time.

First Principles Approach – based on firepower and stove efficiency (Table 8):

CO2balance asserts that further field tests on water boiling events need to be carried out on TSFs that demonstrate lower CoVs before the reported average (7kW) can be applied as a reference value in the methodology. Furthermore, in line with the comments made above related to the approach based on energy requirements- assumed stove efficiency, water loss (%), and time to boil included in the firepower calculation would vary significantly based on local fuel and stove practices, as well as environmental conditions, and therefore applying default values across all projects areas would not give accurate results.

As is highlighted in Table 8, the Wood per Litre calculated in the Firepower approach does also not take into account any water loss from the system, which is not representative of how the system would actually function (in line with energy requirements approach).

CO2balance would welcome more prescriptive guidance on how the BWBT could be conducted. In terms of using firepower as a diagnostic measure, the "*sensible range"* would need to be based on comparable field test data (and cooking events) in line with an acceptable CoV from the data.

REFERENCES

Cha, S., Cho, Y., Kim, SJ., Lee, Y., Choi, S., Asuming, P., Kim, P. & Jin, Y. (2018). Cost-benefit analysis of water source improvements through borehole drilling or rehabilitation: an empirical study based on a cluster randomized controlled trial in the Volta Region, Ghana. *Global Health Action*, 11:1, 1523303

Clean Cooking Alliance (2004). The Water Boiling Test Version 4.2.3, Cookstove Emissions and Efficiency in a Controlled Laboratory Setting. Available: https://www.cleancookingalliance.org/technology-and-fuels/testing/protocols.html

Fewtrell, L., Kaufmann, R. and Prüss-Üstün, A. (2003). Lead: assessing the environmental burden of disease at national and local level. Geneva, World Health Organization. WHO Environmental Burden of Disease Series, No. 2

Fisher, MB., Shields, KF., Chan, TU., Christenson, E., Cronk, RD., and Leker, H.. (2015). Understanding handpump sustainability: Determinants of rural water source functionality in the Greater Afram Plains region of Ghana. *Water Resources Research* 51:8431–8449; doi:10.1002/2014WR016770

Gold Standard (2015). Technologies and Practices to Displace Decentralized Thermal Energy Consumption (TPDDTEC), Version 3.1. Available: https://globalgoals.goldstandard.org/

Harvey, P.A. and Reed, R.A. (2004). Rural Water Supply in Africa: Building Blocks for Handpump Sustainability WEDC, Loughborough University, UK.

Ingram, W. and Memon, FA. (2020). Rural Water Collection Patterns: Combining Smart Meter Data with User Experiences in Tanzania. *Water*. 12(4):1164.

International Red Cross International Water and Sanitation Centre (2013). Applying a life-cycle costs approach to water Costs and service levels in rural and small town areas in Andhra Pradesh (India), Burkina Faso, Ghana and Mozambique. *January 2013, Full Working Paper*

Kelly, E., Shields, K.F., Cronk, R., Lee, K., Behnke, N. and Klug, T. (2018). Seasonality, water use and community management of water systems in rural settings: Qualitative evidence from Ghana, Kenya, and Zambia. *Sci Total Environ* 628-629:715-721

Martínez-Santos, P., Martín-Loeches, M., Díaz-Alcaide, S. and Danert, K. (2020). Manual Borehole Drilling as a Cost-Effective Solution for Drinking Water Access in Low-Income Contexts. *Water*. 2020, 12, 1981

McNicholl, D., Hope, R., Money, A., Lane, A., Armstrong, A., van der Wilk, N., Dupuis, M., Harvey, A., Nyaga, C., Womble, S., Favre, D., Allen, J., Katuva., J., Barbotte, T., Buhungiro, E., Thomson, P. and Koehler, J. (2019). Performance-based funding for reliable rural water services in Africa. *Uptime consortium, Working Paper 1*. Available at: https://www.smithschool.ox.ac.uk/research/water/report-performance-based-funding.html

Oxford/RFL (2014). From Rights to Results in Rural Water Services – Evidence from Kyuso, Kenya. *Smith School of Enterprise and the Environment, Water Programme, Working Paper 1*, Oxford University, UK.

Republic of Malawi (2012). Integrated Household Survey 2010-2011 - Household Socio-Economic Characteristic Report. September 2012. Available at https://tbinternet.ohchr.org/Treaties/CEDAW/Shared%20Documents/MWI/INT_CEDAW_ADR_MWI_19519_E.pdf

SDSN TReNDS (2018). Handpump Data Improves Water Access. Available at https://www.sdsntrends.org/research/2018/11/27/case-study-smart-handpump-project

Taffere, GR., Beyene, A., Vuai, S.A.H., Gasana, J. & Seleshi. Y (2016). Reliability analysis of roof rainwater harvesting systems in a semi-arid region of sub-Saharan Africa: case study of Mekelle, Ethiopia. *Hydrological Sciences Journal*. 61:6, 1135-1140. DOI: 10.1080/02626667.2015.1061195

Thomas, T. (2020). Roofwater Harvesting in Africa. *Waterlines*, vol. 24, no. 4, 2006, pp. 12–14. JSTOR, www.jstor.org/stable/24684814. Accessed 13 Oct. 2020

Thomson, P., Bradley, D., Katilu, A., Katuva, J., Lanzoni, M. and Koehler, J. (2019). Rainfall and groundwater use in rural Kenya. *Sci Total Enviro*. 649:722-730

UNESCO World Water Assessment Programme (2006). Uganda National Water Development Report – 2005. UN-WATER/WWAP/2006/9. 220 p., maps

World Health Organization. (2006). Guidelines for drinking-water quality: incorporating first addendum. Vol. 1, Recommendations. – 3rd ed.

World Health Organization. (2017). WHO Guidelines for drinking-water quality: fourth edition incorporating the first addendum. Geneva: World Health Organization. Licence: CC BY-NC-SA 3.0 IGO.

Response by Likano Project Development GmbH on the report submitted by Berkeley Air Monitoring Group (BAMG) on "Potential over-issuance risk from Safe Water Supply projects"

LIKANO is a project development a company based in Austria focusing on the development of Gold Standard projects in rural areas of Sub-Saharan Africa. We welcome the opportunity to comment on the report prepared by the Berkeley Air Monitoring Group as part of the grievance process. We have carried out an analysis of the report mainly in regard of the literature cited in this report, as the cited research papers are the basis of the argumentation by the authors of the report. Main findings of the authors were also subject to this analysis. We have cross-checked main parameters towards scientific plausibility, interpretation by the authors and applicability for SWS projects according to the Gold Standard TPDDTEC methodology.

Introduction

Handpumps are the key technology to provide access to safe water in rural areas of Sub-Saharan Africa. The following statements confirm this fact, but also raise challenges: maintenance and sustainability:

- *However, 663 million people still lack access to an improved source of drinking water, more than 300 million in sub-Saharan Africa alone [WHO/UNICEF, 2015].*
- *In rural sub-Saharan Africa, most improved water sources are boreholes with handpumps. (Fischer et al., 2015). The majority of those who enjoy access to water from an improved source rely on boreholes with handpumps (Sansom and Koestler, 2009).*
- *Around 30% of handpumps in Sub-Saharan Africa break down within a year of installation, and of those that do not, many are left unused or non-functional for a large part of the year (Alexander et al., 2015; Foster, 2013; IRC, 2018).*
- *With the lack of sustainability so well known, why are the same methods for service provision continually implemented? This issue points to a disconnect between knowledge and action when it comes to the effort to achieve the Sustainable Development Goal (SDG) for safe water. If we continue to invest in building infrastructure but do not focus on the sustainability of the systems we construct, we will never achieve access to water and sanitation for all. (Lane, M., 2018.)*

LIKANO has been active in the provision of safe water access for a number of years. In all our projects, we rehabilitate existing boreholes, which have been built years ago (the majority of them with donor money), but broke down over time due to lack of manpower and financial sources. We provide the required resources for rehabilitation and secure the proper operation and maintenance of these boreholes through the provision of financial resources, technical capacity and training.

The main issue with the report published by BAMG is that it lacks the acknowledgment of basic differences in underlying operational models. It compares the baseline of safe water supply with the projects implemented under the Gold Standard. Under the baseline, boreholes are being drilled (as mentioned mainly supported by donor funding), training is being provided, but there is a lack of maintenance, technical capacity and financial funding for all these activities. Over time, boreholes are not being maintained properly, there is no funding for repairs and in the end many boreholes stop operation.

In SWS projects under the GS, companies like LIKANO work under a performance-based system, where technical input, training and financial funds are being provided to ensure the operation of water supply points over the crediting period of the projects. Whereas users benefit from the access to safe water, LIKANO benefits from the generation of carbon credits, which secure the financing of the safe water access.

The studies cited by BAMG in the report are based on the traditional model of funding being provided for the installation or rehabilitation of boreholes, with little or no support for ongoing operation. This is confirmed by Truslove, who states: ""*One-time investment" approaches adopted by NGOs and donors, and investments into new assets, risk leaving service providers who struggle to provide the maintenance and major repairs required for sustainability unsupported.*" It is self-explanatory that a system, where little or no funding is provided for O&M delivers considerably worse results than a system where the project developer has a vetted interest in maximizing the provision and use of safe water. We are highly surprised that this important differentiation is not reflected in the report.

Chapter II.A: Characteristics of the SWS portfolio

Figure 2 on page 7 shows under GS5047 10 projects with start of crediting period in 2017 and 12 projects in 2020. This is incorrect, as no SWS projects with start of crediting period in 2017 are included in GS5047.

Chapter IV.A.1: Number of Users

Background
The overall portfolio of borehole projects had an average of 544 users per borehole. The authors state that "*handpumps used in the SWS portfolio suggest they are designed to serve a population of 300 users*".

Finding 1
From the report (page 16):
One study in Malawi estimates the number of households up to 155 per source. (Holm et al. 2017).

Comment:
According to the Malawi Demographic and Health Survey (Key Findings, 2015-2016) the average household size in Malawi is 4.5, adding up to almost 700 users per borehole.

Finding 2
From the report (page 16):
Interestingly, the study in Ghana referenced a maximum borehole user standard of 300 users per source, with 7% of sources in the study exceeding that threshold. (Fisher et al. 2015)

Comment:
Original text from the referenced study for comparison:
"*Approximately 7% (*n *= 33) of boreholes with handpumps exceeded the **national** standard [of Ghana] for the maximum **allowable** number of users per borehole (300).*"

If there is a national standard defining the maximum number of users, it is misleading that only 7% of the boreholes exceeded that maximum number. Additionally, if there is a standard defining a maximum user number, this cannot be compared with countries, where no such limit exists. A limitation of 300 users will lead to a might tighter network of boreholes than in countries where no such limitation exists.

In the study of Fisher et al. (2015) a total of 1,509 water sources were enumerated but only 508 interviewees had knowledge of the number of users (37%), the vast majority (63%) did not indicate the number of users. There are two possible conclusions: if only around a third of interviewees can state the number of users for their borehole, this raises serious doubts about the accuracy of numbers stated. Also, it is possible that the majority of boreholes without indication of the number of users exceed the national allowable number of 300.

Finding 3
From the report (page 17):
Of the three case studies we examined closely, no change in the number of households was reported between monitoring periods.

Comment:
This is interesting, as the LIKANO project (which is one of the three case studies) finalized the first monitoring in spring 2020 and no information on the second monitoring period has been shared up to now.

Moreover, given the high population growth rates in Sub-Saharan Countries (including Rwanda) it is most likely that the number of borehole users is rather increasing than decreasing from year to year.

Chapter IV.A.2: Usage percentage

Background
The authors capped the usage rate for the rainy season, assuming a 75% use of boreholes according to a research paper from Kelly et al. 2018.

Finding 1
From the report (page 18):
A qualitative study of rural water systems across Ghana, Kenya, and Zambia, confirmed seasonality

as a significant factor with improved groundwater sources being used more often in the dry season (Kelly et al. 2018)

Comment:
In the study of Kelly et al. (2018) also solar powered borehole pumps were used, which did not work in the rainy season and therefore people had no access to water from the borehole. No borehole in our portfolio uses solar powered pumps.

"Four (4 of 18, 22%) of the communities included in this study had solar-powered mechanized water systems. During the rainy season, solar-powered systems did not operate or operated for reduced hours due to increased cloud cover. One community member in Kenya explained "We are almost in July, when the weather will actually change to cloudy; at that time this [solar-powered] system cannot work because it depends on light intensity. That means at that time people will not access this water." (Kelly et al. 2018)

Finding 2
From the report (page 18):
A study in Kenya and Ethiopia with electric pumps similarly found a 23 % increase in borehole runtime following weeks with no rainfall (Thomas et al. 2019)

Comment:
The study states:
"We collected groundwater extraction data from 221 water points across northern Kenya and Afar Region, Ethiopia serving a total of over 1.34 million people. 171 of these sites were monitored with satellite connected sensors. Additionally, approximately 50 sites were monitored with cellular connected sensors, however between November 2017 and April 2018, the Ethiopian government restricted cellular data service in the region, rendering the data collection from these sites non-viable for this period. Each of these extraction points consisted of electrically powered motorized pumps, extracting water from depths ranging between 4 and 276 m. Each site distributed water to elevated tanks and local distribution networks. Water uses include livestock, irrigation, and drinking water, and we estimate the pumps sampled in this study were used by an estimated average of 2000 people in Kenya and 250 people in Ethiopia."; "Estimated water production yields per borehole based on survey data indicated 18 m^3 per hour in Ethiopia and 19 m^3 per hour in Kenya, with high variability between sites." (Thomas et al. 2019)

It is obvious, that findings from study about a groundwater extraction with borehole depths of up to 276 m, a water production yield of 19,000 liters per hour (!) and water uses for mainly livestock and irrigation should not be comparable to those boreholes in the SWS projects.

Finding 3:
From the report (page 18):
A study in Kenya reported 86% of households naming boreholes as their primary water source during the dry season, with over half reporting the borehole as their only source, but in the rainy season the percentage dropped to 6% (Thomas et al. 2019).

Comment:
The area where the study in Kenya was conducted had an extensive use of rainwater harvesting (66% of households), which enables many households to be self-sufficient during the wettest months and provide some buffer into drier months.

"Observation data from the transect survey was consistent with these responses, with 35% of households having 'extensive' rainwater harvesting (RWH) infrastructure, defined by having a metal roof, guttering and dedicated harvesting tank. The mean storage capacity of these systems was 183 l. A further 31% of households had 'some' RWH infrastructure." (Thomas et al. 2019)

Chapter IV-A.3: Project Technology days

Background
The authors reduced technology days by 25%, assuming that 1 in 4 handpumps in sub-Saharan Africa are non-functioning according to (Foster et al. 2019).

Finding 1
From the report (page 20):
A comprehensive review of handpump functionality found approximately one in four handpumps in sub-Saharan Africa to be non-functional at any point in time (Foster et al. 2019).

Comment:
This is the situation at the start of the implementation of SWS projects, **before** any borehole was rehabilitated – the project **baseline**. This is confirmed by literature, for example, a study in Ghana found strong dependencies of functionality on implementer, pump type, management, and the availability of tools, with synergistic effects from management determinants on functionality, increasing the likelihood of a source being functional from a **baseline of 72% to more than 97%** with optimal management and available tools. (Fisher et al., 2015)

Finding 2
From the report (page 20):
In Rwanda, a comparison study of three different maintenance models showed a decrease in handpumps downtime from 152 days under a "nominal maintenance" model to 57 days under a "circuit rider" model and 21 days when real-time sensor monitoring was used to dispatch technicians. (Nagel et al. 2015)

Comment:
The study included additional data on functionality, which is relevant, as well as even much better results in Kenya.
"In the study period, the nominal maintenance group had a median time to successful repair of approximately 152 days, with a mean per pump functionality of about 68%. In the circuit rider group, the median time to successful repair was nearly 57 days, with a per pump functionality mean of nearly 73%. In the ambulance service group, the successful repair interval was nearly 21 days with ***a functionality mean of nearly 91%.";*** *"While cost estimates per pump and per person are roughly similar in this study, the benefits of the increased functionality on a per pump basis may exceed this simple accounting. In particular, a similar study recently conducted in rural Kenya*

levered sensors to demonstrate an order of magnitude improvement in pump reliability (handpump downtime reduced from 27 to 2.6 days on average). (Nagel et al. 2015)

Finding 3
From the report (page 20):
In Malawi, a sample of approximately 15,000 boreholes found about a 5% increase in functionality, from 66.5% to 71.4%, when a regular service provider was present. (Truslove et al. 2019)

Comment:
The author interpreted the wrong numbers. The functionality with service provider was 71.4% and without 57.6%, an **increase of about 14%** instead of 5% (Truslove et al. 2019, Table 1, page 7).

It is also important to look at the definition of "service provider" in the study. Service providers can be centralized or decentralized and when decentralized these included "*area or water mechanic, community members, an institution, local government, NGO, self-supplied, public operator, water point committee or water user association.*"

Truslove concludes on page 14: "*The drive for decentralisation or "community-led" management of these rural water supplies has left the rural populations of Malawi with the burden of maintaining these assets. However, it is well established that service providers struggle to provide the maintenance and major repairs required to keep services operational sustainably. The reactive approach to the O&M and CapManEx of supplies contributes to the decline of functional assets, which is compounded by the notable effect of depreciating infrastructure across the MDG era. This has produced a growing need for rehabilitation exercises to bring the supplies that were implemented primarily during the early MDG era back to an operational standard. Proactive approaches to adequately maintain these supplies are necessary to prevent or postpone these costly rehabilitation exercises, which are an unsustainable practice due to their reliance on external support and the limited capacity for local governments to fund.*"

It is important to understand that project developers like LIKANO provide the required funding and technical support to avoid problems identified in Malawi. We have an active approach to O&M with regular visits to boreholes and provide funding for necessary investments, including the replacement of entire pumps in case of failure. Therefore, the situation in Malawi with decentralized service providers with lack of funding cannot be compared to the situation Rwanda.

Work and services provided to boreholes include among other activities:

- Minor repairs and maintenance on the boreholes as and when required, with technicians located close to the borehole and trained by the in-country service provider. This includes the following repairs: handle replacement, bearing replacement, axle pin replacement, any minor above-ground repair or adjustment not requiring replacement of parts.
- Bi-monthly visits to all boreholes rehabilitated and recording the condition of the borehole and feedback received from the Water Resource Committee in a standard document.
- Contact with a member of the Water Resource Committee at each borehole in the districts covered during the month that the borehole in question is not visited and record the condition of the borehole and feedback received from the Water Resource Committee in a standard document.

- Inform LIKANO within 48 hours of any lapse of functionality of a borehole and engage technicians to conduct repairs, with costs for any activities other than regular O&M costs to be budgeted separately.

Chapter IV.A.4: Fuel per liter of water boiled

Finding 1
From the report (page 21):
However, laboratory testing of traditional open wood stoves has indicated lower specific consumption estimates (Teune et al. 2020)

Comment:
There are a number of issues related with citing from this study:

- As stated in the report, the results are from "laboratory testing", while results from Water Boiling Tests are carried out in the field. Obviously, these lead to different results and it is incorrect to compare laboratory results with field results.
- The tests under the Teune study where carried out in Vietnam (!), whereas the "95% of all registered and issued projects are in 11 Sub-Saharan countries" (page 6 of the report). We would have expected that authors are aware of the difference between stoves used in Vietnam and Sub-Saharan Africa and don't base conclusions on results from Vietnam.
- Teune stated in his study on page 9 the following: "*The study faces several limitations that are worth highlighting. First, this study may lack external validity, which means results seen here may not be applicable in other settings. For example, the iron bar baseline stoves commonly used in rural Vietnamese communities are somewhat different than simple three-stone fires common in Sub-Saharan Africa*.". It leaves us somewhat speechless that this limitation is not at all mentioned by the authors in the report.

Chapter IV.B.1: Treatment capacity

Finding 1
From the report (page 28):
The technical specifications additionally assume a household size of 3-5 and a population serviced of 300 people at 15-20 liters per person per day (Rural Water Supply Network, 2007a, 2007b)

Comment:
Firstly, the report refers to "manufacturers technical specifications", however, there is not a single reference to a manufacturer. All references in terms of treatment capacity are based on 2 documents by RWSN from 2007 defining technical specifications including drawings (and not analyzing treatment capacity).

Secondly, the documents from RWSN provide information on the discharge to be provided by pump. The "Installation and Maintenance Manual for the Afridev Handpump" (http://rural-water-supply.net/en/resources/details/286) provides in chapter 7.1.4 Discharge Test the following information on the discharge per minute: "*The water collected should be generally **not less than 15 liters***". On page 28 the study concludes that "*16 liters per minute falls on the very upper end of possible values that could be abstracted from the groundwater, with published field-based results*

as much as 10 times lower than those reported by GS projects". So within the same page, the report provides evidence of a required minimum discharge of 15 liters per minute and defines 16 liters as the very upper end of possible values. This is highly contradictory.

Finding 2
From the report (page 28):
The authors indicate the technical specifications of hand pumps sourced from the Rural Water Supply Network (RWSN).

Comment
The parameter „Treatment Capacity" according to the Version 3.1 of the TPDDTEC methodology has the data unit "**Liter per day**" and the source of this data is the "**Manufacturer specification/design specification**".

RWSN is **no manufacturer** of hand pumps and therefore **not qualified** to provide information regarding the treatment capacity parameter according the TPDDTEC methodology.

A simple internet research provides following **manufacturer specifications**:

pump	manufacturer	Approximate discharge	depth	weblink
India MKII	Dayliff	Up to 1,300 l/h	Up to 45 m	https://www.dayliff.com/media/com_hikashop/upload/safe/afridev_704290054.pdf
Afridev	Dayliff	900 l/h	Up to 40m	https://www.dayliff.com/media/com_hikashop/upload/safe/india_686661466.pdf
Afridev	Ajay	16.5 l/min	Up to 45 m	http://www.afridevhandpump.com/afridev-handpump.html
Afridev	Apex International	900 - 1,200 l/h	Up to 45 m	https://www.apex-international.org/hand-pumps.html

Finding 3
The authors capped the users per borehole based on the assumption that a pump can produce 6,000 l/day (page 28 and 35) according to technical specifications of the Rural Water Supply Network

Comment
Again, the Rural Water Supply Network is no manufacturer and not qualified to provide technical specifications according the TPDDTEC methodology. We have contacted several manufacturers, the president of Apex International wrote about the Afridev pump that "you can get around 10,800 l of water per day", which is consistent with the data provided in the table above.

Finding 4
The authors refer to a study of MacDonald et al. (2012) and indicate the handpump abstraction of 6-18 liters per minute.

Comment:
Mac Donald et al. (2012) examined in their study the groundwater storage for [whole!] Africa based on the effective porosity and saturated aquifer thickness; for a community water supply fitted with a handpump, a borehole must be able to sustain a supply of 6 l/min and **preferably** 18 l/min. They found out, that the aquifer productivity map of whole Africa shows, that for many African countries appropriately sited and constructed boreholes will be able to sustain community handpumps. Furthermore, the potential for yields exceeding 300 l/min is not widespread, the potential for intermediate boreholes yields of 30 l/m to 300 l/min is much higher. Rwanda has a moderate aquifer productivity of 60 – 300 liters per minute.

Finding 5
From the report (page 28):
A Study in Kenya of 300 handpumps equipped with Water Transmitters, which continually measure water use, reported an average mean abstraction of 1500 liters per day (Thomson et al. 2019).

Comment:
There are a number of issues related with citing from this study:

- The study reports an average abstraction of 1,500 liters per day, but does give no information or indication on the number of users for these boreholes. As a result, this study is totally unsuitable to give an indication on treatment capacity and we are not sure why this study was cited as the first reference in the chapter "comparison with literature".
- The Waterpoint Data Transmitter (WDT) applied in the study does NOT measure water consumption, but only measures the movements of the pump handle: "*Developed at Oxford University, the WDT uses a low-cost solid-state accelerometer to sense changes in the movement of the pump handle in order to measure pump use and estimate volumetric abstraction (Thomson et al., 2012).*" Also, the WDT is called "*experimental*" in the Thomson study. So the statement in the study that the transmitters "continually measure water use" is just wrong.
- The study carried out by Thomson aimed at "*test the hypothesis that rainfall influences handpump use, and to characterise this relationship*". The purpose of the study was not to investigate treatment capacity, so again we are not sure why this study was cited under the chapter "treatment capacity".

Chapter IV.B.2: Water quality

Finding 1
From the report (page 30):
…creating a risk that the development impacts of the SWS portfolio are being overestimated. It is possible that a portion of the SWS portfolio is actually introducing the health risk of lead poisoning, but this possibility cannot be assessed using the currently available information.

Comment:
The statements made in this section are misleading for a number of reasons:

- If the authors would have had a closer look at the development impacts, they would have found evidence of close to 100% of users confirming improvement of the health situation of users. As the authors didn't have a closer look, we are not clear why accusations are being made about overestimation of development impacts.
- The chapters "reported practices and values" followed by "comparison to literature" insinuates that there is some default on the side of project developers. We are following the requirements of the methodology. If that methodology is not sufficient, please clearly state that the methodology needs to be improved, but don't transfer the responsibility to project developers.
- The report mentions a risk of lead poisoning and the authors state that there is not sufficient information to assess that potential risk. If this is not substantiated, we are not clear why this allegation is being made.

Chapter IV.C.2: Case study results

Table 6 states for the Rwandan borehole project 1,351 users per borehole, leading to uncapped emission reductions of 19,679 tons per year, leading to an overestimation of 762%. The way this is being presented indicates that LIKANO tried to claim these emission reductions, which is not correct and highly misleading. LIKANO included 4 boreholes in each of its VPAs with the full understanding that due to the number of users per borehole emission reductions will be capped. This has been a topic of discussion with SC during the validation process. Taking the cap of 10,000 tons into consideration, the verification of the project was based on 687 users per borehole (1,351 / 19,678 x 10,000).

Cover page:

The cover page of the BAMG report includes 3 pictures taken from the LIKANO website. Whereas the pictures indicate very positive effects for the population, especially children and women, the report is heavily criticizing SWS project. Therefore, we would kindly ask BAMG to refrain from using pictures taken by LIKANO in the report.

Independent evaluation by Objective Observer

In June 2020 Mr. Jerome Bugingo has been appointed by Gold Standard as Objective Observer (OO) to visit the SWS project in Nyagatare District, Rwanda. The report can be accessed through Gold Standard.

Conclusion from the report: "*There were many people waiting to collect water at the time of visit. They however confirmed that they all get enough water. The OO assessed the usage of the boreholes and realized that for some boreholes there are more people collecting water and was interested to know if those boreholes serve them effectively, through the interviews; the OO learned that even if the users are many for some boreholes but the water is enough for their households consumption (drinking, basic personal, hygiene, food preparation and cooking). The water resources committees make sure that every household is served and, in some communities, they have a person who help the users throughout the day to bring order in water collection and make sure everyone is served.*"

References:

See BAMAG Report, additional references:

Alexander, K. T., Tesfaye, Y., Dreibelbis, R., Abaire, B. & Freeman, M. C., 2015. Governance and functionality of community water schemes in rural Ethiopia. Int J Public Health, 60, 977-86.

Lane, M., 2018. Community-based management of handpumps in rural Ghana: a quantitative analysis of what needs to change.

Sansom, K., and L. Koestler (2009), African Handpump Market Mapping Study, UNICEF, N. Y

WHO/UNICEF (2015), Progress on sanitation and drinking water: 2015 update and MDG assessment, World Health Organ. and U.N. Children´s Fund, Geneva, Switzerland

Florence, Italy
16th October 2020

Dear Gold Standard SWS Grievance Working Group,

Thank you for the opportunity for giving feedback on Investigation report assessing the potential over-issuance risk from Safe Water Supply projects and thus way give our support for improving the future GS guidelines and procedures.

Please find below our feedback targeting, as requested, the accuracy of findings of fact within the Section 2 of the report.

Page 2-25: "c) comparison with literature" and page 2-38: "table 5"

- The comparison with the literature is referring only boreholes with handpumps.
- We do consider that the findings based on the handpump functionality should not be used for making assumptions regarding the functionality of the other kind of technologies like solar powered pumps used by many SWS activities.
- Thus, the reference estimate of "274 days/year" presented in the table 5 should not be considered representative even as "illustrative purpose" for SWS projects applying other technologies than handpumps.

Pages 2-26 to 2-30: "c) comparison with literature"

- The comparison with the literature is referring only wood usage (three stone fires or stick fed stoves). No literature references to charcoal consume or firepowers are provided.
- We do consider that the findings made based on the assessment on the wood baseline should not be used for making assumptions regarding the SWS activities with charcoal baseline.

Page 2-36: "reported practices and values"

- The examples provided regarding the project developers practices is not mentioning that part of the SWS developers are recoding the GSPs points of the interviewed households both during the WCFT as well as during the Usage/Project survey.
- Thus, the statement that "project developers currently record the household's distance to the borehole during the water consumption filed test (WCFT)" is not describing in complete manner the procedures applied by project developers.

Page 2-39 to 2-24: "2. Case study results"

- Case study results suggest that the overall ERs were estimated to be 327% to 762% (139%-338%) greater than the reference estimates.
- We would like to note that the illustrative magnitude estimates made based on 3 sample projects which might give biased indications regarding the magnitude estimation. For example, we do assume that all the sample projects are applying the oldest TPDDTEC versions and made under the PoA registered several years ago i.e. not under PoA applying

the latest guidelines and requirements of GS. Moreover, all of the 3 sample projects are having ER levels in significant manner over the micro-scale cap of 10,000 tCO2e/year which, based on our own experience of the SC review process, would not any more pass the validation.

Page 2-38: “table 5”

- Reference estimate for usage rate is set as 75%. There is anyhow not provided a clear and accurate reference on how this value was defined. Thus, the referred estimate not should not be used as reference value (or as default cap value). Indeed, usage rate is strictly linked to the project implementation area, context and the availability of other water sources nearby.

Best regards,

Ulla Mauno
Climate Project Manager
Carbonsink

Dear GS Team,

I understand that we missed the deadline of 16th Oct, I would like to request please consider our feedback if possible.

Our feedback on the investigation report is as follows:

Overall feedback: In general, the investigation report talks overestimation. One of the stated reasons for which in the report is suppressed demand as well. But the investigative results seem worded in a way where this situation appears to be ignored or not included in write up.

Feedback on specific points:
Number of users:
o As mentioned in the investigation report that, it was observed from some handpump studies that number of users served in various countries was less as compared to what was reported under different projects. However, one thing which should be kept in mind is that this value may vary regionally as well. Although these values could be capped to the reasonable values. However, at times, if project results which can be corroborated with valid evidence might provide the accurate values.

Usage Percentage
o As rightly mentioned in the investigation report that seasonal effects and access to other resources should be considered, at the same time there is also a need to keep the suppressed demand parameter in mind and More clear guidance is required for the incorporating this parameter.

Methodological Issues
o Although mentioned briefly in the report, there is a need to emphasize more on the boiling of water in the baseline scenario in a suppressed demand situation. As there is no specific guidance on how exactly this situation should be proved.

Shortcomings of the project developer monitoring
o It is mentioned that some parameters could be subject to social desirability bias where respondents may over report usage in hopes of receiving further benefits. However, at times it is also observed that the respondents try to list out the setbacks faced by them during usage in hopes of receiving a better service in future.

Recommendations
o Water Quality and Project technology days: Although sensor-based approach is the most accurate however, for some project developers and owners it could be a costly affair. Thus, guidelines could be provided to further refine the simple observational sampling.
o Number of users: Measuring and reporting the transition in family size of the users is challenging. Supplementing the project developers with guidelines for this would be useful.
o Usage percentage: As mentioned, that community-based projects should have the same level of guidance as the household treatment projects. However, it is imperative to note that the questions mentioned in Annex-9 completely neglect the suppressed demand scenario.

Thus, questions under this annex are further refined or an addendum to the existing questions to meet the requirements of both suppressed demand and community-based projects.

Regards,

Rohit Garg
Senior Project Manager | Carbon and Renewables

south pole · your climate journey starts here
2nd Floor · 136 B · Safdarjung Enclave · Block B6· Humayunpur · New Delhi -110029
Phone: · · r.garg@southpole.com

October 16, 2020

The Gold Standard Safe Water Supply Grievance Working Group
c/o The Gold Standard Foundation
Chemin de Balexert
1219 Chatelaine
International Environment House 2
Geneva, Switzerland

VIA EMAIL

Dear Safe Water Supply Grievance Working Group:

The purpose of this letter is to respond to the report written by Berkeley Air Monitoring Group ("BAMG") titled "INVESTIGATION OF THE GOLD STANDARD GRIEVANCE, 'Potential over-issuance risk from Safe Water Supply projects'" ("Report") and the supplemental report ("Supplemental Report" and, together with Report, the "Reports") issued by the Safe Water Supply Grievance Working Group ("SWSWG").

I. EXECUTIVE SUMMARY

The SWSWG has requested representations concerning inaccurate findings of fact. We believe the Reports incorrectly find that SustainCERT (i) failed to correct values when it suspected overestimated claims, (ii) allowed project developers to misuse the microscale scheme, and (iii) failed to properly manage risk by using Objective Observers ("OOs") instead of Validation & Verification Bodies ("VVBs") for microscale projects. These findings are rooted in faulty assumptions that require correction as follows:

1. SustainCERT could not unilaterally fix suspected overestimated emissions reductions;
2. There is evidence that the microscale scheme is not responsible for the overestimation of emissions reductions; and
3. There is evidence that third party VVBs do not reduce risk more than SustainCERT's internal officers, who are trained VVBs.

At its core, The Gold Standard is a risk management system. We offer three recommendations for improving risk management across standard-setting and certification:

1. The standard, certification, and verification must be digitized;
2. More prescriptive methodologies with reference values should be developed; and
3. The Gold Standard and SustainCERT should work together to improve technical governance.

The Reports unearthed systemic issues in the Gold Standard requirements, SustainCERT certification, and TAC governance. We applaud BAMG's detailed work and support their findings

and methodological recommendations. On behalf of SustainCERT, and as discussed more fully below, we acknowledge the shortcomings of certification identified in the Reports. In our opinion, project developers, including the first-movers, acted in good faith in applying the methodology's confusing and sometimes contradictory rules, but we agree they must improve their monitoring.[1] This moment is an opportunity for the Secretariat, TAC, SustainCERT, and project developers to work together to make the necessary improvements to ensure The Gold Standard continues to represent best practice.

II. REPRESENTATIONS

A. SustainCERT could not unilaterally fix the suspected overestimated emissions reductions.

Although the grievance was positioned as an investigation of "safe water supply" projects, BAMG focused primarily on borehole projects, which use the TPDDTEC methodology. The Report contains a thorough analysis of TPDDTEC and we agree with the findings regarding the gaps in the methodological guidance. We request that the SWSWG consider an additional finding: unlike improved cookstoves or water filters, which are household-level technologies that also use TPDDTEC, a borehole is a community-level intervention. This is an important distinction. It helps explain (1) why TPDDTEC does not adequately apply to borehole projects, which represent 90% of the SWS portfolio, and (2) why certain values were approved by certification, to wit:

Project Technology Days

Overestimation is attributed in part to project developers claiming a borehole is in use for more project technology days than is supported by the literature. But "[r]elatively little guidance is provided in the TPDDTEC with regard to the monitoring of Project Technology Days, even though it is a key parameter in the emissions reduction calculation." See Report at 19. We agree that there is little guidance for monitoring project technology days, but the Report misses one point: TPDDTEC instructs safe water projects to calculate person days - not project technology days – using the Water Consumption Field Test. Calculating "person days" for a borehole is extremely challenging. Unlike water filters where one technology serves one household, the number of households served varies from borehole to borehole, potentially from year to year. The WCFT is therefore not a suitable format to collect the number of person days for boreholes.

Certification did not identify this problem as early as it should have. Moreover, without reference values or more prescriptive guidance in the methodology, certification used its discretion to approve these values. We take responsibility for these errors. Our TPDDTEC reviewers (like many of the borehole project developers) have backgrounds in improved cookstoves. Because cookstove projects routinely claim 365 project technology days, reviewers likely believed the reported values

[1] BAMG expressly determined that there was no evidence of intentional misreporting by the project developers they interviewed. See Report at 41 ("[T]he findings show developers conforming to the rules and methodology requirements and the certification process.") It is dangerous to infer wrongdoing without direct evidence and we will not do so here.

were reasonable. In April 2020 we began notifying safe water project developers that their previous values claiming zero downtime for maintenance will no longer be accepted. As a result, project developers have started to improve reporting on this parameter.

Treatment Capacity

The Report states that TPDDTEC Versions 1.0 and 2.0 do not require or address treatment capacity. See Report at 27. This is slightly incorrect. A rule update, dated 23rd July 2015, retroactively applied treatment capacity requirements to Version 2.0 projects. However, Version 1.0 projects were not included in the rule update. As a result, certification did not apply the treatment capacity requirements to the 76% of safe water supply projects that are registered under Version 1.0.

Number of Users

Number of users also presents unique challenges for community-level projects using TPDDTEC. In household-level projects (e.g., cookstoves and water filters), you simply record how many technologies are distributed and sold. The number of users is limited by the units sold and is therefore easily checked and known. Furthermore, the number of users per household is a minor consideration for cookstove project emissions reduction calculations, but is critical for safe water supply projects because emissions reductions are calculated per person. Once the link between number of technologies and number of households is broken, emissions reductions can be unintentionally overestimated in community-level projects in a way that is not possible at household level.

If treatment capacity were part of the methodology, this would have re-instated the link between technologies, number of households, and users for community-level projects. In the absence of this requirement, our reviewers used their discretion to approve the number of users without awareness of the significance.

Usage Survey

TPDDTEC Version 2.0 (via a rule update dated 29th April 2015) provides guidance for usage surveys for household water filtration technologies. The guidelines expressly exclude water supply projects like boreholes. See Report at 17 (incorrectly identifying the version as Version 3.1). It is important to state that no additional usage survey guidelines for community-level safe water interventions have been published, even though they were promised in 2015. See id. Instead, our reviewers used their discretion in reviewing usage surveys.

By highlighting this tension in TPDDTEC, we do not intend to absolve SustainCERT of responsibility. We take responsibility for lapses in certification where appropriate. But we strongly disagree with two criticisms. First, it is alleged that SustainCERT has a "clear view that the Design Review and Certification stage [*sic*] was seen as more rigorous than the Performance Review including verification and, therefore, no revisions to parameters were needed despite verification findings." See Supplemental Report at 3-8. The SWSWG has confused input it received from

SustainCERT. At the design certification stage, we check, among other things, the eligibility, calculations, monitoring plan, and conformity with all Gold Standard rules (the components of the design). At verification we check that the results have been obtained in accordance with the approved design. Both processes are rigorous and it is important to understand that they serve different objectives. We also routinely reduce emissions reductions from the volume estimated at Design Certification.[2]

Second, it is stated that we should have imposed stricter standards on safe water supply projects in August 2019 when we first suspected that water boiling test values were high. See Supplemental Report at 3-10. At that time we requested The Gold Standard's assistance because we believed the problem was rooted in the methodology: the baseline protocols for certain POAs were approved in 2013 in accordance with TPDDTEC and many projects and VPAs repeatedly issued Gold Standard credits over several years because they complied with TPDDTEC. While we could make adjustments on the margins, we needed a change in the standard to justify a significant downward adjustment. Otherwise, the suggestion seems to be that SustainCERT usurp the role of The Gold Standard by making up new rules on its own and based only on suspicion. That should be unacceptable to everyone. Certification must follow the standard – not vice versa. Otherwise, there is little need for a standard-setting body.

B. The microscale scheme is not responsible for the overestimation of emissions reductions.

The Reports posit that the project developers' use of the microscale scheme is responsible for the proliferation of safe water supply projects with overestimated emissions reductions. See Report at 45; see also Supplemental Report at 3-4 & 3-5. The implication is that the risk of overestimation would be substantially reduced if these projects were small/large scale because they would have undergone a VVB audit and site visit. However, there is clear evidence that non-microscale projects have submitted, and received approval for, similar values.

The focus on the use of the microscale scheme is therefore a red herring because the form of the project is irrelevant; TPDDTEC requirements are the same regardless of project size. Below are two large scale borehole projects from two different project developers that have VVB-approved values above the BAMG reference values:

	BAMG ref	Project A	Project B
Usage Rate	75%	**100%**	**58.56% #**
WBT	0.4	**0.89**	**0.73**
Users per Technology	300	**981***	**873***

[2] For example, a group of cloned microscale borehole VPAs projected 65,620 credits at Design Certification. At the start of verification they requested 50,125 credits, but we verified 30,085 credits (a difference of 20,040 credits).

* Capped by developer according to treatment capacity in later versions of TPDDTEC

Design certified and VVB-approved estimate was 95%. This value is uncharacteristically low for any TPDDTEC project. In this case the developer reported that the project experienced unexpected challenges in user sensitization in the first year. It was not a VVB intervention.

Both Project A and Project B underwent third party validation and verification with different VVBs for each review. Accordingly, the scale of the project is irrelevant to the overestimation of emissions reductions.[3]

C. VVBs do not reduce risk more than SustainCERT's internal officers, who are trained VVBs.

The Reports repeatedly take issue with SustainCERT's decision to use internal officers and OOs rather than hire third party VVBs to audit microscale projects. See e.g., Report at 45; Supplemental Report at 3-4 & 3-5. Allegedly, this practice has resulted in poor risk management. We disagree. First, the Reports ignore the fact that the SustainCERT internal officers auditing microscale projects are trained VVBs. Before joining SustainCERT, they were hired by The Gold Standard Foundation for their training, experience, and expertise in auditing community service activities. Second, the finding that SustainCERT's reluctance to use VVBs has resulted in poor risk management is incorrect. The two sample projects above reveal that VVB-approved values also exceed the Report's reference values. The VVBs auditing these projects likely approved these values because the projects complied with the methodology.

Finally, we would like to clarify the procedures we follow when hiring an OO. We seek candidates with experience in the project technology (here, Water Sanitation and Hygiene or WASH) who speak the local language to provide an unbiased and informed view. We have found that OOs often reveal facts about a project that a VVB may overlook because the VVB does not speak the local language (and must rely on a project developer to translate), a VVB will only visit monitored households (determined by the project developer), or a VVB may not be familiar with the project technology.

We aggressively screen OO candidates for conflicts of interest. All OO resumes are carefully checked to identify any history with the project developer or its competitors - we have rejected OO candidates for this reason. Each OO is required to declare any conflict of interest prior to the preparation of a workplan, and thereafter confirm that there is no conflict of interest in the contract and on the final report. Most importantly, the contractual relationship is between SustainCERT and the OO, not the project developer. The OO is therefore incentivized to thoroughly address all of our concerns about a project so we may hire them again in the future.

[3] An additional stated concern with the microscale scheme is that certain project developers have reaped economies of scale because there is a lack of upper issuance limits. See Supplemental Report at 3-6. This makes no sense. There would be no upper issuance limit if the projects were large scale. The same number of laborious and expensive studies are required regardless of scale.

While third party VVBs play an important risk management role in the Gold Standard system, an increased use of VVBs for microscale projects would not have prevented the overestimation of emissions reductions in safe water supply projects.

III. RECOMMENDATIONS

While the first part of this letter focuses on correcting findings, we turn to our three recommendations for improving The Gold Standard system for all of its users: digitization, prescriptive methodologies with clear guidance and reference values, and improved technical governance.

A. The Report is a call to action for the digitization of the standard, certification, and verification.

Several years ago, the Secretariat identified the need to upgrade its complex standard and cumbersome certification process. The Gold Standard founded SustainCERT to make the necessary investments to create software-enabled certification. Now, the Report provides a clear call to action to expedite these investments and expand SustainCERT's vision to include software-enabled verification.

The Report identifies several problems with self-reported monitoring data.[4] Project developers manually collect data on usage, project technology days, and water quantity. See e.g., Report at 16, 19, and 28. The Report recommends investigating sensor-based approaches that may be more reliable than project developers' procedures for manual data collection. We agree.

Collecting monitoring data via sensors may also provide other benefits. In addition to enhancing the quality of the data collected, monitoring data can be analyzed across a portfolio of projects, rather than on a project-by-project basis, to identify outliers. To date SustainCERT – and The Gold Standard before it – has never done that type of analysis.[5] We agree that this approach would improve risk management and SustainCERT will use its mandate to invest in software to do so.

B. The Gold Standard should consider developing more prescriptive methodologies with reference values.

Project developers place their reliance and trust in The Gold Standard system when they choose to invest in all of the tools required to achieve Gold Standard certification, including laborious field studies, human capital, and monitoring procedures. It is difficult to know whether an approach or value is correct is if the system is complex, no clear guidance is provided, and no reference values

[4] Although the Report focused on borehole projects, this problem exists across all Community Service Activity project types.

[5] Currently, the questions certification seeks to answer are whether the project complies with the standard and the applicable methodology. See Report at 40 ("[W]e found that both the SC reviewers and the accredited VVB … approached project review with the priority goal of ensuring the rules of the standard and methodology requirements had been met.") If so, the project continues through certification to issuance.

are available. This problem is amplified for innovative projects The Gold Standard was created to incentivize.

The Report found "significant variation in the norms around how to assess reasonableness and what actions to take as a result of this assessment." See Report at 40. More prescriptive methodologies would eliminate variable interpretations by project developers, VVBs, and reviewers. There should be a range of acceptable reference values for a project technology; if the project developer can provide evidence of results above a reference value, SustainCERT or VVBs should impose a higher level of scrutiny before rejecting or approving a value as reasonable.

Post grievance, it will be very tempting to provide yet more rules and guidance to follow. It is essential that rules are simple, concise, and appropriate. Clarity is needed for all users on what is important so we can align to produce credible outcomes. Moreover, The Gold Standard and SustainCERT should work together to create a continuous feedback and improvement loop to ensure methodologies reflect the most up-to-date thinking and practice.

C. The Gold Standard and SustainCERT should work together to improve technical governance.

Going forward, The Gold Standard and SustainCERT should work together to bolster the effectiveness of their technical governance models. For The Gold Standard, this could mean re-designing the role of TAC; for SustainCERT, simplifying project documentation, strengthening performance, improving training regimes, and performance managing VVBs, OOs, and reviewers.

1. The Gold Standard should consider re-designing the role of TAC.

As noted by the SWSWG, the expectations of TAC have shifted over the years. Given the depth of the issues to be resolved, we suggest re-designing TAC's role to bring it closer to certification, narrowing its focus to continuous improvement in the standard and methodologies, and supporting the Secretariat in simplifying the standard.

In the early years, TAC was expected to review project documentation for projects meeting certain milestones. As the pipeline grew over time, that expectation became unrealistic. For example, in 2019, we sent approximately 450 project notifications to TAC; we received 1 response. We agree that it is not reasonable to expect unpaid TAC members to review project notifications regularly. This responsibility should be removed from their Terms of Reference.

TAC's level of engagement with certification has other implications. Each month TAC approves new or revised rules, but doesn't always appreciate the complexities of certification or project development. Instead, SustainCERT is expected to address the unintended consequences of TAC decisions. TAC must better understand the challenges of certification and adapt the standard to these challenges. If ISEAL allows, SustainCERT should be given a voting seat on TAC, rather than its current observer role, so that certification has a stronger voice when decisions are made by the group. Alternatively, SustainCERT can report quarterly to the TAC on certification challenges.

Another important role for TAC is to supplement the Secretariat's expertise. TAC is supposed to be comprised of sectoral experts who drive continuous improvements in the methodologies to ensure they are updated and fit-for-purpose. We agree with the SWSWG that there should be a schedule of methodology reviews, and it should be driven by TAC.

Finally, TAC, the Secretariat and SustainCERT should work together to support the simplification of the standard. Every rule should be examined with simplification in mind to eliminate the complex and sometimes contradictory rules that project developers are expected to follow and SustainCERT is expected to check. Simplification is a critical risk management tool for Gold Standard projects and it should be prioritized by all three parties.

2. SustainCERT must simplify project documents, strengthen training regimes, and performance manage VVBs and reviewers.

SustainCERT has a number of initiatives in progress to improve our technical governance. We are in the process of hiring one staff member dedicated to the oversight of these initiatives.

Simplified Documents and Review Checklists

We have received feedback that template documents cause great confusion to project developers, VVBs, and SustainCERT reviewers. We agree with this feedback. Clarity is vital for a project that faces several different interlinked, complex and sometime contradictory documents to absorb and follow. We recently completed the first of several iterations to provide more clarity. For example, document sections are now linked to the relevant rules for ease of reference. We have also built checklists to follow the new templates. These checklists provide guidance to reviewers on how to deal with parameter values in TPDDTEC.

Training

SustainCERT currently provides quarterly trainings for project developers, reviewers, and VVBs. We acknowledge and agree that SustainCERT needs to improve its training curriculum for these stakeholders, as well as add regimes for new TAC members and OOs. For this to lead to transformational change, it is essential that training goes beyond repeating the rules to tackle the areas that will make a material difference between claims and *credible* claims. To that end, we plan to develop SustainCERT University, an online platform to facilitate improved training. It will have a reference library, videos, and online exams. A curriculum will be developed for each stakeholder group, and it will be continuously updated to keep pace with changes in the standard. We anticipate hiring one staff member dedicated to SustainCERT University.

Reviewer & VVB Performance Management

With the release of new documents and checklists, SustainCERT will soon implement a new performance management framework for reviewers and VVBs. The objective is to increase the quality of reviews and at the same time hold individuals accountable for their work. Low scoring reviewers or VVBs will be targeted for performance improvement and, if necessary, removed from

our roster (reviewers) or have their Gold Standard accreditation suspended (VVBs). We have one existing staff member dedicated to performance management of reviewers and VVBs.

IV. CONCLUSION

We hope these representations are helpful to the SWSWG as it concludes the grievance. We look forward to working with all parties on the necessary improvements moving forward.

Sincerely,

Marion Verles
Chief Executive Officer
SustainCERT

Gold Standard
International Environment House 2
Chemin de Balexert 7-9
1219 Châtelaine
Tel: +41 (0) 22 788 70 80
Fax: +41 (0) 22 788 70 82
info@goldstandard.org
www.goldstandard.org

RE: Gold Standard representation submission to Safe Water Supply (SWS) grievance investigation

Dear sirs,

Gold Standard acknowledges the investigation report provided by BAMG and the additional sections produced by the independent TAC Working Group. Gold Standard thanks BAMG and the independent TAC Working Group for their hard work in providing said report. Gold Standard would also like to thank all participants to the investigation to date.

Overall, the significance of the findings is considered to have fully validated the grievance process and QC/QA process.

Gold Standard observes that over-estimation is the result of three main elements:

1. The inaccurate/over-estimated reporting (noting that results observed were considered 'improbable');
2. A lack of guidance in the methodology;
3. Lapses/failures in the assurance process to adequately review and challenge reported figures leading to approval of likely erroneous results.

The accuracy of reported data is a project developer responsibility. Gold Standard observes that the reporting, when compared to the reference values provided by BAMG appears significantly inaccurate and over-estimated. Likewise, when compared with other SWS projects it is clear that the reported results in the borehole portfolio are beyond reasonable expectation.

Gold Standard therefore observes that:

- Projects have a clear opportunity to correct their own parameters to a more reasonable level in light of these findings, in order to prevent knowingly over-issuing credits;

Gold Standard
International Environment House 2
Chemin de Balexert 7-9
1219 Châtelaine
Tel: +41 (0) 22 788 70 80
Fax: +41 (0) 22 788 70 82
info@goldstandard.org
www.goldstandard.org

- SustainCERT and VVBs have the ability to apply the reference values provided in the ongoing assurance of the existing portfolio in order to prevent knowingly over-issuing credits. Gold Standard acknowledges that treatment of monitored parameters for existing projects may differ from baseline, with the latter considered during Design Certification only;
- Gold Standard will work with all parties to find solutions for existing projects and to build upon the recommendations towards new methodological provisions for new projects.

The following specific inaccuracies are highlighted for clarity:

- Whilst Gold Standard recognises that ISO standards have developed and evolved, none of the current ISO standards are related to the current Gold Standard requirements and there is a significant gap in capacity to follow these standards that would prevent their robust application at this time. Gold Standard deliberately decided not to adopt ISO standard as a requirement for these reasons, following scrutiny and decision making by TAC;
- On the 'Process & Procedure' section of the report, page 8 there is factual incompleteness in the information reported stating that 'no issuance has ever been rejected'. While it's true that no issuance request has been rejected in full, there have been several instances where issuance volumes requested for issuance have been reduced by large amounts based on findings in review;
- The SWS component of the methodology was introduced in 2010 and not 2007/8 as stated on P38 of the BAMG report.

Yours sincerely and on behalf of the Gold Standard Secretariat

Margaret Kim
Chief Executive Officer

CC Gold Standard Board and Management Team