Assessment of cookstove stacking in Northern Ghana using surveys and stove use monitors

Ricardo Piedrahita a,⁎, Katherine L. Dickinson e,f, Ernest Kanyomse d, Evan Coffey a, Rex Alirigia d, Yolanda Hagar b, Isaac Rivera c, Abraham Oduro d, Vanja Dukić b, Christine Wiedinmyer e, Michael Hannigan a

a University of Colorado Boulder, Department of Mechanical Engineering, United States
b University of Colorado Boulder, Applied Mathematics Department, United States
c University of Colorado Boulder, Geography Department, United States
d Navrongo Health Research Centre, Ghana
e National Center for Atmospheric Science, United States
f University of Colorado Boulder, United States

A R T I C L E I N F O

Article history:
Received 17 March 2016
Revised 28 July 2016
Accepted 28 July 2016
Available online xxxx

A B S T R A C T

Biomass burning for home energy use is a major health and environmental concern. While transitioning to cleaner cooking technologies has the potential to generate significant health and environmental benefits, prior efforts to introduce improved cookstoves have encountered many hurdles. Here, we focus on the increased stove use hurdle; households tend to use improved stoves alongside their traditional stoves rather than replacing them entirely, a phenomenon called cookstove “stacking.” This work provides a systematic, multi-method assessment of households’ cooking behaviors and cookstove stacking in the context of a 200-home randomized cookstove intervention study in Northern Ghana. Two stoves were selected for the intervention, a locally made rocket stove (Gyapa) and the Philips HD4012 LS gasifier stove. There were four intervention groups: a control group, a group given two Gyapa stoves, a group given two Philips stoves, and a group given one of each. Two stoves were distributed to each home in an attempt to induce more substitution away from traditional stoves. Adoption and usage patterns were quantified using temperature loggers at a subset of homes, as well as quarterly surveying in all households. We find that using multiple stoves each day is common practice within each intervention group, and that the two groups given at least one Gyapa had the largest reductions in traditional stove use relative to the control group, though use of traditional stoves remained high in all groups.

© 2016 International Energy Initiative. Published by Elsevier Inc. All rights reserved.

Background

Motivation

2.8 billion people burn solid fuels for cooking (Bonjour et al., 2013) and the resulting air pollution is the third leading risk factor for the global burden of disease, contributing to 4 million premature deaths per year (Lim et al., 2012). The environmental impacts from this activity are substantial. In addition to contributing to regional deforestation and forest degradation (Chidumayo and Cumbo, 2013), residential combustion (including wood, agricultural waste, animal waste, and coal) contributes an estimated 32% of particulate black carbon, and 64% of particulate organic carbon to global non-open burning emissions (Bond et al., 2013).

To address these issues, cookstove distribution programs and studies to replace traditional cooking methods with cleaner, more efficient ones continue to grow in scope and magnitude. Measurement of cookstove adoption is critical in determining the feasibility and likelihood of success of these programs. There are many factors involved in the decision to adopt a new stove, among them income, education, availability of viable clean cookstoves, fuel availability, financing, location, and cultural norms (Barnes et al., 1993; Pine et al., 2011; Ruiz-Mercado et al., 2011; Jan, 2012; Jeuland and Pattanayak, 2012; Lewis and Pattanayak, 2012; Rehfues et al., 2013; Malla and Timilsina, 2014). Previous studies have found evidence that even when intervention cookstoves are used regularly, households often maintain regular use of their traditional stoves, a practice known as stove stacking (Pillarisetti et al., 2014; Stanistreet et al., 2015).

Research on Emissions, Air quality, Climate, and Cooking Technologies in Northern Ghana (REACCTING) (Dickinson et al., 2015) is a 200-home cookstove intervention study in the Kassena-Nankana (K-N) Districts of Northern Ghana, designed to learn about cooking...
behaviors and their impacts in this region. Past personal air pollution exposure studies in Ghana have measured worryingly high levels of CO (Burwen and Levine, 2012) and PM (Arku et al., 2008; Rooney et al., 2012; Van Vliet et al., 2013) due to cooking and other combustion sources. The ecological motivation is also strong, as the study is located within a climatically sensitive region at high risk of drought and forced migration (Warner, 2009; Antwi-Agyei et al., 2012). Ghana as a whole is experiencing alarming deforestation rates, with 33.7% of forest area (2.5E6 ha) lost since 1990, and a 2.19% annual deforestation rate from 2005 to 2010 (FAO, 2010). The Upper East region, encompassing our study area, is almost entirely categorized as a high-risk region for desertification (Adanu et al., 2013). Assessment of adoption and stacking has not been undertaken in this region of Africa, where the mix of remoteness and indoor/outdoor cooking offers new challenges.

Measuring Stove Use

In REACTING, we determine the extent of stove use and stove stacking using two methods, stove usage monitoring with temperature data loggers (here referred to as stove usage monitors, or SUMs), and quantitative surveying. Both types of data have strengths and limitations. Stove usage monitoring allows identification of cooking events from extended time series of stove temperature (Ruiz-Mercado et al., 2012; Mukhopadhyay et al., 2012; Graham et al., 2014). SUMs have the advantage of eliminating the biases associated with self-reporting that have been observed in some studies (Thomas et al., 2013; Wilson et al., 2015). However, other sources of bias and measurement error are still possible with SUMs, including reactivity effects (higher use due to the knowledge of being monitored – see Thomas et al., 2016). Considerable uncertainty also remains in detecting cooking events using SUMs data, particularly for the traditional 3-stone fires (TSFs). In addition, SUMs data collection is costly. As a result, we were only able to collect SUMs data for a subset of study households rather than the entire sample. Meanwhile, surveys were conducted in all households at multiple discrete time points (quarterly), and provide us with detailed contextual information along with (potentially mis-reported) stove use information. Survey information such as foods cooked and fuel types used with each stove shed light on how and why certain stoves are being used by different households. Used in combination, survey and SUMs data can more effectively inform future cookstove and fuel improvement efforts in the region.

Our study makes an important contribution to the literature by examining cooking behaviors in a region that has received relatively little attention: Northern Ghana. In addition, this work is among the first to publish results on the use of multiple intervention stoves alongside traditional stoves. Previously, Loo et al. (2016) performed a study in Kenya assessing user perspectives on six different improved combustion stoves (ICSs) rotated through homes for two-week periods.

Methods

Study population and design

The REACTING study ran from November 2013 to January 2016. The study population consisted of households in the K-N Districts that 1) were classified as rural, 2) used biofuels as their main cooking fuel source, 3) had at least one woman of childbearing age (18–55) and one child under five, 4) used a borehole as their primary water source, and 5) did not have electricity in the home. Using data from the district-wide Health and Demographic Surveillance Survey (HDSS) (Oduro et al. 2012), we identified the sample frame of households that met these eligibility criteria, and then used a cluster random sampling method to select 200 households for inclusion in the study. Detailed information on study design and sample selection is presented in Dickinson et al. (2015).

A baseline survey conducted in all 200 households prior to the stove intervention provided detailed information about local cooking practices that confirmed observations the study team made during the 2 years prior to the start of the study, and which informed the design of the REACTING intervention. Even before the introduction of any new stoves, households in this area were cooking with multiple stoves, and with a mix of cooking technologies (Fig. 1). The most common cooking technology in this area is a traditional wood-fired 3-stone stove, but the majority of households (70%) owned at least one charcoal stove, locally known as a “coal pot,” as well. Only 10% of households relied on a single stove to meet their cooking needs; 38% of households had two stoves at baseline, and the remaining 53% had three or more stoves.

Further, we observed that local cooking practices link stove and fuel types to specific foods. The dishes that are commonly prepared and eaten in this area determine households’ stove needs; these staple dishes and associated cooking methods are listed in Table 1. The items

![Fig. 1. Baseline (pre-intervention) cookstove technology mix among study households.](image-url)
on this list can be broken into two groups according to their cooking method or requirements. Dishes that require rapid boiling (such as soup or rice) fall into one category, while dishes requiring more intensive effort such as vigorous stirring (Tuo Zaaf, or TZ for short) are in the second. The baseline survey, conducted from November–December 2013, indicated that by far the most common dishes in this area are TZ and vegetable soup, which are often consumed together, and the next most common dish is rice. We also observed fairly consistent patterns in the types of stoves used to cook each dish. The TSF is the most commonly used stove for all types of dishes. However, a sizeable portion of vegetable soup meals are cooked over charcoal stoves, and these are also used to cook rice on some occasions. Meanwhile, TZ is cooked almost exclusively on TSFs.

Given local cooking practices and previous observations of stove stacking behavior in other contexts, the REACTING study intervention was designed to distribute two stoves to each household randomized into an intervention group in order to create greater potential for households to substitute away from their traditional stoves while continuing to meet their cooking needs. Furthermore, we selected two different stove technologies for our intervention based on two key considerations. First, as discussed in Dickinson et al. (2015), we hoped to contribute to an ongoing debate among cookstove researchers and policymakers over whether movements up the stove technology ladder should be made incrementally (i.e., starting with locally made, affordable, low-tech stoves) or transformationally (moving directly to the cleanest technologies available). Thus, we decided to compare adoption and performance of a locally made ceramic and metal rocket stove (Gyapa) alongside the theoretically cleaner Philips stove. Philips is more expensive, and of the forced draft design, requiring battery charging with the provided solar panel every few days. Second, we suspected that households might use these two technologies differently, with each being suited to meeting different cooking needs. In particular, in piloting several types of cookstoves in the study area, households expressed doubts about being able to cook TZ on some of the stove models given the need for vigorous stirring. These concerns informed the design of the Gyapa stove, which was developed specifically for this project (though it bears resemblance to the one used in a study from Accra by Pennise et al. (2009)), as well as the design of a rebar stand to increase the stability of the Philips stove (Fig. 2).

In November of 2013, the households were randomly placed into one of four intervention groups: one with two Gyapa stoves, one with two Philips stoves, one with each of those stoves, and a control group (no new stoves until the conclusion of the two year study). During stove distribution assemblies, retired nurses working with the project team educated participants on the health, timesaving, and financial and environmental benefits of using the improved stoves, and those receiving new stoves were encouraged to stop using their traditional stoves.

**SUM methods**

SUMs were initially deployed in January/February of 2013, on 103 stoves distributed over 45 households. They were placed on the improved cookstoves and the most-used traditional cookstove at 10–12 households from each study group. More SUMs were added in late

<table>
<thead>
<tr>
<th>Dish</th>
<th>Description</th>
<th>Cooking Method</th>
<th>% Of Households Cooking Dish</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tuo Zaaf (TZ)</td>
<td>Thick porridge made with millet or maize flour, often served with vegetable soup</td>
<td>A mixture of water and millet flour is added to boiling water. Cold water is then added, and after the mixture is heated, half is moved to another bowl. More millet flour is then added, all the while stirring vigorously. The separated portion is added back in, and it is served hot.</td>
<td>57.0% 55.5% 1.5%</td>
</tr>
<tr>
<td>Vegetable soup</td>
<td>Soup made with bennef-kanzaga, vio, alefu, yambola, okro, all local vegetables. Fish or meat is sometimes added depending on availability.</td>
<td>Cut vegetables are boiled 10–15 min. Groundnut paste or bean flour is added, sometimes with fish or meat. Cooked until it boils again.</td>
<td>56.5% 43.0% 19.0%</td>
</tr>
<tr>
<td>Rice</td>
<td>Multiple varieties are available and are used to make jollof rice, rice balls, or plain rice, often with a sauce.</td>
<td>Cooked in metal or earthenware pots</td>
<td>39.0% 33.0% 6.5%</td>
</tr>
<tr>
<td>Beans</td>
<td>Many varieties are commonly eaten in the K-N districts, in various dishes</td>
<td>Varies</td>
<td>6.0% 6.0% 0.0%</td>
</tr>
<tr>
<td>Other dishes</td>
<td>Pompuka (3.0%), Bombara beans (1.5%), Tubani (1.5%), Tubers (1.5%), Corn (1.5%), Porridge (1.5%), Banku (0.5%)</td>
<td>Varies</td>
<td>10.5% 7.5% 3.5%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dish Description</th>
<th>Cooking Method</th>
<th>% Of Households Cooking Dish</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vegetable soup</td>
<td></td>
<td>56.5% 43.0% 19.0%</td>
</tr>
<tr>
<td>Rice</td>
<td></td>
<td>39.0% 33.0% 6.5%</td>
</tr>
<tr>
<td>Beans</td>
<td></td>
<td>6.0% 6.0% 0.0%</td>
</tr>
<tr>
<td>Other dishes</td>
<td></td>
<td>10.5% 7.5% 3.5%</td>
</tr>
</tbody>
</table>

Data on % of households cooking each dish are from REACTING baseline survey questions that asked about use of each stove in the household on the day prior to the survey. Households may have cooked a dish multiple times on that day using different types of stoves, so that the 3-stone (TSF) and charcoal columns do not necessarily equal the “total” column.

Fig. 2. Digit-TL SUM placements on the Gyapa, Philips, and 3-stone fires, from left to right. The Philips stove is shown with the specially designed rebar pot stand. The 3-stone fire at right is shown with wood and millet stalks for fuel.
2014 and early 2015 to replace broken ones and to monitor more stoves in each home, as most homes had 3 or more stoves at baseline (Fig. 1). This paper presents the first year of SUM and survey data, ranging from January 2014 – January 2015. The SUMs deployment time series is shown Fig. 3, organized by the number of households monitored in each intervention group on the left panels, and the number of each type of stove monitored in each intervention group in the right panels. A one-month ramp-up time is apparent as SUMs were deployed, with a reduction in data coverage in the second half of the year in most groups, due to lost or damaged SUMs, or mistakes in SUM data management.

Labjack Digit-TL temperature, light, and humidity monitors (LabJack, Lakewood CO) were used as SUMs. Humidity logging was disabled in favor of using a waterproof enclosure for the SUMs, while light logging was enabled on a subset of SUMs. The US$35 Digit-TLs measure temperature from $-40$ to 85 °C, with 0.5 °C resolution and reported uncertainty of 1.0 °C. Battery life is specified at over 3 years, with onboard memory capacity of 500,000 records. This allowed field workers to visit the homes every 4–6 months, at the selected 1-min sampling rate. However, these SUM features also resulted in lost data because deployment errors or broken SUMs were not identified until the next scheduled visit, often months away.

The SUMs were placed strategically on the stoves to avoid surpassing the SUM temperature limit of 85 °C. SUMs placement on each stove type was tested prior to deployment for the Philips and Gyapa stoves, but were placed conservatively to ensure they were not in danger of over-heating. This resulted in more challenging data analysis, as the peaks were less pronounced than in some other studies. On the Gyapa stove, SUMs were affixed with a metal strap near the center of the stove, while on the Philips they were attached with a hose clamp, which was screwed into one of the existing screw holes near the top of the stove. On the TSFs, the SUMs were wedged behind the largest stone in the fire with metal stakes. We were not able to put them underneath a stone as done in other studies, because the residents use plaster finishes in cooking areas that would have been damaged in the process of installing our relatively large SUMs. Typical placements are shown in Fig. 2.

Cooking events were identified using a modified version of the algorithm described in Ruiz-Mercado et al. (2012), here referred to as the RMM algorithm. The 1-min data from the SUMs were first smoothed using a 2nd order polynomial with a 10-min window. High variability remained from sample to sample even after this step, so the algorithm was modified to identify the start and end of a cooking event within 90 min of its peak to ensure cooking events were not mistakenly prolonged. This value was selected based on uncontrolled cooking tests performed concurrently during the study, which had a mean cooking time of 92.5 min, and maximum

---

Fig. 3. Digit-TL SUM deployment time series for each stove intervention group. The left panels show the number of households monitored for each stove type, while the right panels show the number of stoves monitored for each stove type. Note the difference in y-axis scales from left to right panels, as households often have multiple stoves monitored of the same type.
cooking time of 160 min (60 tests total). The thresholds required in the RMM algorithm were selected based on visual analysis of performance. The slope thresholds selected were based on the 0.1st and 99.9th percentile of the differences in consecutive slopes of the ambient temperature, or 0.09°C/min and −0.12°C/min for rise and decay, respectively. This was higher than the thresholds selected using the 99th percentile of non-cooking days by Ruiz-Mercado et al. (2012). Our study experienced hotter and more dynamic ambient conditions, so we adjusted to capture events more effectively. Peaks within 1 h of each other were grouped into single cooking events.

Individual SUMs cooking event identification was somewhat sensitive to parameter selection in the RMM algorithm, but overall usage trends did not change substantially when varying the parameters. Results were most sensitive to the threshold slopes for entering and departing a peak. In a subset of SUM time series, it was also difficult to identify cooking events with a high degree of confidence, mostly due to indoor-outdoor movement of stoves and a preference for cooking outdoors during the non-rainy season. We may have falsely identified cooking events due to fast radiative heating from sunlight, predominantly during the midday hours. SUMs in direct sunlight were found to reach temperatures of up to 60 °C, whereas cooking events could reach peaks between 30 °C and 130 °C.

We assessed the uncertainty of the SUMs results by validating the SUMs cooking event identification algorithms using calibration data sets, in which thermocouples were collocated with SUMs on multiple stoves in the field. The training sets show that quick cooking events can be missed by the SUM if the peak identification threshold temperature is set too high, as such we set it at 30 °C for this analysis. On the longest training data set with a Gyapa stove, 39 cooking events were identified with the thermocouple, and the RMM algorithm identified 22 of those, with 11 false positives. This stove was left outdoors most days, making it one of the most challenging cases, so we expect this to be the upper limit of misclassification performance.

The performance for TSFs is likely similar to this in the worst cases, as TSFs are challenging to monitor (Burwen and Levine, 2012). However, even with a bias in the cooking period estimates from the stove types, the biases are expected to be comparable within stove types, so for example, comparisons and usage patterns over time for all Gyapa stoves can be considered valid.

Resource limitations prevented us from monitoring every stove in every home monitored, so cooking events with charcoal stoves and secondary TSFs were rarely measured. Some categories, such as the TSFs in the Philips/Philips group, had limited data due to damage to the SUM, theft of the SUM, or insufficient temperature variability during cooking events from poor SUM placement. The complete picture of cooking in some of these households may thus be obscured. To assess the impact of the low sample size in some of these categories, we randomly removed all cooking events from 1 household (jackknifing), to test the reliability of the use predictions on that particular dataset (Fig. 4). The error bars on each stove-use category are the 5th and 95th percentiles of the 100 jackknifed data sets. The uncertainties associated with each group do not change our conclusions about use of the intervention stoves, but low sample size and high uncertainty of the traditional stove use make our stacking and replacement results more tentative, as discussed later.

Survey methods

All 200 households participating in the study were surveyed at multiple time points throughout the study to measure stove use and preferences, among other topics. The first survey round was conducted in November/December of 2013, prior to the stove distribution, called baseline herein. Subsequent survey rounds were conducted in March, May/June, and August of 2014, and December 2014/January 2015. These survey rounds (rounds 2–5) were then used for comparison with SUM data (December 2013–January 2015). As part of each survey visit, interviewers completed a stove use questionnaire for all stoves in the household (Supplementary Information). That is, questionnaires were completed for each “old” stove used by the household prior to the study (charcoal stove and/or TSF), as well as each new stove (Philips and/or Gyapa). For each stove, respondents were asked to estimate a category of how many times the stove was used in the last week (0, 1–3, 4–6, 7 times). If the respondent reported that the stove was not used at all in the past week, the remaining questions for that stove were skipped and the interviewer moved to the next stove in the household. For stoves that were reportedly used at least once, the interviewer recorded whether the stove was in use “now” (i.e., at the time of the survey), and also asked whether the stove had been used “yesterday.” For current cooking and the prior day’s cooking, additional questions asked what dishes were cooked and what fuels were used, as well as who did the cooking and how many people were fed by the meals cooked on the stove in question.

Since cooking is a daily activity for most households, the “yesterday” question was designed to provide a good snapshot of cooking practices across the sample.

Self-reported behavioral data always raise concerns about possible mis-reporting. For example, in a study similar to ours in certain ways, Thomas et al. (2013) measured cookstove and water filter adoption in Rwanda using both surveying and quantitative monitoring, and found that respondents overreported use of both new technologies relative to monitor data. In the case of cookstoves, reported improved stove uses in the last week were 40% higher than measured with SUMs. In our study, we chose short recall periods (i.e., past week, yesterday) to minimize recall bias and make it easier for respondents to provide specific answers about their cooking practices. In an effort to limit reporting bias and reactivity effects, project staff employed as part of the measurement and survey teams were trained to maintain a neutral attitude and encourage truthful reporting of stove use by participants, emphasizing the importance of collecting accurate information about users’ experiences in order to improve the stoves and their usefulness in the future. Thus, during these visits participants were not explicitly encouraged to use their improved stoves, or use them exclusively, and there was no required cooking demonstration or instruction on the use of the stoves. However, it is of course possible that reporting biases remain in our data despite these efforts. In particular, we hypothesize that households in the intervention groups given stoves would tend to overreport their use of new (Gyapa and Philips) stoves, and underreport use of their traditional stoves. Potential for misreporting of stove use seems lower in the control group.

Results

Stove use across intervention groups

Fig. 4 shows usage rates for the Gyapa, Philips, TSF, and charcoal stoves (charcoal SUM results are excluded here due to low SUM coverage, but survey results are presented). For these four stove types, the plot shows the rate at which households use any stoves of this type, as measured by SUMs and reported in surveys. For Gyapa and Philips stoves, survey-based estimates of stove usage rates are consistently higher than SUMs-based measurements. This largest discrepancy was 6.8%, for use of Philips stoves in the Philips/Philips group, where 61.5% of survey respondents said they used the stove yesterday, while SUMs showed that they were used on 54.7% of days. The smallest difference, 2.4%, was for Philips use in the Philips/Gyapa group, where surveys showed 51.1% of users used the stove ‘yesterday’, and SUMs showed use on 48.7% of days. Despite these discrepancies, the surveys and SUMs tell a consistent overall story regarding
the relative patterns of use of these two stove types across intervention groups. Specifically, Gyapa stoves are used at substantially higher rates than Philips stoves. Gyapa/Gyapa households used a Gyapa stove on 82.6% of days according to surveys, or 77.9% according to SUMs. The Gyapa use rate in the Gyapa/Philips group was lower, at 61.5% and 54.7% of days according to the surveys and SUMs, respectively. In contrast, the Philips stoves from the Gyapa/Philips group were used on 25.8% and 23.8% percent of days (surveys, SUMs). The Philips/Philips group used a Philips stove on 48.7% and 51.1% of days (surveys, SUMs). This was close to twice the rate of Philips use from the Gyapa/Philips groups, and substantially less than the Gyapa use from the Gyapa/Gyapa group.

### Traditional stove replacement

A primary objective of this and other cookstove interventions is to produce a reduction in use of traditional stoves. Thus, comparing use of TSFs across groups is a key outcome measure. Unfortunately, the combination of SUMs event detection uncertainty and possible reporting biases leads to difficulty making strong conclusions when it comes to TSF and charcoal stove use. Most puzzling is that while the control group households reported using their TSFs on 98.9% of days measured, SUMs detected TSF cooking events in these households on just 81.3% of days. In this case, the SUMs estimate seems implausibly low for three reasons: 1) these households rely on traditional stoves for all of their cooking needs, 2) the control group is expected to have little motivation for overreporting their stove use, and 3) additional TSFs may be in use while not being monitored. Thus, it is likely that the discrepancy between the SUMs and survey results is due to the challenges of measuring TSFs with SUMs.

To add complexity, reporting biases (specifically, underreporting) are expected to be a factor in survey-based measurements of traditional stove use for the other three groups, such that in these groups it is likely that both types of measurement are low. The direction of the survey-SUMs discrepancies is consistent with under-reporting TSF use, though the magnitude of the difference between surveys and SUMs varies from 1.1–15.8%. Surveys and SUMs were in the closest agreement for the Gyapa/Philips group, with 57.7% of days used for the SUMs, and 56.6% of days used according to surveys. The Philips/Philips group had the second best agreement between methods, with 81.5% of days (SUMs) and 73.0% (surveys), while the Gyapa/Gyapa daily use rates were 64.5% (SUMs) and 48.7% (surveys).

Use of TSFs remained high among all three of our stove intervention groups, and was consistently higher in the Philips/Philips group compared with the two groups that received at least one Gyapa stove. Regarding the question of which of the Gyapa groups had the lowest rate of TSF use, surveys and SUMs are in disagreement: reported use is lowest in the Gyapa/Gyapa group, but SUMs-measured use is lowest for the Gyapa/Philips group. Unfortunately, without further assumptions, it is unclear which of these conclusions is correct.

SUM data are extremely limited for the charcoal stove category, so we must rely entirely on survey data. These data showed a strong reduction in charcoal stove use in the intervention groups relative to the control group, with 29.6% of control group homes reporting use ‘yesterday’ compared to 16.3% of intervention group households.

For a subset of cases, we have both SUMs and survey measurements for the same households on the same days, allowing us to directly observe the method agreement rate for measured vs. reported stove use in each stove group and stove type. Specifically, we compared stove use measured by SUMs ‘yesterday’, the day before the survey, with survey responses on stove use ‘yesterday’. We found moderate agreement on stove use ‘yesterday’, with error likely due to both misreporting and SUM uncertainty. Method percent agreement was highest for detection of TSF stove use (81.4%, n = 43), and Gyapa stove use (62.1%, n = 58). Method percent agreement here is relative to survey results; in other words, the percent of SUM results in agreement with the survey results. The lowest agreement was when the survey indicated that the TSF was not used on the survey day (25.0%, n = 44). Agreement between all SUMs and surveys declined slightly over the three survey periods, possibly indicative of SUM attrition. There was a higher rate of SUM agreement with survey data for cooking-events than non-events – i.e., when the respondent reported that they used a particular stove yesterday (Figure shown in SI). The fact that we see lower agreement on a case-by-case basis compared to the relatively high agreement between the aggregate measures of use presented in Fig. 4 is likely due to the low sample sizes for these case-by-case comparisons and the fact that there is measurement error in both data sources (reporting bias on the part of the respondents, and a systematic underestimation of cooking events by the SUMs).

We also analyzed reactivity to enumerator visits in terms of both total number of SUM-detected cooking events with each stove, and number of SUM-detected days each stove was used. Here, reactivity was measured by comparing those metrics in a three-day window before the visit was scheduled, with a three-day window lagged at various intervals post-visit (the 3-day windows of post-visit analysis started 1, 4, 7, and 10 days after the visit). We found subtle increases in intervention stove use for both metrics, with concomitant decreases in TSF use in the windows after the visits, appearing to stabilize in the 7–10 day windows after the visit (Figure shown in the Supplementary Information), indicating that there may have been a subtle enumerator reactivity effect in our study. Further analysis enumerator reactivity, and reactivity due to the knowledge of being monitored by SUMs will be assessed in future work.

### Stove stacking

Fig. 5 shows the percent of days (as measured by both surveys and SUMs) on which households in different intervention groups used any of the 10 possible combinations of two types of stoves among the Gyapa, Philips, TSF, and charcoal stoves. Use of two Gyapas in the Gyapa/Gyapa group occurred more frequently (28.8% and 25.0% for

---

**Fig. 4.** Use of different stove types among households in different intervention groups, as measured by SUMs and surveys. Figure shows aggregate results using all data available after stove deployment (Nov’13-Jan. ’15).

**Fig. 5.** Use of different stove types among households in different intervention groups, as measured by SUMs and surveys. Figure shows aggregate results using all data available after stove deployment (Nov’13-Jan. ’15).
SUMs and surveys) than use of a Gyapa and Philips in the Philips/Gyapa group (20.2% and 14.3% for SUMs and surveys) or use of two Philips in the Philips/Philips group (14.4% and 8.4% for SUMs and surveys).

Stacking of the Gyapa or Philips stoves with TSFs is much higher from SUM data than survey in the Gyapa/Gyapa (19.6% higher) and Philips/Philips group (12.4% higher). The previously mentioned sources of uncertainty may account for this, and it should be noted that the estimated confidence intervals on both of these groups are large due to the low SUM coverage on these groups. We also find a consistently higher use rate from the SUM data than from surveys, which could be due to the sources of our uncertainties. For example, certain climatic conditions could produce false positives in multiple stoves at a household, or alternatively, co-heating could raise the temperature of multiple SUMs in a cooking area, when not all stoves with SUMs are actually used. Again, due to low coverage or failures of SUMs on TSF and charcoal stoves, we rely on survey results for indication of stacking with charcoal stoves. We found substantially similar results for the three non-control intervention groups, with lower stacking rates of multiple traditional stoves than the control group. Less than 13% of homes from those three intervention groups reported stacking multiple TSFs 'yesterday', while 24% did for the control group. Stacking of a charcoal stove and TSF was reported as less than 10.1% of days for all intervention groups, and 28.4% for the control group. Stacking of two charcoal stoves was very uncommon for all groups, less than 1.5% of days. Similarly, stacking of intervention stoves with charcoal stoves was low, with all groups reporting fewer than 9.8% of days used.

This can be contrasted with the work by Loo et al. (2016), who found that users in Kenya preferred the Philips stove among the six ICSs tested, while Lozier et al. found that even with the preferred Philips stove, it was only used exclusively on 24% days, while traditional methods were used exclusively on 25% of days, and stacking occurred on 45% of days.

**Temporal stove usage trends**

In the ideal scenario of perfect accuracy from the SUMs and from the surveys, the key benefit of SUMs data is higher temporal resolution. Fig. 6 shows a time series of stove use for each type of stove, for the daily SUM data (smoothed with a spline and with shaded area 95% CIs) and for the quarterly survey data (dots). The general trends are in good agreement, except for the charcoal stoves, where the SUM data are extremely sparse. Interestingly, the use of most stoves appeared to be in phases, with strong continued use of a stove for a period of time, followed by very little use. There were no obvious patterns or periodicity to this, and may simply have followed each home’s fuel collection. There are also large fluctuations in usage for the Gyapa and TSFs from October to December 2014, which partially occur in between survey periods, and during the end of the rainy season when harvesting typically occurs, which could impact fuel use.

The intervention did not introduce any new fuel types to the homes, and homes appear to have continued using wood, charcoal, and millet stalks primarily. Woods commonly used for cooking include Neem, Shea, Mango, and varieties locally known as Zanka, Sesibe, and others. Millet stalks, shown in Fig. 2, are usually 1–3 cm diameter, often up to 8 ft in length, collected from agricultural byproducts. They have high lignin content, and a higher heating value of 18.05 MJ/kg, similar to agri-wastes like wheat straw and bagasse (Nhuchhen and Abdul Salam, 2012). Survey results indicate that the Philips stoves were used with...
charcoal about 1/3 of the time, with wood making up the remaining fraction. The Gyapa stove has reportedly been used with wood 80% of the time, with millet stalks making up the rest. More detailed analysis of fuels and fuel stacking, a critical component of health and environmental impacts (Masera et al., 2015), will be undertaken when surveying is completed.

There was no clear initial adjustment period for the intervention stoves, as seen in other studies, but that may have been due to the relatively low sample size for each group and the difficulty of removing seasonal variations in a one-year data set. This will be analyzed formally with the complete two-year data set, but the survey data do show a lower TSF use rate for the Gyapa/Gyapa and Philips/Philips groups, followed by sustained use on about 50% of days on subsequent surveys. We observed decreasing use of all Gyapa and Philips stove categories over the year, though we have not yet quantified the trend. Charcoal stove use also appeared to decrease over the course of the study, except for the Philips/Gyapa group, though seasonal and adjustment periods may be driving their use as well.

There was no significant day of week trend for cooking with any of the stoves. This was likely because the main temporal periodicity for such rural households only is due to local markets, which cycle locations every 3 days. Cooking peak distributions (Figures in SI) show that the stove groups use different stoves at different times of day, further illustrating the reasons for stacking. The Gyapa stoves, for example, appear to be used preferentially later in the day compared to the others.

**Daily stove use durations**

Although estimation of cooking time from the SUM data has large uncertainties due to our experimental set up, we did observe it to be lower in our study compared with previous work undertaken in the Sissala district in Northwestern Ghana (Burwen and Levine, 2012). Our control group was found to cook with TSFs 4.6 h/day (we have insufficient data from charcoal stoves to measure cooking time), while in Sissala, the control group was found to cook an average of 10.7 h/day. Burwen and Levine (2012) also found substantial stacking of traditional methods with the brick and clay stove their study introduced. There, the treatment group cooked with the intervention stove 26% of the time, 2.5 h/day with intervention stoves and 7.1 h/day with traditional stoves. Stacking ratios were higher in our study, with the intervention stoves accounting for 61.5% of total cooking time for the Gyapa/Gyapa group, 59% for the Gyapa/Philips group, and 32% for the Philips/Philips group. We also found that when two identical stoves were provided, they were used nearly the same amount, likely to prolong their lives or minimize wear. In the case of the Philips stove, this would also help mitigate downtime caused by the need to charge the stove with solar panels during daylight hours.

**Links between stove type and type of food cooked**

Fig. 7 presents survey data on the dishes households in the different intervention groups reported cooking with each stove type during the Dec. ’14/Jan. ’15 survey round, approximately one year after receiving their new stoves. At baseline, we noted that TZ was cooked exclusively over TSFs, while vegetable soup and, to a lesser degree, rice, were also cooked over charcoal stoves (Table 1). As Fig. 7 shows, this pattern persists in the control group households. However, there is variation in the extent to which new stoves replaced these existing technologies to cook these three dishes in the intervention groups. Across all groups, TZ continued to be cooked over TSFs in at least two thirds of cases in which this dish was cooked. Use of Gypas to cook TZ was more common than use of Philips stoves for this dish. Specifically, in this survey round none of the households in the Gyapa/Philips group reported cooking TZ on a Philips stove. This was despite our effort to provide metal support stands for the Philips that we believed would help make them more stable for cooking dishes like TZ, which requires vigorous mixing (Dickinson et al., 2015).

Meanwhile, larger shares of the two other dishes were reportedly cooked over new stoves. Gypas were used to prepare about 65% of

![Fig. 7. Percent of households cooking different dishes with each stove type, one year after intervention (Dec ’14/Jan’15), by stove intervention group and type of stove.](image-url)
vegetable soup dishes and 50% of rice dishes in the Gyapa/Gyapa group. The shares of these dishes cooked over Philips stoves in the Philips/Gyapa group are lower: about 40% for vegetable soup and 30% for rice. In the Gyapa/Philips group, Gyapas were used more frequently than Philips stoves to prepare both of these dishes; use of the two new stoves together comprised about 50% of vegetable soup cooking and 70% of rice cooking in these households. Overall, these results indicate that households found both types of new stoves better suited for cooking dishes like vegetable soup and rice than TZ, and even for these less involved dishes, a sizeable fraction of cooking continued to be done over TSFs and charcoal stoves. These usage patterns also suggest that cleaner cooking fuels like LPG or electric stoves could face similar barriers to use, and versatile and heavy-duty designs may be required to better replace traditional stoves and fuels.

Conclusions

SUM and survey data from a cookstove intervention study in Northern Ghana were combined to assess stove stacking and adoption behavior. Use of multiple methods has previously been identified as a valuable tool for technology and behavior adoption assessment (Stanistreet et al., 2015). Our study provides further evidence of this value. Each of the assessment methods used here faced certain challenges, and there were unexplained discrepancies in results between them. Despite the discrepancies, the two approaches tell a fairly consistent story about patterns of stove use in our different intervention groups. We observed generally high rates of use of the Gyapa stoves, and lower rates for the Philips stoves. There was substantial continued use of traditional TSFs and charcoal stoves among the three intervention groups, though their use was lower than for the control group. This continued reliance on TSFs is problematic, as Johnson and Chiang (2015) found that use of a TSF for as little as 10 min per day was enough to surpass the WHO interim 1 24-h PM$_{2.5}$ limit of 35 μg/m$^3$ (WHO, 2006). This then suggests that we may not see substantial personal particulate matter exposure reductions with the stove use patterns observed in this study.

To effectively reduce use of TSFs and achieve health improvements, a better understanding of the drivers of traditional and improved stove use will be required. A mixed methods approach will likely be needed to examine these drivers. In our case, information from surveys on the types of foods cooked may help explain the continued reliance on TSF stoves. Specifically, it appears that neither of the new stove models was seen as an adequate replacement for TSFs for the task of cooking the staple porridge TZ. This thinking is brought into focus by Ruiz-Mercado and Masera (2015), who note that households rely on traditional charcoal-based stoves and TSFs to conduct a wide variety of cooking tasks and meet a diverse set of needs. Generally, any one improved stove may be well suited to some of these tasks, but less so for others, such that stacking to meet household energy needs is inevitable, and potentially optimal if appropriate and efficient technologies are stacked.

Reporting bias is always a concern with survey-based measurement of technology use. However, for the new stove technologies, survey/SUM discrepancies in use rates are not as large in this study as in others. Specifically, our primary metric of interest, use-days, held relatively constant for both surveys across all stove technologies, though recall bias relative to longer measures (e.g., reported uses in the past week). For this primary metric, we find discrepancies between reported and SUMs-measured use of new stoves that range from 2.4% to 6.8%, suggesting that well-designed survey methods may be able to generate estimates of stove use that do not differ greatly from objectively measured results.

Finally, we find that more method validation data is needed to improve cooking event identification from the SUM data. Measurement methods for temperature monitoring of TSFs and frequently moved stoves are needed for more robust and reliable stove usage estimation. Thermocouple or infrared SUMs would provide more accurate results, and this option is becoming increasingly attractive. Thermocouple loggers are available for $25–100 per unit, now comparable to iButtons (Ruiz-Mercado et al., 2008) and Digit-TLs ($18–35 per unit, respectively). Other device manufacturers have also identified and filled this gap with custom configured thermocouples as with the Nexcel StoveSense Usage Sensor (http://nexel.org/technology/cookstove-usage-sensor), SWEETSense AIR (http://www.sweetensors.com/applications/energy/), and Berkeley Air Monitoring Group K-SUM.

Ethical Considerations

This work was reviewed and approved by the Institutional Review Boards at the University of Colorado Boulder, the National Center for Atmospheric Research, and the Navrongo Health Research Center, a part of the Ghana Health Service.

Acknowledgements

This work was funded by a grant from the US National Science Foundation (Award #83542401), and by a grant from the US Environmental Protection Agency (Award #1211668).

The authors declare no conflicts of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.esd.2016.07.007.

References


Warmer K. In search of Shelter: mapping the effects of climate change on displacement and migration, a joint effort of the United Nations University Institute for Environment and human security. CARE International and the Center for International Earth Science Information Network at the Earth Institute of Columbia University New York.
