

# Cooking the books: Pervasive over-crediting from cookstoves offset methodologies

Annelise Gill-Wiehl (✉ [agillwiehl@berkeley.edu](mailto:agillwiehl@berkeley.edu))

University of California, Berkeley <https://orcid.org/0000-0002-3908-7903>

Daniel Kammen

University of California, Berkeley <https://orcid.org/0000-0003-2984-7777>

Barbara Haya

Goldman School of Public Policy & California Institute for Energy and Environment, University of California, Berkeley <https://orcid.org/0000-0002-7010-3908>

---

## Analysis

### Keywords:

**Posted Date:** February 23rd, 2023

**DOI:** <https://doi.org/10.21203/rs.3.rs-2606020/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

## Abstract

Carbon offsets from improved cookstove projects could advance Sustainable Development Goals 13 (climate), 7 (energy), 5 (gender), and 3 (health). To legitimately "offset" emissions, methodologies must accurately or conservatively quantify climate impact. We conduct the first comprehensive, quantitative over/under crediting analysis of five cookstove methodologies, comparing them against published literature and our own analysis. We find misalignment, in order of importance, with: fraction of non-renewable biomass, fuel consumption, stove adoption, usage, and stacking, emission factors, rebound, and firewood-charcoal conversion factor. Additionality and leakage require more research. We estimate that our project sample, on average, is over-credited by 6.3 times. Gold Standard's Metered and Measured methodology, which directly monitors fuel use, is most aligned with our estimates (only 1.3 times over-credited) and is best suited for fuel switching projects which provide the most abatement potential and health benefit. We provide specific recommendations for aligning all methodologies with current science.

## Main Text

Roughly 2.8 billion people globally cook with smoky solid fuels or kerosene<sup>1</sup>, contributing to 2-3 million premature deaths annually<sup>2,3</sup> and roughly 2% of global greenhouse gas (GHG) emissions<sup>4</sup>. Efficient cookstoves can reduce GHG emissions while also supporting Sustainable Development Goals (SDGs) 13, 7, 5, and 3, related to climate change, energy access, gender equality, and health<sup>5</sup>. Monetizing the abated GHG emissions from efficient cookstove projects through the voluntary carbon market (VCM) can support these SDGs. Efficient stoves can reduce emissions by (1) using less fuel (efficiency gain) and/or a less GHG-intensive fuel (fuel switch) and (2) achieving more complete combustion which emits less methane and other pollutants. While efficient stove projects are often touted for their health benefits, only solar, electric, gas, ethanol, and select types of forced-draft pellet stoves meet the World Health Organization (WHO) definition of "clean" for carbon monoxide and particulate matter emissions, by reducing exposures enough to meaningfully improve health<sup>6,7</sup>.

Cookstove projects with credits on the VCM are registered under the Gold Standard (GS) and Verified Carbon Standard (VCS) offset registries and estimate climate benefits using methodologies primarily developed by GS and the Clean Development Mechanism (CDM). Cookstoves represent 992 out of the 7185 project activities (individually registered or included in a Program of Activities) on the VCM<sup>8</sup> and generated ~63 million total issued credits (as of November 9th, 2022). Most VCM cookstove project activities replace traditional stoves (i.e., three stone fires or inefficient biomass stoves) with improved firewood stoves, while 43 project activities only distribute a WHO-defined clean stove or fuel (Figure 1).

Studies have evaluated the effectiveness of cookstoves projects at reducing GHGs<sup>9,10</sup>, measuring those reductions<sup>11-15</sup>, reducing exposures<sup>16</sup>, and improving health and well-being<sup>17,18</sup>. While no single study has comprehensively assessed the quality of cookstoves offset credits and the methodologies used to calculate them, published studies have identified several sources of over-crediting<sup>11-15,19-21</sup>.

Drawing on published literature and our own analysis, we offer the first comprehensive over/under crediting analysis of cookstove offset projects and methodologies. We assess the four most prominent cookstoves offset methodologies (GS-TPDDTEC<sup>22</sup>, GS-Simplified<sup>23</sup>, CDM-AMS-II-G<sup>24</sup>, and CDM-AMS-I-E<sup>25</sup>) and the recently released GS-Metered<sup>26</sup> methodology (Box 1) to:

1. compare the number of credits generated with our estimated GHG reductions
2. suggest methodological updates to achieve credit quality

We discuss the accuracy and conservativeness of all factors used by the five methodologies to estimate project emissions reductions (see methods). We then recalculate the climate benefit of a purposive sample of 36 cookstoves projects, addressing ranges of uncertainty using a Monte Carlo Analysis (MCA) Framework. Our sample spans 22 countries and represents 37% of cookstove credits generated on the VCM (Figure 2). Grouping GS-TPDDTEC and GS-Simplified together, our analysis focuses on seven methodology-stove combinations: (1) GS-Firewood, (2) GS-Charcoal, (3) GS-LPG, (4) CDM AMS-II-G-Firewood, (5) CDM-AMS-II-G-Charcoal, (6) CDM-AMS-I-E-Ethanol, and (7) GS-Metered-Pellets (Tier 4+) (See Box 1 and Methods).

## Results & Analysis

### Review of factors affecting cookstove offset quality

### Adoption, usage, and stacking rates

Efficient cookstove projects reduce emissions to the extent that users: (1) "adopt" a more efficient project stove defined as the % of distributed stoves actually in use; (2) use the project stove, where "usage" is defined as the % of meals cooked using the project stove; and (3) stop or reduce "stacking," defined as % of meals cooked using the baseline stove(s) in concert with the project stove. These rates determine the change between pre- and post-project fuel use.

Methods for monitoring adoption, usage, and stacking fall into three categories. AMS-I-E, GS-Simplified, and AMS-II-G track them through short cross-sectional surveys. GS-TPDDTEC requires in-field multi-day kitchen performance tests (KPTs) for a sample of households, capturing both usage and stacking rates by directly measuring daily fuel usage. Results are then applied to the full set of project households that indicate through a survey that they use the stove (adoption). GS-Metered uses the most robust approach, directly tracking project stove and fuel use in all participating households through meters or fuel sales data.

The default surveys for all methodologies are infrequent, simplistic, and vulnerable to social desirability<sup>27–30</sup> and recall<sup>30,31</sup> biases. For example, AMS-II-G's default survey simply asks households if they used the improved stove in the last week or month. Credits are generated for all households that reply "yes" as if they used the stove 100% of the time for the entire one-to-two-year crediting period, with a small discount if they also reported using the baseline stove in the last week or month. The surveys are not designed to avoid social desirability bias, well-documented to occur when participants provide responses (such as inflated adoption and usage rates) which they believe the surveyors (hired by the cookstove project developer) want to hear. Survey-based methods are further complicated as households may suffer from recall bias in remembering stove use over the past year<sup>30,32,33</sup>.

KPTs, if done well, are reasonably robust, yet still have weaknesses. As a form of social desirability bias called the Hawthorne effect, households may change their behavior in the presence of project staff who can observe their stove choices while weighing the fuel<sup>34</sup>. Due to cost, KPTs are only required biennially on a sample of households; however, stove usage, stacking, and fuel quality and availability can be seasonal and highly variable<sup>35</sup>. Thus, KPTs might not accurately represent stove use across the participant pool over the two-year crediting period.

The projects in our sample that use surveys report adoption and usage rates much higher than rates documented in the literature (90% adoption rate and 99% usage rate compared to 53% and 48% in our and published literature reviews<sup>36</sup>), and stacking rates that are much lower (26% stacking rate compared to 76% in a literature review<sup>35</sup>). These empirical studies, performed on cookstoves projects very similar to those participating in the studied offset programs are constructed to avoid bias with frequent, comprehensive, longitudinal surveys, triangulated with photos, field-tests, and/or stove monitors. Since the offset project surveys have known biases, to estimate project climate benefits, we replace all survey-derived adoption and usage rates with literature values as the best data available. We use empirical ranges in a MCA using a triangle distribution: adoption 53% [40%, 65%]<sup>37–45</sup>, usage 48% [16%, 80%]<sup>36</sup>, and stacking 76% [28%, 100%]<sup>35</sup>. We discount KPT-derived (i.e., GS-TPDDTEC) usage and stacking rates with a MCA using a uniform distribution with the maximum based on an empirical study estimating the Hawthorne effect (-53% in usage and 29% in stacking)<sup>34</sup>. We do not correct GS-Metered.

## Fuel consumption

Methodologies use three approaches to estimate the difference between baseline and project fuel consumption. AMS-II-G and GS-Simplified determine baseline fuel use, and then use stove efficiencies to estimate fuel use savings based on surveyed adoption and usage rates. GS-TPDDTEC determines baseline and project fuel consumption separately and calculates emissions reductions as the difference between the two. GS-Metered/AMS-I-E back calculate baseline fuel consumption from the measured/surveyed project fuel use assuming the equivalent energy would have been used in the baseline by the less efficient baseline stove.

Methodologies give projects several options to determine most inputs into each of these approaches. AMS-II-G, GS-Simplified, and GS-TPDDTEC allow projects to determine total baseline fuel use using a default value (0.4<sup>46</sup>-0.5 tons of firewood/capita/year), literature, national/project survey data, or a KPT (rarely chosen)<sup>13,47</sup> (Box 1). AMS-II-G and GS-Simplified use default values for the baseline stove efficiency and determine the project stove efficiency with a laboratory test. GS-Metered and AMS-I-E determine baseline fuel consumption with default values, literature, or surveys. GS-TPDDTEC and GS-Metered allow for no flexibility in estimating project consumption, requiring KPTs and metered or sales data respectively.

Old CDM default baseline stove efficiencies are lower than those found in the literature<sup>48</sup>, while laboratory derived project stove efficiencies are higher than actual performance in the field<sup>49</sup>. For projects that use default efficiencies, we update them to the CDM Methodology Panel's 2022 recommendations, which reflect current literature<sup>46</sup>.

Baselines constructed with project-led and national<sup>50–53</sup> fuel consumption surveys are vulnerable to social desirability<sup>27–30</sup> and recall<sup>30,31</sup> biases as households may want to present affluence and struggle to estimate kilograms of fuel used<sup>30</sup>. These biases can result in abnormally high baseline and/or low consumption values, especially when used together. Without a way to ground truth fuel consumption, we simply confine fuel consumption values to a reasonable literature-derived range of 2–4 MJ/capita/day<sup>54–56</sup> energy delivered to the pot.

## Fraction of non-renewable biomass (fNRB)

Projects that reduce biomass use should only be credited for the proportion of  $CO_2$  emissions reduced from non-renewable sources. Previously, all methodologies relied on inaccurate CDM fNRB default values. As these defaults have now expired, projects may calculate fNRB values from a CDM tool<sup>57</sup> or assume a 30% default (rarely chosen). Both the earlier defaults and the tool overstate forest degradation compared to published literature<sup>4,58</sup>. Bailis et al. (2015)'s WISDOM model estimates fNRB, accounting for biomass regrowth, and geographical, ecological, and land use heterogeneity at the sub-national level<sup>4,58</sup>. The most robust fNRB approach to date, is a dynamic landscape model, MoFuSS (Modeling Fuelwood

Sustainability Scenarios)<sup>59</sup>, although few national values are currently available. Using an MCA, we replace project fNRB values with Bailis et al. 2015's "Scenario B-Low Yield" "minimum value" as the low boundary, "expected value" as the mode, and 10% over the expected value as the high boundary. On average, the projects' chosen fNRBs are 2.8 [min:1.1, max: 16.4] times Bailis et al. 2015 values<sup>4</sup>.

## Emission factors (EFs)

To translate fuel use into GHG emissions, GS uses outdated 2006 IPCC default EFs and allows, but does not require, the inclusion of upstream emissions. Counter-intuitively, to work around an early agreement prohibiting the crediting of reduced deforestation, CDM cookstoves methodologies apply a baseline EF assuming future fossil fuel use rather than biomass. This is a source of under-crediting<sup>60</sup>. We replace each approach with cooking-fuel specific EFs, including upstream emissions, from Floess et al.<sup>61</sup>, the most comprehensive, up-to-date cooking fuel EF database<sup>61</sup>. We also update all GWPs to the most recent IPCC values, accounting for distinctions for renewable/nonrenewable biomass<sup>14</sup>. Due to high uncertainty around the climate impacts of black carbon (BC) emissions from cookstoves projects<sup>62,63</sup>, we, like the current methodologies, exclude BC.

## Firewood-charcoal conversion

All methodologies allow projects replacing charcoal to use a firewood-charcoal conversion factor to estimate the amount of firewood (on a wet basis) needed to produce the equivalent weight of charcoal (on a dry basis). All used a default of 6, which has been updated to 4, based on literature<sup>46</sup>. Alternatively, methodologies allow projects to use literature to establish this conversion factor. All projects using this conversion used a value of 6 or higher. However, conversion efficiency is highly dependent on the specific location and charcoal production practices<sup>64</sup>. In our recalculations of project impact, we do not use a firewood-charcoal conversion factor but instead use charcoal upstream and point of use EFs from Floess et al.<sup>61</sup>.

## Rebound effect

Households commonly increase their overall cooking energy consumption with access to an improved stove (e.g.,<sup>65,66</sup>). The improved stove lowers the "cost" of cooking and provides another burner, allowing the household to increase their fuel consumption. Only projects that utilize KPTs capture this increase<sup>47</sup>. We reduced our emissions reduction estimation by 22% for projects that do not utilize KPTs, which we calculate with published literature<sup>39,65,67-69</sup>, and corroborate with data from three projects in our sample that reported both KPT (observed) and laboratory water boiling test (expected) values (averaging 32%).

See methods for a more detailed discussion of each factor and our investigation into additionality and leakage.

## Over/under crediting analysis results

We recalculate the climate benefits of all sample projects using the methods described above. According to our estimates, our sampled cookstove projects are on average 6.3 times over-credited ([95% Confidence Interval: 6.2,6.3]) (Box 2, Fig. 3). That is, the average project generated roughly 6 times more credits than our estimated climate benefits.

To date, our sample generated 7.8 million offset credits (each representing one metric ton of carbon dioxide equivalent (tCO<sub>2</sub>e)). We estimate their climate benefits to be just 1.1 [0.95, 1.5] million tCO<sub>2</sub>e. Extrapolating by our five methodology-stove combination, our analysis finds that out of their 53 million VCM-issued credits, only 8.7 [8.5, 8.9] million (~ 16.4%) likely should have been issued.

Respectively, fNRB, fuel consumption, adoption, usage, and stacking produce the most over-crediting: 2.3, 1.9, 1.6, 1.5 and 1.4 times (Box 2). On average, only correcting the EFs resulted in under-crediting (0.8x), while the firewood-charcoal conversion and rebound minimally affect crediting amounts (1.1x each).

We find that all methodology-stove combinations over-credit (Fig. 3 on a per stove-day basis, and Fig. 4 first panel as a percent of issued credits).

AMS-II-G-Charcoal is the most over-credited project type (26.3 [20.1, 35.4]) stemming from specific project values (fNRB ~ 1.9x, consumption ~ 1.9, firewood-charcoal conversion ~ 1.3x) and the methodology's approach (stacking ~ 3x, usage ~ 2x, adoption ~ 1.7x, and rebound ~ 1.3) that together have a multiplier effect. These sources of over-crediting are the same for AMS-II-G-Firewood, except for the firewood-charcoal conversion, but the projects' reported figures are closer to the literature-derived values (7.9 [7.8, 8.0]). GS-Firewood (12.5[12.0, 13.0]) and GS-Charcoal 2.5 [2.4,2.5]) over-credit similarly for adoption, usage, stacking, fNRB, and fuel, but GS-Charcoal is under-credited from not including upstream charcoal emissions (~ 0.7x) and had more project fuel consumption values within our reasonable range (1.6x compared to 2.2x). AMS-I-E-Ethanol over-credits 2.6 [2.5, 2.9] times from adoption (1.5x), usage (1.6x), fNRB (2.1x), and rebound (1.3x), but has under-crediting from CDM's fossil fuel EF (0.65x). GS-LPG over-credits by 8.3 [7.6, 9.2] times, stemming from fNRB, adoption, usage and EFs. GS-Metered-Pellets have the least over-crediting (1.3 [1.3, 1.3]), stemming only from fNRB and rebound, with slight under-crediting from EFs.

Over-crediting from fNRB stems from the location-specific difference in the project's fNRB values to the Bailis et al. 2015 values (Figure S4). Adoption, usage, and stacking rates affect methodology-stove combinations based on the methodology's requirements (e.g., meters, KPTs, surveys)

and the project's reported values. GS-LPG, AMS-I-E-Ethanol, and GS-Metered-Pellets on average did not report fuel consumption values outside of a reasonable range, likely due to the use of KPTs, meters, or sales data.

EF choices result in overall under-crediting (0.8x) from five methodology-stove combinations: CDM methodologies use the low EF of a projected fossil fuel as the baseline; GS-Charcoal projects do not always include upstream emissions; and GS-Metered-Pellets use a weighted average baseline EF which ultimately is lower than those in Floess et al.<sup>61</sup>. The EFs used by GS-Firewood and GS-LPG for the baseline fuels are slightly higher than Floess et al.<sup>61</sup>, leading to slight over-crediting, stemming from project-chosen values, not the switch to LPG.

Per stove-day, GS-Metered-Pellets and AMS-I-E projects reduce emissions by roughly 0.009 and 0.005 tCO<sub>2</sub>e due to their renewable feedstocks, and thus minimal project emissions. They are followed, on a per stove-day basis, by GS-Charcoal (0.004 tCO<sub>2</sub>e), GS and AMS-II-G-Firewood and GS-LPG (0.001 tCO<sub>2</sub>e), and then CDM-II-G-Charcoal (0.0002 tCO<sub>2</sub>e).

## Discussion And Recommendations

We conduct a comprehensive, quantitative over/under crediting analysis of five cookstove offset methodologies by comparing all elements of each methodology's quantification methods against published literature and our own analysis. The over/under crediting analysis we develop and demonstrate could be used to systematically assess offset quality and inform methodology improvements across all project types.

We estimate the average cookstove project in our sample to be 6.3 [6.2, 6.3] times over-credited, stemming from misalignment across numerous factors that are compounded when multiplied together. Over-crediting occurs across all seven methodology-stove combinations and 22 countries (Figure 3 & Figures S9). Our results support and build on existing literature which has critiqued cookstoves offset projects on individual and groups of factors<sup>11,13–15,19</sup>.

The majority of over-crediting stems from the broad flexibility and poorly designed tools methodologies offer project developers to determine fNRB, adoption, usage, stacking, and fuel consumption. Project developers have incentives to make choices that produce the most credits. This is evident, as robust fNRB values have been published for eight years, yet all projects have opted to use higher CDM tool-derived or default values, and some projects track purchase data, yet fail to use it in reduction estimation. Eliminating the flexibility to use methods that are neither robust nor conservative could reduce over-crediting easily, universally by 1.4-2.3x for each factor.

To accurately and conservatively estimate the climate benefits from cookstove projects, methodologies should require, and until then project developers should choose:

1. fNRB: Bailis et al. 2015's "Scenario B-Low Yield" values
2. Adoption, usage, stacking, and rebound: One of the following options:
  - a. Meters or collect fuel purchase data with a longitudinal stacking/rebound survey; if a project has metered or fuel purchase data, this option is required.
  - b. KPTs, adjusting for the Hawthorne effect with a literature-derived default, monitoring adoption with a robust longitudinal adoption survey or a conservative literature-derived default
  - c. Robust longitudinal surveys
  - d. Conservative literature-derived default values
3. Fuel consumption: Initial and updated baseline KPTs and/or robust project-led surveys; enforce a reasonable range of 2-4 MJ-delivered/capita/day.
4. EFs: Upstream, point of use, and non- EFs for each cooking fuel from Floess et al.<sup>61</sup>, removing the need for a firewood-charcoal conversion factor; use the IPCC's separate renewable/non-renewable GWPs for methane and nitrous oxide emissions, but continue to exclude BC pending future research
5. Regular updates to reflect established research advancements (i.e., MoFuSS fNRB values)<sup>59</sup>

Our findings align with a growing literature documenting high rates of over-crediting for a range of offset project types (e.g.<sup>60,70–73</sup>). Offsets have the potential to direct substantial international finance into cookstove projects that simultaneously reduce emissions, deforestation, and potentially, devastating health impacts. Over-crediting undermines these SDGs by basing the incentive structure on false claims. This creates a reputational risk for credit buyers, weakens climate action by covering on-going emissions with a lesser amount of reductions/removals, and obscures the overall effectiveness of our climate actions. Increases in offset credit prices could make more accurate quantification (i.e., meters, longitudinal surveys, KPTs) more affordable, but there is a feedback loop – poor quality keeps offset prices too low to support accurately credited projects.

Prioritizing metered fuel switch projects would progress SDGs as they are currently the most aligned with our estimates, have the greatest abatement potential, and are typically the only WHO-defined 'clean' stoves. Our results further support Gill-Wiehl and Kammen's call for the VCM to exclusively

fund WHO-defined ‘clean’ stoves<sup>74</sup>, currently only 4% of the VCM. Quality cookstove offsets could meaningfully, instead of performatively, improve the health of people and the planet.

## Online Methods

### Sample Selection

We evaluate the quality of offset credits from the methodologies with the largest number of cookstove offset project activities on the VCM: GS-TPDDTEC, GS-Simplified, CDM-AMS-II-G, and CDM-AM-I-E. We also review GS’s new Methodology for Metered & Measured Cooking Energy Devices (GS-Metered) released October 2021.

The methodologies deploy different project stoves. GS TPDDTEC (previously GS’s *Methodology for Improved Cook-Stoves and Kitchen Regimes*) is the most versatile methodology covering any thermal domestic technology switch that is less GHG intensive, including but not limited to Improved biomass, heat retention, solar, LPG, and electric stoves. CDM-AMS-I-E replaces non-renewable biomass with renewable energy (e.g., renewable biomass, biogas, bioethanol, electric stoves). Designed for smaller projects, GS-Simplified and CDM-AMS-II-G have limited scopes, only allowing for biomass efficiency projects (e.g., traditional fuelwood stove to an improved fuelwood stove). GS-Metered is designed for cookstoves with metered or other direct fuel monitoring (e.g., purchase records) such as electric, LPG, biogas, bioethanol, or advanced biomass pellet stoves.

Most cookstoves projects are structured as Program of Activities (PoAs) in which multiple similar project activities (called voluntary project activities (VPAs) on the VCM and component project activities (CPAs) on the CDM) are bundled together to allow for rapid replication, only requiring a quick check from a validator and not a full registration procedure<sup>75</sup>. In order to reflect the diversity of projects on the VCM, we evaluated VPAs separately. CDM methodologies are used on both the CDM and the VCM, but we limited our scope to only VCM-registered projects (i.e., those certified by GS or Verra).

We then identified the 15 countries with the most credits from cookstove projects on the market and selected the projects with the most credits from each country for each methodology. There were so few projects under AMS-I-E and GS-Metered that we selected all eligible projects that provided enough information to recreate offset credit calculations on a stove-day basis for individual stove types. For the GS-TPDDTEC, GS-Simplified, and CDM-AMS-II-G projects, we prioritized projects that posted at least one monitoring report and provided their exact calculations and the stove days. No AMS-I-E and GS-Metered projects had issued credits; however, we still decided to include all listed projects under these two methodologies because of the importance of these methodologies, in terms of offering other methods for monitoring stove usage and fuel consumption, and because of the greater potential emissions reductions and health benefits from fuel switch projects that these protocols accommodate. We also ensured that our sample covered all types of fuel transitions, except for electric stoves. There were no issued projects actively deploying an electric stove, and the only listed electric project under GS Metered had no files available.

This approach resulted in 36 projects spanning 22 countries and accounts for 37% of all issued credits from these cookstove methodologies on the VCM (as of November 9th, 2022). Our sample contains seven methodology-project type combinations: (1) G-Firewood, (2) GS-Charcoal, (3) GS-LPG, (4) CDM AMS-II-G-Firewood, (5) CDM-AMS-II-G-Charcoal, (6) CDM-AMS-II-E-Ethanol, and (7) GS-Metered-Pellet (WHO Tier 4+ Biomass Pellet Stove). We have no reason to believe that these projects are not representative of the entire pool of cookstove credits on the VCM. The 20 GS standard projects represent 42% of the GS credits on the VCM. The 12 AMS-II-G contain 23% of that methodologies’ credits. The 4 AMS-I-E projects did not have any issued credits as of March 31st, 2022 (when we selected the sample), and the only AMS-I-E projects that had been issued credits were also credited under AMS-II.

### ***Factors affecting cookstove offset quality***

Due to the nature of this analysis, the methods we chose to utilize in our over/under crediting analysis are also our results and inform our recommendations. Thus, our methods are summarized in the main text. Here, in the expanded methods section, we include a full standalone explanation and justification for the methods we used to estimate the climate benefit of cookstoves offset projects and our recommendations for amending the cookstoves methodologies.

### ***Note on uncertainty***

Quantification of emission reductions from offset programs is inherently uncertain. Emission reductions must be estimated against an immeasurable counterfactual scenario. Other factors, notably fraction of non-renewable biomass, upstream emissions, and leakage are also difficult to measure, and with limited research to date, involve substantial uncertainty. Since offset credits often are used to “offset” or trade with direct emissions reduction, in order to maintain the integrity of an emissions reduction claim, offset programs are tasked with estimating program impacts conservatively when there is uncertainty. Here *conservative* means more likely to under-credit than to over-credit. Our analysis uses the most rigorous and up-to-date values from the literature when available (e.g., fNRB). Instead of choosing conservative methods for all factors, we do not or minimally correct factors with little published research, notably additionality, leakage, and fuel consumption.

### ***Note on methodology versions***

All methodologies, except for recently released GS-Metered, have undergone considerable updates over the years of credit generation that affect the methodological factors we study. Our recommendations and discussion below focus on the most recent version of each methodology and any updates proposed by the registry. However, most credits on the VCM including those still available for purchase are issued under previous methodology versions. Therefore, our quantitative over/under crediting analysis assesses the credits generated regardless of the methodology version used. We note in the main text and detail below where updated methodologies address over-crediting.

#### *Adoption, usage, and stacking rates*

**Definition:** Efficient cookstove projects reduce emissions to the extent that users: (1) “adopt” a more efficient project stove, defined as the % of distributed stoves actually in use; (2) use the project stove, where “usage” is defined as the % of meals cooked using the project stove; and (3) stop or reduce “stacking,” defined as % of meals cooked using the baseline stove(s) in concert with the project stove. These rates determine the change between pre- and post-project fuel use.

We note that with these definitions, usage and stacking do not sum to 100% as a household can use the household stove for every meal (usage=100%) and simultaneously use the baseline stove for every meal (stacking=100%). These rates determine the change in biomass or alternative fuel use induced by the project. Methodologies define these rates differently; most refer to adoption as “usage,” and some incorrectly include stacking as a source of leakage (see Box 1). In this paper and analysis, we always use the terms as defined above for clarity and consistency. Projects typically report a % of baseline stove use, which is then incorporated into the fuel consumption calculation. Thus, we use the project’s documentation to fit their reporting to our definition of stacking. For example, we would remove the deduction from fuel consumption or leakage. We then exactly replicate their calculated emission reductions using our definitions before implementing our adjustment.

**Approaches:** Methods for monitoring adoption, usage, and stacking fall into three categories. AMS-I-E, GS-Simplified, and AMS-II-G track them through short cross-sectional surveys. GS-TPDDETEC requires in-field, multi-day kitchen performance tests (KPTs) for a sample of households, capturing both usage and stacking rates by directly measuring daily fuel usage. Results are then applied to the full set of project households, according to their surveyed use of the stove (adoption). GS-Metered uses the most robust approach, directly tracking project stove and fuel use in all participating households through meters or fuel sales data.

**Weaknesses in the methods used by cookstoves offset projects:** Default surveys for all methodologies, commonly used by projects, are infrequent, simplistic, and vulnerable to social desirability<sup>27–30</sup> and recall<sup>30,31</sup> biases.

The various default surveys ask 14-40 questions, which also includes general information and information on leakage and end user training. Only 1-2 questions are explicitly identified to calculate adoption, usage, and/or stacking rates. Universally, all sample surveys ask specifically about the number of times the stoves were used over a time period, without the context of how many total cooking events were conducted, the nature of those cooking events (e.g., tea, dinner), etc. For example, AMS-II-G’s default survey simply asks households if they used the improved stove in the last week or month. Credits are generated for all households that reply “yes” as if they used the stove 100% of the time for the entire crediting period, typically a year or two. Stacking is estimated separately. AMS-II-G surveys stacking through two questions, “Do you use your traditional (baseline) cookstove also?” If yes, “how many meals did you prepare using the traditional (baseline) cookstove last week or last month?”. The answers to these questions are then used to discount the amount of biomass saved from the project (see supplemental).

The current approach to use surveys to obtain these values can lead to inaccurate results. For example, households often only use an LPG stove to make tea in the mornings. These households would answer “yes” to the question, “have you used the stove regularly since you installed it?” even though this is not the same level of usage as preparing an entire meal. The survey would capture the use of the baseline stove but equating a meal of chai and porridge in the morning to heavier meals will not result in accurate estimates of fuel reductions. All default surveys ask three to four further questions about adoption/usage/stacking, but the answers are commonly ignored in emissions reduction calculations. GS’s suggested surveys are more comprehensive than CDM’s; however, they still do not collect the information required to estimate stove adoption, usage, and stacking rates (see supplemental for full depiction of each methodology’s survey). GS-Metered provides a default survey; however, this is not required as surveying is only one option to determine the baseline or leakage (see Box 1).

Social desirability bias has been well-documented in survey methods<sup>27–30</sup> whereby participants provide responses (in this case inflated adoption and usage rates) which they believe the surveyors (hired by the cookstove project developer) want to hear. Social desirability bias is more likely when the data collection is a one-off (cross-sectional), simple survey, rather than more intensive monitoring that can build rapport and more definitively confirm the households long term stove use (i.e., it is easier to over-state use once than consistently overstate)<sup>76,77</sup>. Research has established further methods to decrease social desirability bias (i.e., asking about specific days, multiple, carefully crafted questions, triangulation with photos, etc.), yet none of these methods are required by the methodologies (see section on specific survey recommendations)

Projects can inadvertently increase social desirability bias through their educational and marketing campaigns as respondents are keenly aware that being a good customer means using the project stove<sup>30</sup>. Additionally, participating households may know that project developers benefit financially from high rates of adoption and usage, and so may inflate those rates especially when asked simple questions. Since project developers profit from that desirability bias, they are not motivated to minimize it.

The survey based methods commonly used are further compromised by their infrequency as households may suffer from unintentional recall bias<sup>30,32,33</sup>. For example, a household may use a project stove every day for a few weeks, abandon it for a while, and then return to it again. The household may not remember exactly how long each phase lasted, and only respond given their recent experience. The survey asks about the last week or month, but that week or month may not be indicative of the past year or next year for which their response will hold. This uncertainty could thus lead to over- or under-crediting<sup>33</sup>, and is more likely to lead to over-crediting when paired with social desirability bias<sup>30</sup>.

These biases have been documented in both cookstove and another household-level carbon financed technology. In the cookstove literature, studies have compared the most robust and objective monitoring (i.e., wireless sensors<sup>21</sup>, stove use monitors<sup>31</sup>, KPTs<sup>78</sup>, and purchase data<sup>30</sup>) to survey data, finding higher rates of reported stove/fuel use from the surveys. Specifically, Ramanathan et al. 2017<sup>21</sup> found that using wireless monitors to track usage resulted in only 25% of the amount of usage that surveys found. Wilson et al. 2015 compared stove use monitors to self-reported surveys, finding that the surveys reported twice as much stove usage as the monitors<sup>31</sup>. Beyond cooking specifically, Pickering et al. 2017 evaluated a carbon-financed water filter program and found usage rates of 19%, while the project reported 81% over the same 2-3 year time period<sup>79</sup>.

KPTs, if done well, reasonably robustly capture usage and stacking, albeit with a few weaknesses. As a form of social desirability bias called the Hawthorne effect, households may change their behavior in the presence of project staff who can observe their stove choices while weighing the fuel<sup>34</sup>. Due to cost, KPTs are only required every two years on a sample of households; however, stacking and fuel quality and availability can be seasonal<sup>35</sup>, and stove use varies significantly between households. KPTs might not provide an accurate representation of stove use across the participant pool over the two-year crediting period. There are further concerns about KPTs (see fuel consumption).

Infrequent monitoring has not been shown to be indicative of the long-term mean. Studies have shown that even with objective measures (i.e., not surveys or KPTs) such as particulate and temperature sensors<sup>80</sup> and stove use monitors<sup>81</sup>, short-term measurements (e.g., one or two random or consecutive days) did not consistently estimate the long-term mean. Research indicates that adoption and usage diminish drastically over time, both in the short term (few months to a year)<sup>82</sup> and long term (years)<sup>37</sup>. Since estimated adoption, usage, and stacking rates are locked in until another survey (annual) or KPT (every two years) is required, this may lead to some under and over-crediting depending on the monitoring period length.

**Adoption, usage, and stacking rates in the literature:** The projects in our sample that use surveys report adoption and usage rates much higher than rates documented in the literature (90% adoption rate and 99% usage rate). As opposed to the simple, infrequent methodology surveys, research studies rely on highly engaged methods: a combination of KPTs, stove use monitors (SUMs), photos, and/or comprehensive longitudinal surveys (recommended when a high rate of bias is suspected<sup>76</sup>) that together can fully capture adoption, usage, and stacking. Further, the enumerators are trained to explain to respondents that they have no interest in the household's answers, only in the truth. Two recent review studies compiled the usage and stacking rates from studies of improved and clean cookstove programs. Jeuland et al.'s review of cookstove programs found an average usage rate of 48% (minimum (16%) and maximum values (80%)<sup>83</sup>) across eight studies located in South America, Africa, and Asia. A review of stacking identified 11 case studies, spanning Sub-Saharan Africa, Asia and Latin America and found stacking rates ranging from 28%-100%<sup>84</sup>, with an average of 76%. National data on stacking (i.e., beyond specific RCTs) is limited as most national surveys or Demographic Health Surveys only collect data on the primary fuel used (although this is currently changing under WHO's guidance). Shankar et al. argued that with access to multiple options, everybody stacks, even in high-middle income countries (e.g. using a microwave, a gas stove, and occasionally barbecuing)<sup>84</sup>.

Only one empirical study has quantified the Hawthorne effect in KPTs, although this source of bias is widely acknowledged in the cookstove literature<sup>34,85,86</sup>. Simons et al. found a 53% increase in project stove usage and 29% decrease in pre-project devices while being monitored during the KPT as compared to when stove use was monitored with sensors<sup>34</sup>. The study used a peer-reviewed algorithm to translate temperature sensor data into fuel consumption, to make it comparable to the KPT<sup>87</sup>. Cutting edge technology in the cookstove monitoring space includes a Fuel Use Electronic Logger (FUEL) which monitors household fuel consumption, air pollution, and temperature<sup>78,88,89</sup>. It has been shown to be comparable to the KPT<sup>78</sup>; however, the household was still keenly aware that they were being monitored. The FUEL requires an actively involved field team and household<sup>78</sup>. For example, the user must store all firewood in the monitor's holder, remove it immediately before use, and only add additional wood after storing it in the holder for at least a minute. Thus, this work does not yet quantify the Hawthorne effect compared to KPTs as both still are quite intrusive approaches.

To determine a reasonable range for adoption rates, we identified and reviewed nine randomized control trials (RCTs) that implemented improved and clean stove programs across Central and South America<sup>42,44,45</sup>, Asia<sup>37,43</sup>, and Sub-Saharan Africa<sup>38-41</sup>. Our review found a minimum, average, and maximum adoption rate of 40%, 53%, and 65% respectively.

**Approach:** The default surveys used by offset projects have known biases because of how they are written and conducted. If adoption and usage rates were truly as high as the projects report, and stacking as low, the cookstove sector at large would not document and bemoan the barriers to effective interventions<sup>90,91</sup>. Without metering these projects to reveal the true adoption, usage, and stacking rates, and given the known weaknesses in the survey methods used by offset projects, the research studies reviewed above provide the best evidence<sup>38-41</sup>, using multiple robust methods in parallel, to estimate the project's climate benefit. The studies within the adoption<sup>38-41</sup>, usage<sup>83</sup>, and stacking<sup>84</sup> reviews cover a range of stove types, countries, and contexts similar to those of cookstove carbon offset projects (see below for a discussion on research vs. results-based financing



projects). Stove and country specific adoption, usage, and stacking rates would limit the sample we would be able to evaluate. Rather, we use the reviews detailed above to construct a bounded range of the likely adoption, usage, and stacking rates across stove types and contexts. The range allows for uncertainty.

We replace all survey-derived adoption and usage rates with literature values as the best data available. We use empirical ranges in a Monte Carlo Analysis (MCA) using a triangle distribution: adoption 53% [40%, 65%]<sup>38-41</sup>, usage 48% [16%, 80%]<sup>83</sup>, and stacking 76% [28%, 100%]<sup>84</sup>. We discount KPT-derived (i.e., GS-TPDDTEC) usage and stacking rates with MCA with the maximum based on an empirical study estimating the Hawthorne effect (-53% in usage and 23% in stacking)<sup>34</sup>. We use a uniform distribution for the Hawthorne effect because of the greater uncertainty in the true value, given that there is only one study. We do not correct GS-Metered.

For two projects in our sample, we take a slightly different approach. One project required the destruction of the baseline stove. Another project built the improved stove in the exact place as the baseline stove. There is limited research on the effect of requiring the easily reversible commitment of removing the baseline stove. Harrel et al.'s study in India presented households the option of a second LPG cylinder in return for destroying or filling in their traditional stove. The researchers compared LPG usage and baseline stove use to similar households who obtained the second cylinder without this commitment. They found that baseline stove use decreased with the soft commitment, but 39% of households rebuilt the baseline stove<sup>92</sup>. Harrel et al. suggests that requiring the destruction of the baseline stove can lead to lower rates of stacking; however, these households were already LPG users, which is a very different context than most cookstove carbon offset projects. Further research is needed. While this one study would suggest that the adoption and stacking rates of the projects that require the destruction of the baseline stove are likely lower than a project without this requirement, the projects with this requirement are likely to still have stacking rates higher than their surveys would suggest. With this limited literature, we construct an uncertainty range within a MCA using a uniform distribution with the maximum as the mean of each of the triangle distributions outlined above (adoption: 53%, usage: 48%, and stacking: 76%). These are rare project requirements within cookstove projects on the market; no methodology requires this.

Beyond these two projects, we have no reason to believe that cookstove offset projects' rates of adoption, usage, or stacking would be any different than projects studied in the literature, especially as many projects in the literature work with NGOs similar to those implementing carbon financed projects. An important distinction is that research is often based on research-funded cookstove activities, while VCM projects are results-based schemes (i.e., VCM projects will only generate revenues if they demonstrate usage). Results-based financing should imply that cookstove offset developers are motivated to structure the project so that households adopt and use the stoves; however, obtaining high rates of adoption and usage is difficult as the literature has shown<sup>38-41</sup>. This incentive to create projects that households like is blurred with the incentive to ensure high reported usage rates. This should mean that there is a further imperative on the methodologies to meticulously track actual adoption, usage, and stacking. Yet only GS-Metered requires robust methods. If anything, the literature's low adoption and usage rates are suspected to be even lower (and stacking even higher) had the research team not been so highly engaged, limiting the household's ability to show true valuation or resource allocation (i.e., stove maintenance)<sup>37</sup>. Thus, the values derived from the literature likely represent the higher end of adoption and usage.

**Recommendations:** The methodologies should require projects to choose between the following options:

- a. Use meters or fuel purchase data for adoption and usage and conduct a longitudinal stacking/rebound survey (with observation and photos); if a project has metered or fuel purchase data, this option is required.
- b. Conduct a KPT, adjusting for the Hawthorne effect with a literature-derived default, for usage and stacking, tracking adoption with a robust longitudinal survey (with observation and photos) or a conservative literature-derived default
- c. Capture all with a robust longitudinal survey (with observation and photos)
- d. Assume all conservative literature-derived default values

**Specific survey recommendations:** Longitudinal surveys are favored when social desirability biases are likely to be high<sup>76</sup>, such as when enumerators work for a company or organization with known financial interest in the participants' answers. Well known guidance on survey design to avoid bias suggests (1) asking about specific days<sup>50</sup> (yesterday, the day before yesterday, and within the last seven days), (2) asking specifically when and what they cooked (i.e., deploying time-use strategies), and (3) even having households self-administer the surveys to reduce recall and social desirability bias<sup>93</sup>, which is often not feasible in some low and middle income settings<sup>94</sup>. Triangulating with photos, observation (i.e., is the intervention stove clearly used or put away) and asking similar questions in different forms can also reduce bias. Currently, the surveys suggest kitchen visits, taking photos, etc.; however, only one or two questions are used to calculate usage rates, not mentioning these other checks. Instead, the projects should be required to outline before monitoring how the rates will explicitly be calculated leveraging all available and relevant questionnaire data. For example, project should use the answers to multiple specific questions to support usage and the photos to confirm it. The frequency of follow-ups in robust research ranges from 1-3 weeks to once a year<sup>37,82,95</sup>; however, it would be important to capture seasonality. Stove use monitors, meters, and fuel purchase data improve the accuracy of quantifying project emissions from the new stove<sup>13</sup>; but surveys are still needed to monitor continued use of the baseline stove as stove use monitors often burn up in traditional stoves. Observation can play a large role in monitoring baseline stove use, noting if the traditional stove is warm, contains embers or signs of use, etc. To obtain trustworthy survey results, field teams must be well trained and motivated, and it is recommended that data analysts clean survey data at least every few days. We acknowledge that these are high and expensive

expectations that many projects may not have the resources to perform. However, if survey methods do not avoid bias, developers can simply use data from published studies that use robust methods.

**Notes on methodology requirements:** While KPTs used by some AMS-I-E-Ethanol projects, they are only used to estimate an amount of ethanol used in the project scenario (not all fuels used) to calculate the thermal equivalent amount of biomass that the project is displacing, assuming a static amount of energy consumption. These projects use surveys to determine adoption and usage and so we replace their survey results with literature-derived values as we do with other projects that use surveys. GS-Metered does not require the monitoring of stacking, as they only credit for project fuel usage and assume a static amount of energy consumption, like AMS-I-E. GS-Metered and AMS-I-E thus are vulnerable to over-crediting from the unmonitored use of the baseline stove only in the context of rebound (see section below). GS Metered requires that projects demonstrate continual purchase of the new fuel for use in the new stove (i.e., customers are not burning pellets for warmth instead of cooking).

Not all projects release their surveys, therefore, we do not know whether some projects use more robust survey methods than the methodologies' default surveys; however, no project is required to do so. Even if the projects use better surveys than the defaults, if these surveys did not follow established guidance on reducing social desirability and recall bias, we will still recommend they use literature values.

#### *Fuel consumption*

Methodologies use three approaches to estimate the difference between baseline and project fuel consumption. AMS-II-G and GS-Simplified determine baseline fuel use, and then use stove efficiencies to estimate fuel use savings. GS-TPDDTEC determines baseline and project fuel consumption separately, and calculates emissions reductions as the difference between the two. GS-Metered/AMS-I-E back calculate baseline fuel consumption from the measured/surveyed project fuel use assuming the equivalent energy would have been used in the baseline by the less efficient baseline stove.

Methodologies allow projects to select between several options to determine most inputs into each of these approaches. AMS-II-G, GS-Simplified, and GS-TPDDTEC allow projects to determine total baseline fuel use using a default value (0.4<sup>46</sup>-0.5 tons of firewood/capita/year), literature or national survey data, their own survey, or a KPT (rarely chosen)<sup>22,24</sup> (Box 1). AMS-II-G and GS-Simplified use default values for the baseline stove efficiency and determine the project stove efficiency with a laboratory test. GS-Metered and AMS-I-E determine baseline fuel consumption with default values, literature, or surveys. GS-TPDDTEC and GS-Metered allow for no flexibility in estimating project consumption, requiring KPTs and metered or sales data respectively.

**Weaknesses in the methods used by cookstoves offset projects:** We identified a few inaccuracies within these options to estimate fuel consumption. Old CDM default baseline stove efficiencies are lower than those found in the literature. There are well documented discrepancies between stove performance in the lab and when in use in households, often with higher performance and efficiency noted in the lab where fuel and combustion conditions are controlled and/or optimized (e.g., <sup>49</sup>). Other studies have shown that the improved stove in the field consumes the same or more biomass as the baseline stove<sup>49,96</sup>.

Numerous factors contribute to an improved stove's poor performance in the field as perfectly controlling combustion is difficult in real world scenarios. However, the largest two sources of poor in-field performance relate to the fuel and to the user. In the field, fuelwood has varied quality (i.e., moisture, wood-type, etc.) which will affect the stove's performance, while the fuelwood used in a laboratory is typically of uniform quality. Lower efficiencies in the field are also a product of users not following manufacturer guidance. For example, some stoves require small pieces of logs or branches, which creates further time and effort for the household<sup>97</sup>. A study of an improved stove in Malawi found that to avoid this extra hassle users placed larger wood pieces into the stove, which then stuck out of the combustion chamber, allowing for incomplete combustion and decreasing the stove's efficiency<sup>49</sup>.

The testing protocols (e.g., WBT 4.2.3, CCT 2.0) give little guidance regarding the specific actions of the operator during the laboratory test, and a review of thermal efficiency protocols found systematic errors associated with the test operator's behavior<sup>98</sup>. Zhang et al. states that in China's stove testing facilities, there is a saying that, "Thirty percent depend on the stove, and seventy percent rely on the operator", which implies the operator has a greater influence on the performance rating than any inherent characteristics of the stove itself<sup>98</sup>. We also note that some projects estimate project stove efficiency with the use of pot skirts, which can increase the efficiency by 5-10%; however, pot skirts are not fixed and users often remove them, especially to accommodate larger pots<sup>97</sup>. These inaccuracies together can create artificially high baseline and low project consumption values.

As described above, baselines constructed with project-led fuel consumption surveys are vulnerable to social desirability<sup>27-30</sup> and recall<sup>30,31</sup> biases, as households struggle to estimate kilograms of fuel used<sup>30</sup>, especially since use is seasonal. Kar et al.'s assessment of an LPG program in India found that surveyed consumption values can be far higher or lower than tracked fuel usage<sup>33</sup>, but other research has found that self-reported cooking fuel consumption values were significantly and statistically larger than the tracked consumption from purchase data<sup>30</sup> or stove use monitors<sup>99</sup>.

These biases can also be present in national survey data<sup>50-53</sup>. For example, in a national survey, households again may struggle to estimate the weight of firewood used daily<sup>33</sup> or inflate their use of an improved or clean stove to present themselves as having a higher level of affluence. Further,

lengthy national surveys often do not have the resources, time, or need to triangulate between questions, photos, and monitoring techniques to estimate these cooking statistics that the research studies we cite use. National surveys may track fuel consumption by asking, “Over the last month, did the household purchase/acquire/consume firewood, charcoal, or other cooking fuel?” and then ask for that amount in grams, kilograms, milliliters, liters, or kWhs<sup>100</sup>. Beyond bias, the use of national surveys may also compromise the baseline as a national average may not represent the project’s specific target population.

Further, no methodology requires the baseline value to be updated throughout the crediting period. Urbanization, decreasing household sizes, and increasing incomes trend towards less GHG intensive fuels and stoves and thus lower baseline emissions. However, project baselines are typically fixed for 5-10 years.

In theory, energy delivered to the pot should be similar or even increase (i.e., rebound effect) in the project compared to the baseline; however, most projects only show a decrease in energy consumption. We do not assume that energy needs with the baseline and project stoves are the same, since multiple factors affect a household’s cooking demand and practice. Adoption of a more efficient stove is not a guarantee that fuel consumption will decrease. Inherently, equation based, non-field methods will lead to lower fuel consumption after the improved stove is introduced. These equations assume static delivered energy, not considering true daily fuel usage, or rebound effect.

*Further detail on potential bias in field-based methods:* Even field-based methods can lead to unreasonable baseline and project fuel consumption values. KPTs, although better than surveyed consumption, are highly variable<sup>47,101</sup>. Related to our discussion on why improved stoves may perform differently in the field, field tests are more accurate to real world conditions, but less precise as those conditions introduce a lot of variability from factors such as fuel quality and user behavior. An improved stove’s performance may differ between the lab and the field, but also between households. Unlike in laboratory conditions, the tester cannot control factors such as temperature, air flow, fuel quality, cooking techniques, etc. between households<sup>47,101</sup>.

Additionally, household fuel consumption is very specific to a households’ context (size, ages of family members, types of foods cooked, occupations, etc.). Thus, sample households chosen for KPTs must be carefully selected. The variability and uncertainty in measuring a specific household’s daily fuel use can lead to KPT-derived consumption values outside of the range. Therefore, field-based methods are valuable, but still vulnerable to some bias and uncertainty.

*A potential source of under-crediting for stoves with a high turndown ratio:* A potential source of under-crediting for GS Metered and AMS-I-E is the nuance between energy delivered to the pot vs. useful energy, which are equated in the methodology. To establish baseline useful energy, the methodologies use the efficiency of the baseline stove (as determined in laboratory settings). The result is not useful energy, but energy that went into the pot. Thus, energy delivered to the pot still can contain wasted energy. The amount of wasted energy relates to the power level households were using to cook. For the most time-consuming cooking tasks, the baseline stove may only “need” a certain power; however, since baseline stoves cannot be turned down to this level, the stove may be operating at a higher power to perform the task, wasting energy. This is most apparent when it comes to keeping water boiling, which is one of the most common long-range cooking tasks. For example: a baseline stove may only need to operate at 2kW to keep water boiling, but its minimum power may be 4 or even 6kW power, leading to 2-3x higher energy delivered than what was needed. This means that improved stoves that can be turned down to deliver heat at a low level can save greater amounts of fuel than methodologies estimate, implying under-crediting.

To incorporate this source of under-crediting, we would have to establish the type and quantity of food cooked in the baseline scenario, and how much of the new fuel would be needed to cook that food in the new stove. We would have to assume that the user would in fact turn down the stove appropriately. We only note this source of under crediting. These methodologies could keep this source of under-crediting for stoves with high turndown ratios as a needed buffer of conservatism.

**Approach:** For projects that use default stove efficiency values to calculation emissions reductions, we update those figures to the 2022 CDM Methodology Panel’s recommendations, which reflect current literature<sup>46</sup>. They update the baseline firewood stove’s efficiency from 10% to 15% and a baseline charcoal stove’s efficiency from 20 to 25%.

Without a way to ground truth fuel consumption, we simply confine fuel consumption values to a reasonable literature-derived range of 2-4 MJ/capita/day<sup>54–56</sup> energy delivered to the pot. This range is reasonable beyond the literature as GS-TPDDTEC includes guidance that 0.75 ton/capita/year (~3.2 MJ-delivered/capita/day) is a ‘threshold’ or nearing the maximum guidance and that 0.95 ton/capita/year (~4MJ/capita/day) is the maximum allowable baseline. In cases of projects reporting multiple fuels, we allowed for the individual fuels to be under 2 MJ-delivered/capita/day as the total cooking fuel use was over 2 MJ-delivered/capita/day. In cases of low project consumption, we set the project value at 2 MJ-delivered/capita/day of the fuel and then recalculated the baseline value based on the project’s reported fuel savings as a percent. We are minimal in our corrections. There is likely further over-crediting than we are able to adjust.

*Fraction of non-renewable biomass (fNRB)*

Projects that reduce biomass use should only be credited for the proportion of emissions reduced from non-renewable sources. This is because projects should only be credited for emissions reductions from biomass use associated with the loss of carbon storage in natural systems over the project crediting period.

Previously, all methods relied on inaccurate CDM fNRB default values. As these defaults have now expired, projects may calculate fNRB values from a CDM tool<sup>57</sup> or assume a 30% default (rarely chosen). Both the earlier defaults and the tool rely on a simple equation and Food and Agriculture Organization (FAO) data<sup>102</sup>. Both options assume trees and bushes have no regenerative capacity, overstating forest degradation as compared to published literature<sup>4,58</sup>. Bailis et al. (2015)'s WISDOM model determines fNRB estimates that account for biomass regrowth and geographical, ecological, and land use heterogeneity<sup>4,58</sup>. The WISDOM model can be implemented at a minimum administrative unit of analysis (i.e., villages, counties, states) and has been used to estimate fNRB for 25 countries. Using fNRB values from WISDOM at the regional level, Bailis et al. (2017)<sup>11</sup> estimate abated carbon as 41-59% less than estimates using CDM and GS protocols. Johnson, Edwards, and Masera (2010)<sup>103</sup> and Bailis et al. (2017)<sup>11</sup> advocate for the use of community-level fNRB values in emission reduction calculations. National level estimates lack the resolution needed to reflect fuelwood consumption for specific communities, requisite for small-scale projects. The WISDOM model incorporates cost-distance maps to limit the uncertainty through specifically evaluating the land cover most likely impacted by the project<sup>4</sup>.

The most robust fNRB approach to date, is a dynamic landscape model, MoFuSS (Modeling Fuelwood Sustainability Scenarios)<sup>59</sup>, although only a few national values are currently available. Protocols and projects should monitor new MoFuSS values and update their crediting appropriately. In our sample, 11 projects used the CDM default values, 18 used the CDM tool, four used previous surveys from the literature, and two used preliminary values to be determined upon verification. The average fNRB used in our sample of projects was 88% [min:61%, max: 100%], while the average, if the projects had used the Bailis et al. 2015 value, would be 38% [min: 6%, max: 65%]. On average, the projects' chosen fNRB values are 2.8 [min:1.1, max: 16.4] times the Bailis et al. 2015 values<sup>4</sup>.

The CDM Methodology Panel as of June 30th, 2022<sup>46</sup> recommended that projects which individually determine the percentage through a CDM tool, must address, compare, and justify any difference to established academic literature, specifically noting Bailis et al. 2015<sup>4</sup>. However, this change has not yet been implemented. There is no guidance for which fNRB value within Bailis et al. 2015<sup>4</sup> to use, and projects still have leeway to determine their own fNRB as the Bailis et al. 2015<sup>4</sup> numbers are not enforced.

**Approach:** Within a Monte Carlo Framework, we utilize a triangle distribution which draws on Bailis et al.'s fNRB estimates for each subnational (if the subnational location is listed in the PDD) or national level, in cases where the project's scope is the entire country<sup>4</sup>. We utilize the fNRB values from Bailis et al.'s second, low-yield scenario which considers that the byproducts of deforestation are used to meet a fraction of the wood demand and then the other component represents other sources when the byproducts of deforestation are exhausted, assuming productivity from the Forest Service of India<sup>4</sup>. For the mode of the triangle distribution, we utilize Bailis et al.'s "expected" fNRB value, which Bailis et al. constructed as a linear transformation from their estimated minimum value of fNRB<sup>4</sup>. We utilize the low yield scenario's minimum value as the triangle distribution's minimum. Since we have no information about the shape of the distribution after this expected value, we set the maximum of the triangle distribution as 10% greater than the expected value.

#### *Emission factors*

All cooking fuels have and non- emissions at the point of use, and all except for collected firewood have upstream emissions. To translate baseline and project fuel use into GHG emissions, GS and CDM methodologies take two very different approaches.

GS uses outdated (i.e., 2006) IPCC default emission factors (EFs) for both and non- gasses. GS methodologies make lifecycle emissions accounting optional, a source of under-crediting when excluded, and as a result only eight projects include it. Counterintuitively, the CDM, instead of using an emissions factor for the fuel wood or charcoal used by the baseline stove for all registered projects, uses an EF per unit energy of a fossil fuel that is projected to substitute only the non-renewable biomass portion of past fuel usage. This projected EF is based on a weighted regional value of different fossil fuels: kerosene, liquified petroleum gas, and coal, all of which have lower carbon dioxide EFs than firewood and charcoal per unit of heat delivered to the pot. This avoids the restriction that CDM projects are not permitted to credit reductions in deforestation. This workaround, in theory, provides a conservative estimate of true emissions reductions. It is less complex than directly measuring the baseline technology's emission profile. It does not, however, reflect actual emission reductions. Since 2019, CDM methodologies have included the projected fossil fuel's non-gasses since (version 11.1.), a source of (past) minimal under-crediting. Although recent CDM methodologies (as of v11.1) account for upstream emissions, it is still a source of under-crediting as the upstream emissions of the CDM required are much less than those of the typical baseline fuel (firewood/charcoal) of the projected fossil fuel which is inappropriately used as the baseline.

**EFs in the literature:** Emission profiles (point of use and upstream) of various cooking fuels vary widely in the field, depending on moisture content, wood type, sourcing of feedstock, characteristics of production, distribution, etc. Since the IPCC values were established, there have been a considerable number of studies specifically looking at the lifecycle emissions of cooking fuels. Recently, Floess et al.<sup>61</sup> compiled a database of lifecycle emission factors for cooking fuels, drawing from peer-review literature<sup>10,104,105</sup>, the EPA/CCA's analysis<sup>48</sup>, and the Greenhouse Gases, Regulated Emissions, and Energy Use in Technologies Model (GREET) model<sup>106</sup> (see Floess et al.<sup>61</sup> Supplemental Table S9).

Freeman & Zerrifi and Sanford & Burney argue that the CDM and GS methodologies both result in substantial under-crediting by ignoring reductions in black carbon (BC) emissions, estimated as having 600 times the global warming potential (GWP) of <sup>15,19</sup>. Subsequent research, however, has found the climate benefits of reducing BC to be ambiguous<sup>62,63</sup>. This ambiguity is in part due to the simultaneous reduction of co-emitted species, some of which have climate cooling effects. For example, improved stoves shrink the ratio of EC (a climate cooler) to OC (a climate warmer), an effect that lessens and may negate the climate benefits of reducing BC emissions<sup>107</sup>. Grieshop et al. note that accounting for benefits over shorter time periods could counterbalance these complicating factors, as BC's short lifespan may offer immediate climate relief, while the imbalance of EC to OC has implications over a longer period of time<sup>108</sup>.

Whitman and Lehmann explain that all methodologies over-credit due to an error in the way they account for reductions in methane emissions from renewable biomass<sup>14</sup>. Instead of claiming a reduction in methane using its full 100-year GWP value, the methodologies should use a lower GWP value that accounts for the renewability of the that methane becomes in the atmosphere after its lifetime of 10-20 years. However, since the IPCC GWP figures have since increased, the old GWP values are now a small source of under-crediting.

**Approach:** We replace the methodologies' EF respective approaches with cooking-fuel specific EFs, including upstream and non-CO2 emissions, from Floess et al.<sup>61</sup>, the most comprehensive, recent database.

In practice, this means we add methane and nitrous oxide (i.e., non- gasses) for earlier CDM projects and update the IPCC values used by GS projects with the Floess et al.<sup>61</sup> figures. We incorporate (or update for the projects that already include upstream emissions) all production, processing, and distribution emissions for each cooking fuel from Floess et al.<sup>61</sup>

We updated all methodologies' GWPs to the most recent IPCC values, accounting for the most recent IPCC distinctions for renewable/nonrenewable biomass (i.e., 28 rGWP and 30 GWP).<sup>14</sup> Due to high uncertainty around the climate impacts of black carbon emissions from cookstoves projects<sup>62,63</sup>, we, like the methodologies, exclude black carbon (see online methods).

#### *Rebound effect*

Households commonly increase their overall cooking energy consumption with access to an improved stove (e.g.,<sup>65,66</sup>). The improved stove lowers the "cost" of cooking and provides another burner, allowing the household to increase their fuel consumption. The most recent literature and modeling of clean cooking transitions is moving towards including this increase in cooking energy consumption<sup>109,110</sup>. Therefore, we reviewed and analyzed the literature of improved stove studies, finding estimates of increased cooking fuel consumption which averaged to roughly a 22% increase in consumption. Only projects that utilize KPTs would capture this increase<sup>47</sup>, while laboratory water boiling tests (WBT), thermal output formulas, controlled cooking tests, and metering/purchase data would not capture this phenomenon as they assume that total household energy delivered to the pot is constant.

The average rebound effect from five published studies<sup>39,65,67–69</sup> with available datasets is 22% (i.e., consumption with the KPT was found to be 22% higher than expected consumption with an efficiency equation, assuming static consumption). The three projects in our sample that reported both KPT and WBT values found rebounds of 16%, 32% and 47%. We reduced our estimating emissions reduction estimation by 22% for projects that do not utilize KPTs to monitor project fuel consumption, reflecting the rebound effect implied in published datasets and confirmed by three projects in our sample that reported both KPT (observed) and WBT (expected) values.

#### *Additionality*

Offsets should only be generated for reductions caused by the offset program and should not generate credits from improved or clean cookstoves that would have been purchased regardless of the prospect of income from carbon revenues. The methodologies treat additionality as a binary. Projects in locations with low penetration rates of improved stoves, small projects in certain locations, or projects using specific technologies are automatically considered additional. All other projects must demonstrate that they would not have gone forward without carbon revenues using the CDM additionality tool, which has been criticized for being inaccurate<sup>60,72,111</sup>. In the published literature, only one narrative case study explores the additionality of a cookstoves project, finding strong evidence for the additionality of the studied project<sup>112</sup>. We define additionality as a proportion of participating households who would likely have not used an efficient stove were it not for the offset program. We tracked the reported percentage of control households obtaining the improved or clean stove and then weighted by the time period of the study. We also triangulated with recent projections for clean fuel use up to 2030<sup>113</sup> globally and for the countries within our sample. These results were inconclusive as two RCTs found positive spillover effects from the stove programs, suggesting even greater additionality under our non-binary definition. Others found extremely high rates of control households obtaining the stove, suggesting lower additionality, due to government cooking policy (i.e., India's PMUY policy<sup>43</sup> (see supplemental information). Given that the sources we reviewed did not converge within a reasonable range, we defer to the methodologies' tests for additionality and identify this aspect as a factor for future research.

Our recommendation to establish baselines with required KPTs and/or robust project-led surveys will also hedge against non-additional stoves. Baselines constructed from national statistics may not capture non-additional stoves, particularly in urban, heterogenous contexts where households

might already use or obtain a clean fuel.

### *Leakage*

Leakage can occur when reduced use of firewood or charcoal by project households leads to increased fuel usage by non-project households; GS methodologies include additional sources of emissions within their definition of leakage, which at times includes stove stacking, upstream emissions, etc. CDM methodologies account for leakage either through household surveys or an unjustified default value of 5% of final emission reductions<sup>114</sup>. This default is used by 50% of our sample. GS methodologies use the same 5% default value or allow projects to thoroughly evaluate five potential sources of leakage. The unjustified default of 5% is another source of uncertainty<sup>114</sup>. It is largely unknown whether the introduction of an improved stove program increases the consumption of biomass of non-project households. When projects do track leakage, they rarely report a value larger than 5%. This is not empirical evidence that 5% is conservative as projects are incentivized to find low leakage.

Analogous to our quantitative additionality assessment, we identified three cookstove randomized control trials that tracked the control household's biomass consumption from baseline to endline. Following advice from Calyx Global<sup>115</sup>, specifically Hilda Galt, we also reviewed all the GS projects for each country in our sample to find the projects that actively tracked leakage as opposed to relying on the default value. We do note that some projects inappropriately included upstream emissions or baseline stove use within leakage. We then compiled the percentage of leakage each of these projects found (see supplemental information). These results were also inconclusive as some of the RCTs found that the average control household actually decreased their consumption<sup>37,116</sup> suggesting zero or even negative leakage (although leakage was not isolated as the factor), while another found that control households increased their consumption by 33%<sup>39</sup> (see supplemental information). The GS projects tracking leakage from each country in our sample found a range of 0% to 6% leakage, with an average of 1%. Projects are not unbiased in their assessment as more leakage implies fewer credits. As in the baseline, projects devise their own questions, risk developing leading questions, or rely on a single question to determine leakage. Given that the sources we reviewed were inconclusive and potentially biased, we do not correct any leakage values and identify this aspect as a factor for future research.

### **Adjusting factors**

Using the values listed in the latest verified monitoring report or project documents of these 36 projects, we calculated the number of VERs on a per stove-day basis. We only included projects (or monitoring reports from projects) in our sample if we were exactly able to replicate the number of VERs either in total or on a per stove-time basis. Once we replicated the credits generated under the methodologies, we then adjusted all of the identified factors contributing to either over/under crediting as described above. Then, we conducted analyses isolating each factor. Finally, we conducted one analysis evaluating only ex post values across all protocols (i.e., excluding adoption, usage, and stacking rates) as at the time of sampling GS-Metered-Pellet and AMS-I-E Ethanol projects had generated no credits and thus adoption and stacking rates are ex ante values from the PDD rather than ex post values from monitoring reports as with all other protocols. In all analyses in which we adjusted the emission factors, firewood-charcoal conversion factor, or consumption, we removed GS-Metered and AMS-I-E's calculation approach and calculated the baseline emissions and project emissions separately. For example, we used the baseline and project consumption reported in their PDDs to calculate the difference between baseline and project emissions instead of using their baseline conversion factor approach (see fuel consumption section).

In total, we have analyses in which (1) all factors were adjusted, (2) only adoption rates were adjusted, (3) only usage rates were adjusted, (4) only stacking rates were adjusted, (5) only fNRB values were adjusted, (6) only emission factors (including upstream emissions) were adjusted, (7) only the firewood-charcoal conversion were adjusted, (8) only consumption (baseline and project) values were adjusted, (9) only the rebound adjustment was incorporated, and (10) all factors were adjusted, except adoption, usage, and stacking (see supplemental). For each analysis, we were then able to project the amount of over/under crediting of these cookstove projects and methodologies and on the market.

We then spliced the results by methodology-project type combination and then by country. Our sample contains 7 methodology-project type combinations: (1) GS-Firewood, (2) GS-Charcoal, (3) GS-LPG Stove, (4) CDM-AMS-II-G-Firewood, (5) CDM-AMS-II G-Charcoal, (6) CDM-AMS I-E-Ethanol, and (7) GS-Metered-Pellets.

### **Monte Carlo Framework**

A Monte Carlo Analysis is a statistical framework which calculates possible outcomes when input parameters are randomly varied within a specified range and a given distribution<sup>117</sup>. When used for fNRB, adoption, usage, and stacking rates, the Monte Carlo analysis generates values within our defined limits, following a defined distribution. We specified the simulation to run 10,000 times randomly generating new values for each of these factors and calculating an associated emission reduction. We acknowledge the inherent uncertainty within our factors and bound them within a literature derived range.

### **Extrapolating emission reductions**

We extrapolate our estimated amount of over-crediting from our sample of monitoring reports to each project's total issued credits. We then sum those to compare the amount of VERs we estimate to those issued. To extrapolate to the entire cookstove market, we take the average range of over-

crediting for each methodology-stove combination and then apply these to the total amount of credits issued for each methodology-stove combination.

## Limitations

Our study has some limitations that must moderate our conclusions. We only cover 37% of the market, 22 countries; however, we attempted to have a fully representative sample across methodologies, location, and project type (see methods). We were limited to projects that were transparent enough to provide their exact calculations or stove-days within their monitoring or validation reports. There is uncertainty around the factors, which is why we utilized a Monte Carlo Framework as explained above. We were limited by the details provided by the projects and the standards. For example, numerous projects did not specify the rural or urban setting or more specific administrative units, which is very important for fNRB.

## Commercial Credits

A few of our sampled projects credited stoves used for commercial purposes (restaurants, schools, etc.). We do not adjust their adoption, usage, or stacking rates, or baseline/project fuel consumption. There are still barriers to adoption, usage, and ending stacking for commercial institutions; however, the literature on these rates is limited<sup>118</sup>, and thus an area for future research.

All analysis was conducted in Python 3.

## Declarations

### Acknowledgements

The research that led to these results received funding from the Center for African Studies, University of California, Berkeley (AGW), Carbon Direct (BH), Google.org (DK), Katherine Lau Family Foundation (AGW & DK),

the National Science Foundation's Graduate Researcher Fellowship Program (AGW), the Mulago Foundation/Better Cooking Company (AGW), and the Zaffaroni Family Foundation (DK). We further thank all those who reviewed drafts of this work, including Rob Bailis, Ajay Pillarisetti, Hilda Galt, Quirin Oberpriller, the Carbon Credit Quality Initiative, Mattias Ohlson, Robert Müller, James Mulligan, Kyle Hemes, and Aidan Haya. We also thank Calyx Global and Donna Lee. AGW thanks Dr. Isha Ray for additional advising support.

**Data and materials availability:** All data and code are available in the supplementary information, or publicly available online [will include GitHub link upon acceptance].

**Code availability:** All data and code are available in the supplementary information, or publicly available online [will include GitHub link upon acceptance].

### Additional Information

Supplementary Information is available in the online version of the paper. Correspondence should be addressed to the first author.

### Competing Interests:

AGW has received research support from Better Cooking Company Limited, whose leadership also provided comments on a draft of the manuscript. BKH has received research support from Carbon Direct.

## References

1. Stoner, O. *et al.* Household cooking fuel estimates at global and country level for 1990 to 2030. *Nat. Commun.* **12**, 5793 (2021) doi:10.1038/s41467-021-26036-x.
2. World Health Organization. Household air pollution. (2022).
3. Abbafati, C. *et al.* Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet* **396**, 1223–1249 (2020) doi:10.1016/S0140-6736(20)30752-2.
4. Bailis, R., Drigo, R., Ghilardi, A. & Masera, O. The carbon footprint of traditional woodfuels. *Nat. Clim. Change* **5**, 266–272 (2015) doi:10.1038/nclimate2491.
5. United Nations. *Transforming our world: the 2030 Agenda for Sustainable Development*. <https://sdgs.un.org/goals> (2015).
6. World Health Organization. *WHO global air quality guidelines: particulate matter (PM<sub>2.5</sub> and PM<sub>10</sub>), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide*. <https://apps.who.int/iris/handle/10665/345329>.
7. Burnett, R. T. *et al.* An integrated risk function for estimating the global burden of disease attributable to ambient fine particulate matter exposure. *Environ. Health Perspect.* **122**, 397–403 (2014) doi:10.1289/ehp.1307049.

8. Berkeley Carbon Trading Project. *Voluntary Registry Offsets Database*. (2021).
9. Johnson, M. A. & Chiang, R. A. Quantitative guidance for stove usage and performance to achieve health and environmental targets. *Environ. Health Perspect.* **123**, 820–826 (2015) doi:10.1289/ehp.1408681.
10. Jetter, J. *et al.* 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 (2012) doi:10.1021/es301693f.
11. Bailis, R., Wang, Y., Drigo, R., Ghilardi, A. & Masera, O. Getting the numbers right: Revisiting woodfuel sustainability in the developing world. *Environ. Res. Lett.* **12**, (2017) doi:10.1088/1748-9326/aa83ed.
12. Oeko-Institut, Carbon Limits, Green Gas Management Institute (GHGMI), INFRAS, & Stockholm Environment Institute. *Application of the Oeko-Institut/WWF-US/EDF methodology for assessing the quality of carbon credits*. <https://carboncreditquality.org/download/Assessments/1.3.2%20GS%20TPDDTEC%20%2831%20May%202022%29.pdf> (2022).
13. Lee, C. M., Chandler, C., Lazarus, M. & Johnson, F. X. Assessing the Climate Impacts of Cookstove Projects: Issues in Emissions Accounting. *Chall. Sustain.* **1**, 53–71 (2013) doi:10.12924/cis2013.01020053.
14. Whitman, T. L. & Lehmann, C. J. Systematic under and overestimation of GHG reductions in renewable biomass systems. *Clim. Change* **104**, 415–422 (2011) doi:10.1007/s10584-010-9984-5.
15. Sanford, L. & Burney, J. Cookstoves illustrate the need for a comprehensive carbon market. *Environ. Res. Lett.* **10**, (2015) doi:10.1088/1748-9326/10/8/084026.
16. Pope, D. *et al.* Are cleaner cooking solutions clean enough? A systematic review and meta-analysis of particulate and carbon monoxide concentrations and exposures. *Environ. Res. Lett.* **16**, (2021) doi:10.1088/1748-9326/ac13ec.
17. Smith-Sivertsen, T. *et al.* Effect of Reducing Indoor Air Pollution on Women's Respiratory Symptoms and Lung Function: The RESPIRE Randomized Trial, Guatemala. *Am. J. Epidemiol.* **170**, 211–220 (2009) doi:10.1093/aje/kwp100.
18. Smith, K. R. *et al.* Personal child and mother carbon monoxide exposures and kitchen levels: Methods and results from a randomized trial of woodfired chimney cookstoves in Guatemala (RESPIRE). *J. Expo. Sci. Environ. Epidemiol.* **20**, 406–416 (2010) doi:10.1038/jes.2009.30.
19. Freeman, O. E. & Zerriffi, H. How you count carbon matters: Implications of differing cookstove carbon credit methodologies for climate and development cobenefits. *Environ. Sci. Technol.* **48**, 14112–14120 (2014) doi:10.1021/es503941u.
20. Simon, G. L., Bumpus, A. G. & Mann, P. Win-win scenarios at the climate-development interface: Challenges and opportunities for stove replacement programs through carbon finance. *Glob. Environ. Change* **22**, 275–287 (2012) doi:10.1016/j.gloenvcha.2011.08.007.
21. Ramanathan, T. *et al.* Wireless sensors linked to climate financing for globally affordable clean cooking. *Nat. Clim. Change* **7**, 44–47 (2017) doi:10.1038/nclimate3141.
22. Gold Standard. *REDUCED EMISSIONS FROM COOKING AND HEATING – TECHNOLOGIES AND PRACTICES TO DISPLACE DECENTRALIZED THERMAL ENERGY CONSUMPTION (TPDDTEC)*. (2021).
23. The Gold Standard Foundation. *The Gold Standard Simplified Methodology for Clean and Efficient Cookstoves*. (2022).
24. Clean Development Mechanism. AMS-II.G.: Energy efficiency measures in thermal applications of non-renewable biomass — Version 12.0. (2020).
25. Clean Development Mechanism. AMS-I.E.: Switch from non-renewable biomass for thermal applications by the user — Version 12.0. (2021).
26. The Gold Standard Foundation. *Methodology for Metered & Measured Energy Cooking Devices*. (2021).
27. Krumpal, I. Determinants of social desirability bias in sensitive surveys: a literature review. *Qual. Quant.* **47**, 2025–2047 (2013) doi:10.1007/s11135-011-9640-9.
28. Presser, S. & Stinson, L. L. Data collection mode and social desirability bias in self-reported religious attendance: Church attendance in the United States. *Am. Sociol. Rev.* **63**, 137–145 (1998).
29. Epstein, W. M. Response bias in opinion polls and American social welfare. *Soc. Sci. J.* **43**, 99–110 (2006) doi:10.1016/j.soscij.2005.12.010.
30. Kar, A., Brauer, M., Bailis, R. & Zerriffi, H. The risk of survey bias in self-reports vs. actual consumption of clean cooking fuels. *World Dev. Perspect.* **18**, 100199 (2020) doi:10.1016/j.wdp.2020.100199.
31. Wilson, D. L. *et al.* Comparing Cookstove Usage Measured with Sensors Versus Cell Phone-Based Surveys in Darfur, Sudan BT - Technologies for Development. in (eds. Hostettler, S., Hazboun, E. & Bolay, J.-C.) 211–221 (Springer International Publishing, 2015).
32. Das, J., Hammer, J. & Sanchez-Paramo, C. *The Impact of Recall Periods on Reported Morbidity and Health Seeking Behavior*. <https://openknowledge.worldbank.org/bitstream/handle/10986/3541/WPS5778.pdf?sequence=1>.
33. Brooks, N. *et al.* How much do alternative cookstoves reduce biomass fuel use? Evidence from North India. *Resour. Energy Econ.* **43**, 153–171 (2016) doi:10.1016/j.reseneeco.2015.12.001.
34. Simons, A. M., Beltramo, T., Blalock, G. & Levine, D. I. Using unobtrusive sensors to measure and minimize Hawthorne effects: Evidence from cookstoves. *Spec. Issue Environ. Econ. Dev. Ctries.* **86**, 68–80 (2017) doi:10.1016/j.jeem.2017.05.007.
35. Shankar, A. V. *et al.* Everybody stacks: Lessons from household energy case studies to inform design principles for clean energy transitions. *ENERGY POLICY* **141**, 111468 (2020) doi:10.1016/j.enpol.2020.111468.



36. Jeuland, M., Soo, J.-S. T. & Shindell, D. The need for policies to reduce the costs of cleaner cooking in low income settings: Implications from systematic analysis of costs and benefits. *Energy Policy* **121**, 275–285 (2018) doi:<https://doi.org/10.1016/j.enpol.2018.06.031>.
37. Hanna, R., Duflo, E. & Greenstone, M. Up in smoke: The influence of household behavior on the long-run impact of improved cooking stoves. *Am. Econ. J. Econ. Policy* **8**, 80–114 (2016) doi:[10.1257/pol.20140008](https://doi.org/10.1257/pol.20140008).
38. Burwen, J. & Levine, D. I. A rapid assessment randomized-controlled trial of improved cookstoves in rural Ghana. *Energy Sustain. Dev.* **16**, 328–338 (2012) doi:[10.1016/J.ESD.2012.04.001](https://doi.org/10.1016/J.ESD.2012.04.001).
39. Beltramo, T., Blalock, G., Harrell, S., Levine, D. & Simons, A. M. The Effects of Fuel-Efficient Cookstoves on Fuel Use, Particulate Matter, and Cooking Practices: Results from a Randomized Trial in Rural Uganda. *UC Berkeley Cent. Eff. Glob. Action* (2019).
40. Rosa, G. *et al.* Assessing the Impact of Water Filters and Improved Cook Stoves on Drinking Water Quality and Household Air Pollution: A Randomised Controlled Trial in Rwanda. *PLOS ONE* **9**, e91011 (2014) doi:[10.1371/journal.pone.0091011](https://doi.org/10.1371/journal.pone.0091011).
41. Bensch, G. & Peters, J. The intensive margin of technology adoption – Experimental evidence on improved cooking stoves in rural Senegal. *J. Health Econ.* **42**, 44–63 (2015) doi:<https://doi.org/10.1016/j.jhealeco.2015.03.006>.
42. Ruiz-Mercado, I., Masera, O., Zamora, H. & Smith, K. R. Adoption and sustained use of improved cookstoves. *Energy Policy* **39**, 7557–7566 (2011).
43. Islam, M. M. *et al.* Assessing the Effects of Stove Use Patterns and Kitchen Chimneys on Indoor Air Quality during a Multiyear Cookstove Randomized Control Trial in Rural India. *Environ. Sci. Technol.* **56**, 8326–8337 (2022) doi:[10.1021/acs.est.1c07571](https://doi.org/10.1021/acs.est.1c07571).
44. García-Frapolli, E. *et al.* Beyond fuelwood savings: Valuing the economic benefits of introducing improved biomass cookstoves in the Purépecha region of Mexico. *Ecol. Econ.* **69**, 2598–2605 (2010) doi:[10.1016/j.ecolecon.2010.08.004](https://doi.org/10.1016/j.ecolecon.2010.08.004).
45. Agurto Adrianzén, M. Social Capital and Improved Stoves Usage Decisions in the Northern Peruvian Andes. *World Dev.* **54**, 1–17 (2014) doi:[10.1016/j.worlddev.2013.07.004](https://doi.org/10.1016/j.worlddev.2013.07.004).
46. CDM Methodologies Panel. *MP88: Meeting report/recommendations to the Executive Board*. <https://cdm.unfccc.int/Panels/meth/index.html> (2022).
47. Bailis, R. & Edwards, R. Kitchen Performance Test (KPT) Version 3.0. 1–32 (2007).
48. US EPA. Life Cycle Assessment of Cooking Fuel Systems in India, China, Kenya, and Ghana. [https://cfpub.epa.gov/si/si\\_public\\_record\\_report.cfm?dirEntryId=339679&Lab=NRMRL&simplesearch=0&showcriteria=2&sortby=pubDate&timstype=Published+Report&datebeginpublishedpresented](https://cfpub.epa.gov/si/si_public_record_report.cfm?dirEntryId=339679&Lab=NRMRL&simplesearch=0&showcriteria=2&sortby=pubDate&timstype=Published+Report&datebeginpublishedpresented).
49. Wathore, R., Mortimer, K. & Grieshop, A. P. In-Use Emissions and Estimated Impacts of Traditional, Natural- and Forced-Draft Cookstoves in Rural Malawi. *Environ. Sci. Technol.* **51**, 1929–1938 (2017) doi:[10.1021/acs.est.6b05557](https://doi.org/10.1021/acs.est.6b05557).
50. Stockwell, T. *et al.* Estimating under- and over-reporting of drinking in national surveys of alcohol consumption: identification of consistent biases across four English-speaking countries. *Addiction* **111**, 1203–1213 (2016) doi:[10.1111/add.13373](https://doi.org/10.1111/add.13373).
51. Ezzati, M., Martin, H., Skjold, S., Hoorn, S. V. & Murray, C. J. L. Trends in National and State-Level Obesity in the USA after Correction for Self-Report Bias: Analysis of Health Surveys. *J. R. Soc. Med.* **99**, 250–257 (2006) doi:[10.1177/014107680609900517](https://doi.org/10.1177/014107680609900517).
52. Harel, Y. *et al.* The effects of recall on estimating annual nonfatal injury rates for children and adolescents. *Am. J. Public Health* **84**, 599–605 (1994) doi:[10.2105/AJPH.84.4.599](https://doi.org/10.2105/AJPH.84.4.599).
53. Moreno-Serra, R., Anaya-Montes, M., León-Giraldo, S. & Bernal, O. Addressing recall bias in (post-)conflict data collection and analysis: lessons from a large-scale health survey in Colombia. *Confl. Health* **16**, 14 (2022) doi:[10.1186/s13031-022-00446-0](https://doi.org/10.1186/s13031-022-00446-0).
54. Foster, V., Tre, J.-P., Wodon, Q. & Bank, W. *Energy prices, energy efficiency, and fuel poverty 1*. (2000).
55. Clean Development Mechanism. *Concept Note CDM-MP85-A07. Analysis and options regarding caps used in AMS-I.E, AMS-II.G and TOOL30 Version 01.0*. (2013).
56. Daioglou, V., van Ruijven, B. J. & van Vuuren, D. P. Model projections for household energy use in developing countries. *7th Bienn. Int. Workshop "Advances Energy Stud.* **37**, 601–615 (2012) doi:[10.1016/j.energy.2011.10.044](https://doi.org/10.1016/j.energy.2011.10.044).
57. Clean Development Mechanism. Tool 30: Calculation of the fraction of non-renewable biomass (version 3). (2020).
58. Drigo, R. 'Wisdom' Woodfuel Integrated Supply/Demand Overview Mapping (WISDOM).
59. Ghilardi, A. *et al.* Spatiotemporal modeling of fuelwood environmental impacts: Towards improved accounting for non-renewable biomass. *Environ. Model. Softw.* **82**, 241–254 (2016) doi:[10.1016/j.envsoft.2016.04.023](https://doi.org/10.1016/j.envsoft.2016.04.023).
60. Cames, M. *et al.* *How additional is the Clean Development Mechanism? Analysis of the application of current tools and proposed alternatives. Study prepared for DG CLIMA*. (2016). doi:[10.13140/RG.2.2.23258.54728](https://doi.org/10.13140/RG.2.2.23258.54728).
61. Floess, E. *et al.* Scaling up gas and electric cooking in low- and middle-income countries: climate threat or mitigation strategy with co-benefits? *Environ. Res. Lett.* **18**, 034010 (2023) doi:[10.1088/1748-9326/acb501](https://doi.org/10.1088/1748-9326/acb501).
62. Huang, Y. *et al.* Global radiative effects of solid fuel cookstove aerosol emissions. *Atmospheric Chem. Phys.* **18**, 5219–5233 (2018).
63. Kodros, J. K. *et al.* Uncertainties in global aerosols and climate effect due to biofuel emissions. *Atmospheric Chem. Phys.* **15**, 8577–8596 (2015).

64. Gomez, D. *et al.* 2006 IPCC Guidelines for National Greenhouse Gas Inventories: Chapter 2: Stationary Combustion. <https://www.ipcc-nggip.iges.or.jp/public/2006gl/vol2.html> (2008).
65. Bailis, R. *et al.* Enhancing clean cooking options in peri-urban Kenya: A pilot study of advanced gasifier stove adoption. *Environ. Res. Lett.* **15**, (2020) doi:10.1088/1748-9326/ab865a.
66. Masera, O. R., Saatkamp, B. D. & Kammen, D. M. *From Linear Fuel Switching to Multiple Cooking Strategies: A Critique and Alternative to the Energy Ladder Model.* *World Development* vol. 28 2083–2103 (Pergamon, 2000). doi:10.1016/S0305-750X(00)00076-0.
67. World Bank Group. *Forum on Natural Capital Accounting for Better Policy Decisions: Taking Stock and Moving Forward (English).* <http://documents.worldbank.org/curated/en/904211580129561872/Forum-on-Natural-Capital-Accounting-for-Better-Policy-Decisions-Taking-Stock-and-Moving-Forward>.
68. Dufournaud, C. M., Quinn, J. T. & Harrington, J. J. A Partial Equilibrium Analysis of the Impact of Introducing More-Efficient Wood-Burning Stoves into Households in the Sahelian Region. *Environ. Plan. Econ. Space* **26**, 407–414 (1994) doi:10.1068/a260407.
69. Lambe, F. *et al.* Opening the black pot: A service design-driven approach to understanding the use of cleaner cookstoves in peri-urban Kenya. *Energy Res. Soc. Sci.* **70**, 101754 (2020) doi:10.1016/j.erss.2020.101754.
70. Haya, B. *et al.* Comprehensive Review of Carbon Quantification by Improved Forest Management Offset Protocols. *Front. For. Glob. Change* doi:10.3389/ffgc.2023.958879.
71. Badgley, G. *et al.* Systematic over-crediting in California's forest carbon offsets program. *Glob. Change Biol.* **28**, 1433–1445 (2022) doi:10.1111/gcb.15943.
72. Haya, B. K. Carbon Offsetting: An Efficient Way to Reduce Emissions or to Avoid Reducing Emissions? An Investigation and Analysis of Offsetting Design and Practice in India and China. *ProQuest Dissertations and Theses* (University of California, Berkeley, 2010).
73. West, T. A. P., Börner, J., Sills, E. O. & Kontoleon, A. Overstated carbon emission reductions from voluntary REDD+ projects in the Brazilian Amazon. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 24188–24194 (2020) doi:10.1073/pnas.2004334117.
74. Gill-Wiehl, A. & Kammen, D. M. A pro-health cookstove strategy to advance energy, social and ecological justice. *Nat. Energy* (2022) doi:10.1038/s41560-022-01126-2.
75. Climate Focus. *The Handbook for Programme of Activities: Practical Guidance to Successful Implementation.* <https://climatefocus.com/publications/handbook-programmes-activities-practical-guidance-successful-implementation/> (2011).
76. Rindfleisch, A., Malter, A. J., Ganesan, S. & Moorman, C. Cross-Sectional versus Longitudinal Survey Research: Concepts, Findings, and Guidelines. *J. Mark. Res.* **45**, 261–279 (2008) doi:10.1509/jmkr.45.3.261.
77. Cohen, Louis, Manion, Lawrence, & Morrison, Keith. Surveys, longitudinal, cross-sectional and trend studies. in *Research Methods in Education* 27 (Routledge).
78. Ventrella, J., Lefebvre, O. & MacCarty, N. Techno-economic comparison of the FUEL sensor and Kitchen Performance Test to quantify household fuel consumption with multiple cookstoves and fuels. *Dev. Eng.* **5**, 100047 (2020) doi:10.1016/j.deveng.2020.100047.
79. Pickering Amy J., Arnold Benjamin F., Dentz Holly N., Colford John M., & Null Clair. Climate and Health Co-Benefits in Low-Income Countries: A Case Study of Carbon Financed Water Filters in Kenya and a Call for Independent Monitoring. *Environ. Health Perspect.* **125**, 278–283 (2017) doi:10.1289/EHP342.
80. Pillarisetti, A. *et al.* Repeated assessment of PM<sub>2.5</sub> in Guatemalan kitchens cooking with wood: Implications for measurement strategies. *Atmos. Environ.* **295**, 119533 (2023) doi:10.1016/j.atmosenv.2022.119533.
81. Pillarisetti, A. *et al.* Patterns of stove usage after introduction of an advanced cookstove: The long-term application of household sensors. *Environ. Sci. Technol.* **48**, 14525–14533 (2014) doi:10.1021/es504624c.
82. Gill-Wiehl, A., Sievers, S. & Kammen, D. M. The value of community technology workers for LPG use: A pilot in Shirati, Tanzania. *Energy Sustain. Soc.* **12**, 1–16 (2022) doi:10.1186/s13705-022-00331-x.
83. Jeuland, M., Soo, J.-S. T. & Shindell, D. The need for policies to reduce the costs of cleaner cooking in low income settings: Implications from systematic analysis of costs and benefits. *Energy Policy* **121**, 275–285 (2018) doi:https://doi.org/10.1016/j.enpol.2018.06.031.
84. Shankar, A. V. *et al.* Everybody stacks: Lessons from household energy case studies to inform design principles for clean energy transitions. *ENERGY POLICY* **141**, 111468 (2020) doi:10.1016/j.enpol.2020.111468.
85. Barnes, B. R. The Hawthorne Effect in community trials in developing countries. *Int. J. Soc. Res. Methodol.* **13**, 357–370 (2010) doi:10.1080/13645570903269096.
86. Thomas, E. A. *et al.* Behavioral Reactivity Associated With Electronic Monitoring of Environmental Health Interventions—A Cluster Randomized Trial with Water Filters and Cookstoves. *Environ. Sci. Technol.* **50**, 3773–3780 (2016) doi:10.1021/acs.est.6b00161.
87. Harrell, S. *et al.* What is a “meal”? Comparative methods of auditing carbon offset compliance for fuel-efficient cookstoves. *Ecol. Econ.* **128**, 8–16 (2016) doi:10.1016/j.ecolecon.2016.03.014.
88. Ventrella, J. & MacCarty, N. Monitoring impacts of clean cookstoves and fuels with the Fuel Use Electronic Logger (FUEL): Results of pilot testing. *Energy Sustain. Dev.* **52**, 82–95 (2019) doi:10.1016/j.esd.2019.06.004.

89. Miller, H., Shrestha, J., Lefebvre, O. & MacCarty, N. Use of an integrated suite of sensors to simultaneously monitor fuel consumption, air quality, and adoption provides important insights and validates impact metrics for household stoves. *Dev. Eng.* **7**, 100099 (2022) doi:10.1016/j.deveng.2022.100099.
90. Puzzolo, E., Pope, D., Stanistreet, D., Rehfuess, E. A. & Bruce, N. G. *Clean fuels for resource-poor settings: A systematic review of barriers and enablers to adoption and sustained use. Environmental Research* vol. 146 218–234 (2016).
91. Rehfuess, E. A., Puzzolo, E., Stanistreet, D., Pope, D. & Bruce, N. G. Enablers and barriers to large-scale uptake of improved solid fuel stoves: A systematic review. *Environ. Health Perspect.* **122**, 120–130 (2014) doi:10.1289/ehp.1306639.
92. Harrell, B. S. *et al.* Incentivizing Elimination of Biomass Cooking Fuels with a Reversible Commitment and a Spare LPG Cylinder. *Environ. Sci. Technol.* (2020) doi:10.1021/acs.est.0c01818.
93. Tourangeau, R. & Yan, T. Sensitive questions in surveys. *Psychol. Bull.* **133**, 859–883 (2007) doi:10.1037/0033-2909.133.5.859.
94. Lupu, N. & Michelitch, K. Advances in Survey Methods for the Developing World. *Annu. Rev. Polit. Sci.* **21**, 195–214 (2018) doi:10.1146/annurev-polisci-052115-021432.
95. Harrell, B. S. *et al.* Incentivizing Elimination of Biomass Cooking Fuels with a Reversible Commitment and a Spare LPG Cylinder. *Environ. Sci. Technol.* (2020) doi:10.1021/acs.est.0c01818.
96. Muralidharan, V. *et al.* Field testing of alternative cookstove performance in a rural setting of western India. *Int. J. Environ. Res. Public Health* **12**, 1773–1787 (2015).
97. Gill-Wiehl, A., Price, T. & Kammen, D. M. What's in a stove? A review of the user preferences in improved stove designs. *Energy Res. Soc. Sci.* **81**, 102281 (2021) doi:10.1016/j.erss.2021.102281.
98. Zhang, Z. *et al.* Systematic and conceptual errors in standards and protocols for thermal performance of biomass stoves. *Renew. Sustain. Energy Rev.* **72**, 1343–1354 (2017) doi:10.1016/j.rser.2016.10.037.
99. Piedrahita, R. *et al.* Assessment of cookstove stacking in Northern Ghana using surveys and stove use monitors. *Energy Sustain. Dev.* **34**, 67–76 (2016) doi:https://doi.org/10.1016/j.esd.2016.07.007.
100. Kenya Integrated Household Budget Survey 2015-2016. <https://statistics.knbs.or.ke/nada/index.php/catalog/13/related-materials> (2016).
101. L'Orange, C., DeFoort, M. & Willson, B. Influence of testing parameters on biomass stove performance and development of an improved testing protocol. *Energy Sustain. Dev.* **16**, 3–12 (2012) doi:10.1016/j.esd.2011.10.008.
102. FAO. Global Forest Resources Assessment 2010.
103. Johnson, M., Edwards, R. & Masera, O. Improved stove programs need robust methods to estimate carbon offsets. *Clim. Change* **102**, 641–649 (2010) doi:10.1007/s10584-010-9802-0.
104. Grieshop, A. P., Marshall, J. D. & Kandlikar, M. Health and climate benefits of cookstove replacement options. *Clean Cook. Fuels Technol. Dev. Econ.* **39**, 7530–7542 (2011) doi:10.1016/j.enpol.2011.03.024.
105. Akagi, S. K. *et al.* Emission factors for open and domestic biomass burning for use in atmospheric models. *Atmospheric Chem. Phys.* **11**, 4039–4072 (2011) doi:10.5194/acp-11-4039-2011.
106. Argonne National Laboratory. The Greenhouse Gases, Regulated Emissions, and Energy Use in Transportation (GREET) Model. (2021).
107. Johnson, M., Edwards, R., Alatorre Frenk, C. & Masera, O. In-field greenhouse gas emissions from cookstoves in rural Mexican households. *Atmos. Environ.* **42**, 1206–1222 (2008) doi:10.1016/j.atmosenv.2007.10.034.
108. Grieshop, A. P., Reynolds, C. C. O., Kandlikar, M. & Dowlatabadi, H. A black-carbon mitigation wedge. *Nat. Geosci.* **2**, 533–534 (2009) doi:10.1038/ngeo595.
109. Pachauri, S., Poblete-Cazenave, M., Aktas, A. & Gidden, M. J. Access to clean cooking services in energy and emission scenarios after COVID-19. *Nat. Energy* **6**, 1067–1076 (2021) doi:10.1038/s41560-021-00911-9.
110. Poblete-Cazenave, M., Pachauri, S., Byers, E., Mastrucci, A. & van Ruijven, B. Global scenarios of household access to modern energy services under climate mitigation policy. *Nat. Energy* **6**, 824–833 (2021) doi:10.1038/s41560-021-00871-0.
111. SCHNEIDER, L. Assessing the additionality of CDM projects: practical experiences and lessons learned. *Clim. Policy* **9**, 242–254 (2009) doi:10.3763/cpol.2008.0533.
112. Purdon, M. Opening the Black Box of Carbon Finance 'Additionality': The Political Economy of Carbon Finance Effectiveness across Tanzania, Uganda, and Moldova. *World Dev.* **74**, 462–478 (2015) doi:10.1016/j.worlddev.2015.05.024.
113. Stoner, O. *et al.* Household cooking fuel estimates at global and country level for 1990 to 2030. *Nat. Commun.* **12**, 5793 (2021) doi:10.1038/s41467-021-26036-x.
114. Simon, G. L., Bumpus, A. G. & Mann, P. Win-win scenarios at the climate-development interface: Challenges and opportunities for stove replacement programs through carbon finance. *Glob. Environ. Change* **22**, 275–287 (2012) doi:10.1016/j.gloenvcha.2011.08.007.
115. Calyx Global. <https://www.calyxglobal.com/>.
116. Singh, D., Zerriffi, H., Bailis, R. & LeMay, V. Forest, farms and fuelwood: Measuring changes in fuelwood collection and consumption behavior from a clean cooking intervention. *Energy Sustain. Dev.* **61**, 196–205 (2021) doi:10.1016/j.esd.2021.02.002.

117. Fitzpatrick, D. Chapter 10 - Monte Carlo Analysis. in *Analog Design and Simulation Using OrCAD Capture and PSpice (Second Edition)* (ed. Fitzpatrick, D.) 151–164 (Newnes, 2018). doi:10.1016/B978-0-08-102505-5.00010-0.

118. Robinson, B. L., Clifford, M. J., Hewitt, J. & Jewitt, S. Cooking for communities, children and cows: Lessons learned from institutional cookstoves in Nepal. *Energy Sustain. Dev.* **66**, 1–11 (2022) doi:10.1016/j.esd.2021.10.012.

## Boxes

Boxes 1 and 2 are available in the Supplementary Files section.

## Figures

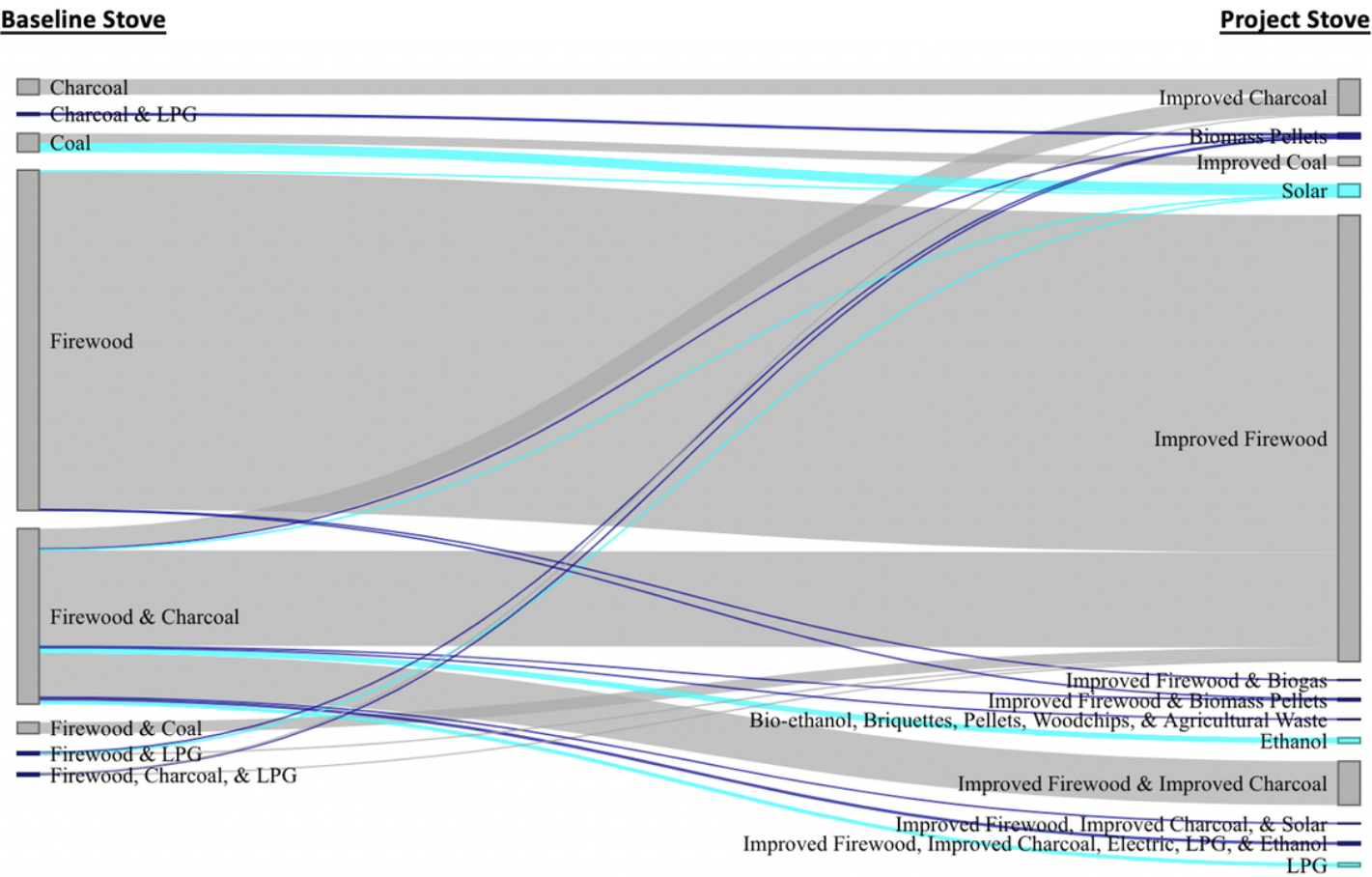
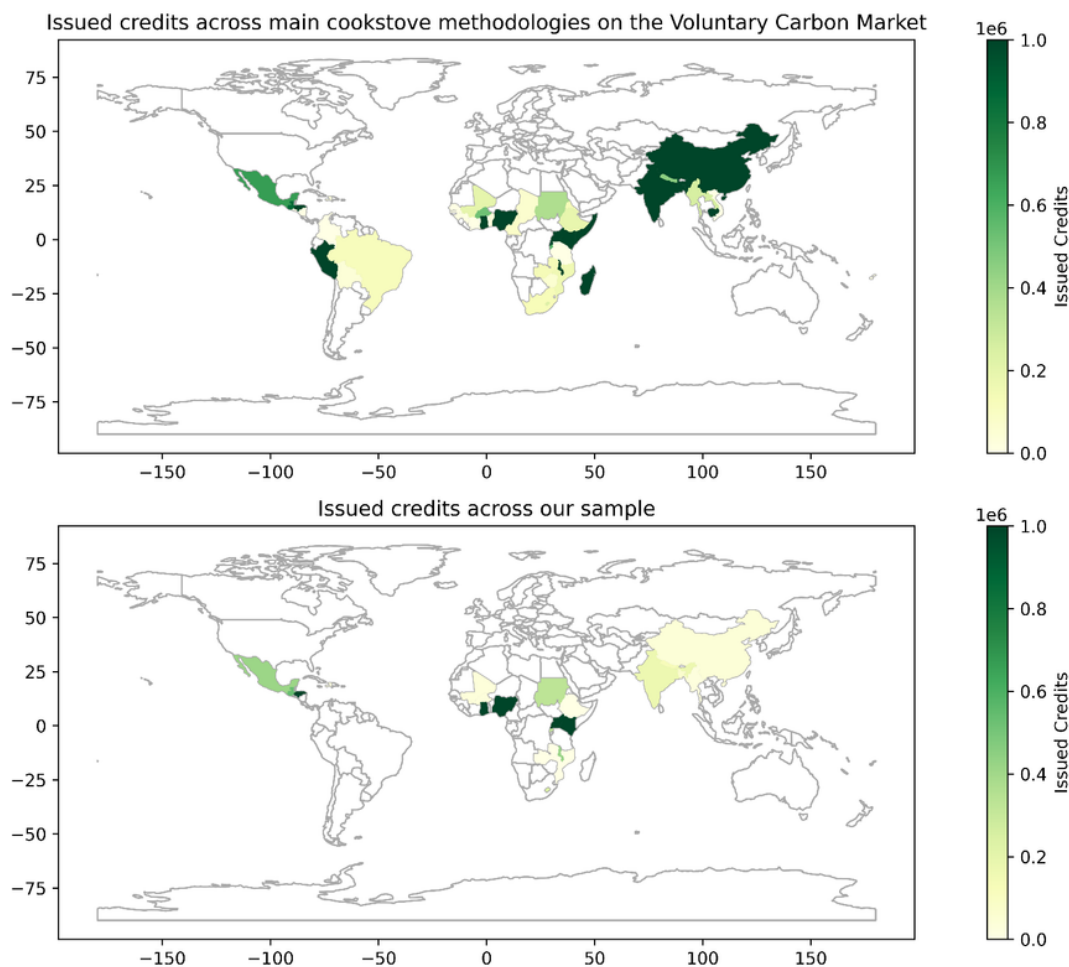


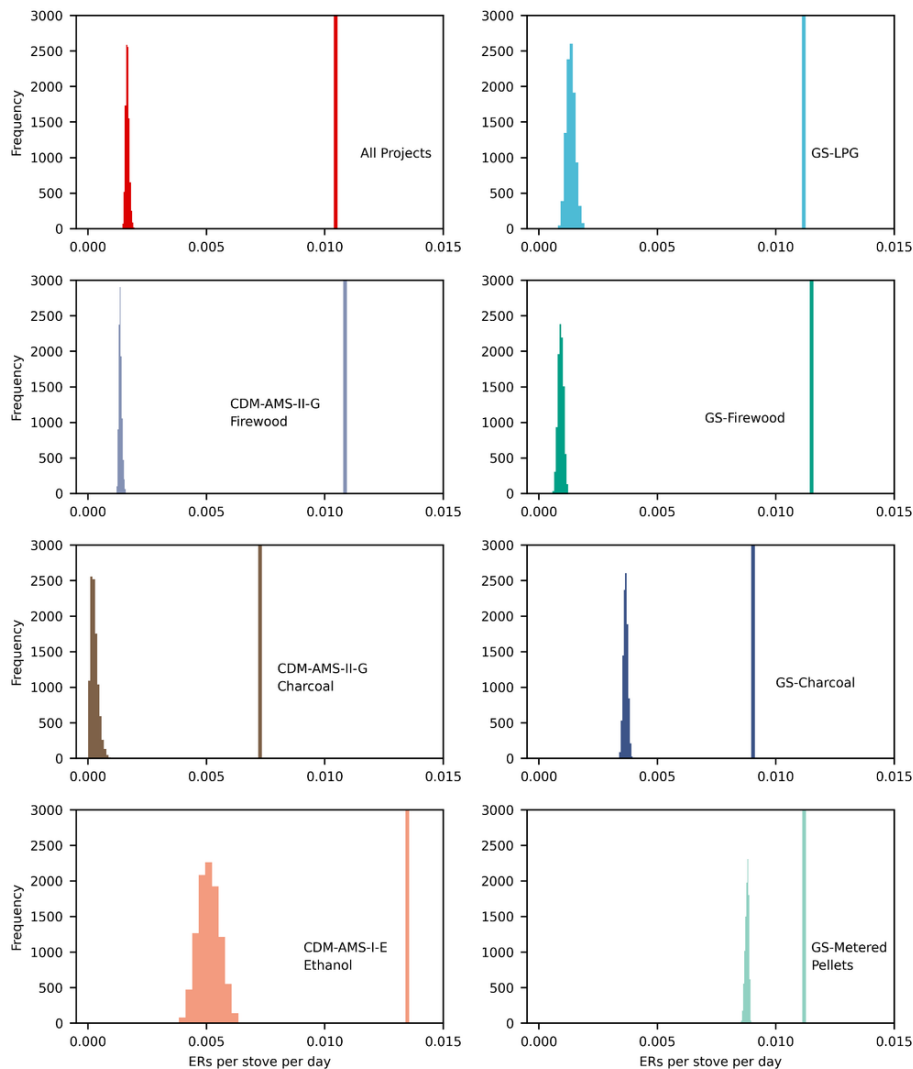
Figure 1

As of November 9th, 2022, 4% of cookstove project activities (43 out of 992 projects) registered on the Voluntary Carbon Market (VCM) distribute only cooking fuels or stoves that meet the WHO’s definition of clean, that is, they meet Tier 5 for carbon monoxide and Tier 4 for particulate matter. The WHO defines stoves that meet Tier 3 as transitional, and Tiers 0-2 as polluting. *The left side of the diagram indicates the majority baseline fuels before intervention, and the right side represents the project fuel/stove that the VCM funded project implemented. The width of the link indicates the number of projects. Grey indicates WHO polluting or transitional fuels or stoves (Tier 0-3). Dark blue indicates a mix of WHO clean, transitional, or polluting fuels and stoves, while cyan indicates only WHO clean fuels or stoves. We exclude 6 projects that do not change the stove, but only replaced firewood with agricultural waste.*



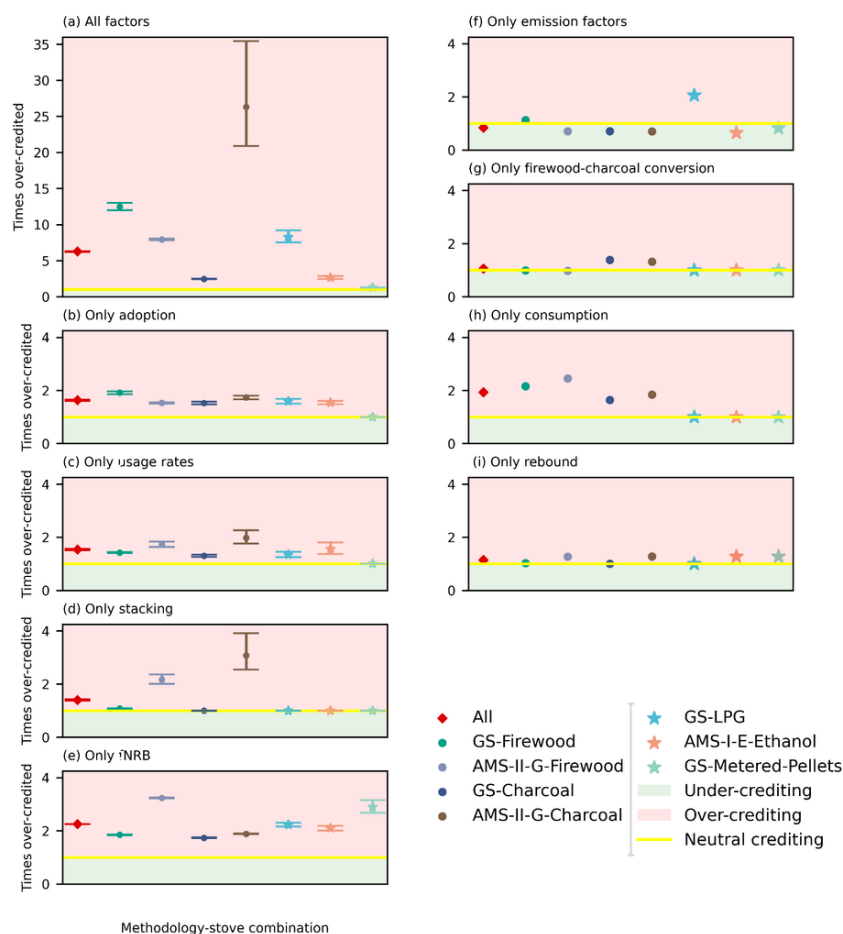
**Figure 2**

*Credits issued to date on the Voluntary Carbon Market across the five methodologies covered as of November 9th, 2022 (top panel) and from the 36 cookstove project activities in our sample (bottom panel). We cover the Gold Standard's (GS) Technologies and Practices to Displace Decentralized Thermal Energy Consumption (GS-TPDDTEC), GS's Simplified Methodology for Clean and Efficient Cookstoves, Clean Development Mechanism's (CDM) Energy efficiency measures in thermal applications of non-renewable biomass (AMS-II-G), CDM's Switch from non-renewable biomass for thermal applications by the user (AMS-I-E), and GS Methodology for Metered & Measured Energy Cooking Devices (GS-Metered).*



**Figure 3**

*Average results of our over/under crediting analysis of our sample of 36 cookstove projects in total and broken down by 7 methodology-stove combinations: CDM-AMS-II-G-Firewood, CDM-AMS-II-G-Charcoal, CDM-AMS-I-E-Ethanol, GS-LPG, GS-Firewood, GS-Charcoal, and GS-Metered-Pellets. We consider the following factors: adoption, usage, and stacking rates, baseline and project consumption, the fraction of non-renewable biomass, emissions factors (point of use, non-carbon dioxide emissions, methane stoichiometry, and upstream emissions), firewood to charcoal ratio, and rebound effect. We utilize a Monte Carlo Analysis Framework to construct a range of the actual emission reductions (ERs) on a per stove-day basis (the histogram) compared to the average ERs on a per stove-day basis claimed by the projects (the solid line).*



**Figure 4**

(a) The amount of over/under crediting after quantifying all factors and (b-i) individual factors by methodology-stove combinations. The points indicate the average over- or under-crediting, while the error bars refer to the 95% confidence interval from the Monte Carlo Analysis Framework applied to adoption, usage, and stacking rates and fraction of non-renewable biomass (fNRB) (see Methods). Emission factors include point of use emissions including non-CO<sub>2</sub> emissions and upstream emissions. Less than 1 (green shading) indicates under-crediting. Red shading indicates over-crediting; Yellow indicates accurate crediting.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [22023GillWiehletal.SupplementalMaterials.docx](#)
- [Boxes.docx](#)