Available online at www.sciencedirect.com

ScienceDirect

http://www.elsevier.com/locate/biombioe

Comparing methods for signal analysis of temperature readings from stove use monitors

Andrew M. Simons [a,*], Theresa Beltramo [b], Garrick Blalock [a], David I. Levine [c]

[a] Dyson School of Applied Economics and Management, Cornell University, Ithaca, NY 14853, USA
[b] Impact Carbon, 47 Kearny St., Suite 600, San Francisco, CA 94108, USA
[c] Haas School of Business, University of California, Berkeley, CA 94720, USA

ARTICLE INFO

Article history:
Received 5 May 2014
Received in revised form
11 August 2014
Accepted 13 August 2014
Available online 6 September 2014

Keywords:
Nontraditional cookstoves
Technology adoption
Biomass fuel
Monitoring and evaluation
Stove use monitor

ABSTRACT

Understanding the daily use patterns of traditional and nontraditional cooking technologies is essential for researchers and policy makers attempting to reduce indoor air pollution and environmental degradation from inefficient cookstoves. This paper describes field methods and proposes a new algorithm for converting temperature data generated from stove use monitors into usage metrics for both traditional and nontraditional stoves. Central to our technique is recording the visual on/off status of a stove anytime research staff observes the stove. The observations are regressed against temperature readings in a logistic regression to estimate the probability that a temperature reading indicates usage. Using this algorithm we correctly predict 89% of three stone fire observations and 94% of Envirofit observations. The logistic regression correctly classifies more observations than published temperature analysis algorithms. This is the first published algorithm for converting temperature data for traditional stoves such as three stone fires.

1. Introduction

This study examines how best to measure the use of nontraditional and traditional stoves, an important step in the effort to mitigate the harm caused by inefficient cookstove designs. Traditional wood-and charcoal-burning stoves both burn inefficiently and produce a great amount of smoke. The smoke leads to respiratory, heart, and other disorders that kill approximately four million people per year [1]. Much of the health burden is concentrated on women and children, as is the time burden collecting biomass fuel [2]. Environmental damage can be significant; by one estimate, household energy use in Africa will produce 6.7 Gt of carbon by 2050 [3]. Furthermore, the incomplete combustion of biomass fuels leads to the release of black carbon (soot) that contributes to current global warming [4,5]. These inefficiencies imply deeper poverty, loss of biodiversity, rapid deforestation, and climate change. These problems have led both policy-makers and practitioners to search for safer and more fuel-efficient stoves that are attractive to consumers as substitutes for traditional stoves.

One challenge to understanding stove use substitution behaviors is to measure stove use precisely, economically and with minimal intrusion into the lives of study participants. Measuring stove usage with minimal observation bias is

* *Corresponding author*. Tel.: +1 214 534 0185.
E-mail addresses: ams727@cornell.edu, andrewsimons_2@yahoo.com (A.M. Simons).
http://dx.doi.org/10.1016/j.biombioe.2014.08.008

necessary to recover unbiased estimates of stove use because study participants behave differently when research staff are present [6] and because study participants over report stove use when it is self-reported [7]. Scientists studying the health effects of cooking technologies need to understand how many hours a day a stove is used to estimate exposure to household air pollution [8,9]. Project financiers need to measure the time cooked on both new and old stoves to allocate the carbon credits that can fund stove subsidies in developing countries [10–12]. Further, "stove stacking" (the use of multiple fuels and stoves) often occurs when multiple cooking options exist [13–15]. To understand any of these issues researchers must know the baseline amount of cooking on existing stoves, and then the time cooked on both the existing and new stove.

Pioneering stove studies have highlighted the harmful effects of indoor air pollution and the health benefits of adopting fuel-efficient nontraditional stoves [16–22]. However these studies were characterized by high levels of interactions between research staff and study participants; for example, in Ref. [16] research staff noted the locations of each household member and the status of how active the cooking fire is every 5–10 min. Observational studies of stove use raise the possibility of observation bias (or Hawthorne effect), which arises when participants act differently than they normally would because they know they are being observed (as noted by Ref. [22]). Our study is similar to [7] in that both use co-located observational and sensor data to calibrate a sensor to measure household product use (in their case, a water filter).

To minimize observation bias, stove use monitor systems (SUMs) can record stove temperatures without the need for an observer to be present. The SUMs used for our project, iButtons™ manufactured by Maxim Integrated Products, Inc., are small stainless steel temperature sensors about the size of a small coin and the thickness of a watch battery that can be affixed to any stove type. The SUMs used in our study record temperatures with an accuracy of ±1.3 °C up to 85 °C (see Fig. 1 for a photograph). Stove Usage Monitors that log stove temperatures were first suggested by Ref. [23]. SUMs offer an unobtrusive, precise, relatively inexpensive (approximately USD $16 each), and objective measure of stove usage [23,24].

We know of two published algorithms for determining stove usage with SUMs. Both studies use SUMs to measure the temperature of nontraditional cookstoves at set time intervals, and then calculate temperature changes between readings. In Ref. [25] the algorithm considers an increase in temperature above a certain threshold (1.52 °C increase over 40 min) as the start of a candidate cooking or refueling event. If the post-increase stove temperature is also above the ambient temperature, the algorithm counts the passage of time until a temperature drop indicates that the stove is cooling down (indicated by a 2.28 °C decrease over 60 min), signaling the end of the cooking event. These slope thresholds are chosen by taking the 99th and 1st percentiles of slope values from the distribution of ambient air temperatures in the research area. Multiple temperature peaks within a 2 h period of each other are considered the same cooking event [25]. A similar technique, combining both change in slope and a temperature threshold to classify when a stove is being used, is used in a similar study, however, the researchers choose different thresholds based on observed local cooking practices and local ambient temperatures [26].

The use of temperature slopes to identify cooking works well for heat-efficient manufactured stoves, such as those used in these two studies [25,26]. Because these stoves heat and cool quickly and have fixed form-factors, temperature slopes are steep and can be consistently measured by SUMs attached to a fixed spot. For traditional stoves, such as a three stone fire, however, the use of temperature slopes can be challenging for two reasons. First, three stone fires heat and cool more slowly, producing more attenuated temperature slopes. Second, each three stone fire has a unique form-factor that can change over time, e.g., a stone is moved. Thus, the same cooking activity could create different measured temperature slopes both across stoves and within the same stove over time. The challenges suggest that a more flexible functional form mapping temperature readings to cooking activities might be advantageous for three stone fires.

We propose a methodology that uses SUMs temperature data gathered in the same way as previous studies, but combines this temperature data with observations made throughout the experiment of when stoves were seen as on or off. We run a logistic regression that fits SUMs temperature data to observations of stoves being on or off. We then take the coefficients from this regression to predict the probability of cooking over the much larger set of SUMs temperature readings without any observational data. The researcher can sum the predicted minutes cooked in a 24-h period to estimate minutes of cooking that day. This measure of cooking can then be used as the basis for other analysis such as comparing cooking behaviors between groups of households with certain characteristics or groups differentiated by experimental design.

Fig. 1 – Photograph of stove use monitoring system (SUMs).

2. Methods

Our technique requires continuous SUMs temperature data for a given stove, and recorded instances of whether that particular stove is seen in use or not (henceforth called "observations"). We matched observations of stove use to SUMs temperature data by time and date stamps. The core of our method is a logistic regression using the lags and leads of the SUMs temperature data to predict observations of stove usage.

2.1. Continuous SUMs temperature data

A stove usage monitor takes a temperature reading at an interval specified by the user. The SUMs used in this experiment hold approximately 2000 data points before the data needs to be downloaded and the device reset. We set the SUM to record temperature every 30 min, giving it six weeks between required servicing. Instantaneous temperature readings from SUMs devices in a given household were matched based on date and timestamp to observations of stove use made in that same household. The observation of use is matched with the temperature reading with the closest date and timestamp. Since our temperature data occurs every 30 min, this ensures that at most a visual observation is ±15 min from the timestamp of the temperature reading.

2.2. Observations of stove use

Each time the data collection team visited a household they observed which stoves were in use (date and timestamp were recorded digitally via handheld device). Enumerators visited a house numerous times during a "measurement week," when we also enumerated a survey and weighed wood for a kitchen performance test [27,28]. Another enumerator visited once every 4–6 weeks to download data and reset the SUM.

A detailed description of the process to match the timestamps of observations of use to SUMs level temperature data is described in the web enabled Supplementary Material, but a summary conclusion of that process is presented in Fig. 2. This presents the outcomes for the visual observations of three stone fires, but the same process was followed for Envirofits as well. The largest challenge was to reconcile households with two SUMs readings, two three stone fires, and an observation of one stove in use. To do this we classified the hotter of the two stoves (within ±60 min of the observation) as 'on', and the colder observation as 'off.' For the cases of households with one SUMs reading, two three stone fires, and an observation of one stove being used, we could not definitively determine if the three stone fire observed as 'on' was the same three stone fire as the one with the SUMs device attached, therefore we had to exclude these observations from use in the subsequent analysis.

We originally intended to use the entire sample of observations. We found, however, a number of anomalies. For example, 3.0% (10 out of 329) of "lit" three stone fire observations had SUM readings below the daily mean temperature (23.8 °C). At the same time 7.8% (105 out of 1339) of "not lit" three stone fire observations had SUM readings over 40 °C, which was the highest ambient air temperature recorded in the observation period. While suspicious, these readings may not necessarily be erroneous, as SUMs can be cool if the stove was just lit in the last few minutes and the SUM has not yet registered the heat increase, or a SUM could still be hot on an unlit stove if the stove was just extinguished.

We suspect a substantial share of such cases were due to errors in observing or recording the lit status of stoves or in errors in matching the time stamps, stoves, or homes of observations and SUMs measurements. Therefore we trim the observation sample based on the hourly ambient temperatures (with a conservative adjustment). We drop observations observed as "lit" below the mean minus 2.0 °C of the ambient temperature for that hour, and we drop observations observed as "not lit" that are above the maximum ambient temperature plus 2.0 °C for that hour. The mean and maximum hourly ambient temperatures are taken across the entire sample of ambient temperature readings. This removes 11.0% (184 out of 1668) and 3.6% (28 out of 782) of the observation samples for three stone fires and Envirofits, respectively.

2.3. Regression specification

There is no theory of what combinations or transforms of current, lagged, and leading temperature readings should be used to predict observed cooking. We test ten specifications and compare the goodness of fit for each specification. (We do not also include slope terms, because they are perfectly collinear with multiple lags or leads.) For simplicity, our desire is to use the same specification on both the three stone fire sample and the Envirofit sample.

3. Data

A series of randomized control trials were executed in rural areas of the Mbarara District in southwestern Uganda (Fig. 3) from February to September 2012, which focused on the use, and adoption of fuel-efficient stoves (see Refs. [29,30] for more details on the experimental design and an overview of the study area). In our study the traditional stove was a three stone fire, (simply three large stones, approximately the same height, on which a cooking pot is balanced) and the

	One TSF at home	Two TSF at home, two SUMs at time of observation	Two TSF at home, only one SUM at time of observation
No stove observed lit in observation	65	672	339
One stove observed lit in observation	23	§263 "on", 263 "off"	†168
Two stoves observed lit in observation	--	38	5

§: We classified the hotter of the two stoves (within +/- 60 minutes of the observation) as'on' for this case. So that means there are 263 'on' observations and 263 'off' observations at the stove level for these households.
†: We dropped these cases in the main regression, as we were unable to definitively determine which stove was on.

Fig. 2 – Stove level temperature data matched with household level observation of stove use data.

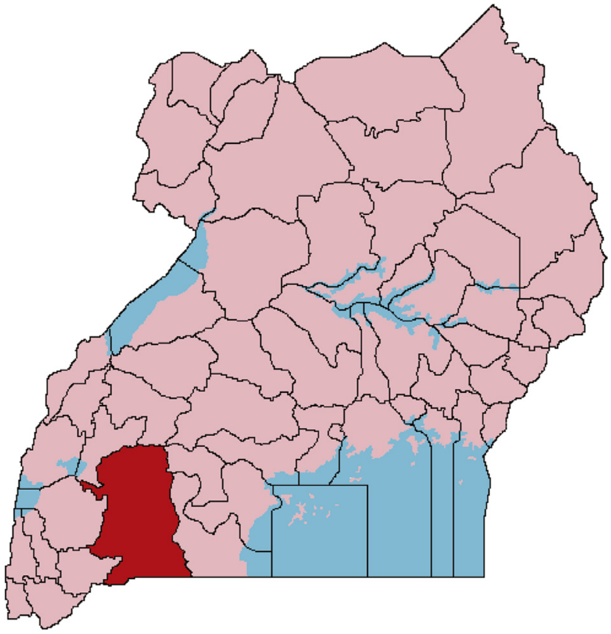


Fig. 3 – Map of Uganda with Mbarara district highlighted.

nontraditional stove was a fuel-efficient Envirofit G-3300 (a engineered stove designed to be heat efficient with the manufacturer reporting 60% biomass fuel reduction, an 80% reduction in harmful smoke and gasses, and reductions of cooking time of up to 50% versus the three stone fire). The study area is characterized by agrarian livelihoods including farming of *matooke* (starchy cooking banana), potatoes, and millet as well as raising livestock. At baseline almost all families cook on a traditional three-stone fire (97%), usually located within a separate cooking hut (62% of households had totally enclosed kitchens with no windows, while 38% had semi-enclosed kitchens with at least one window). Most stove usage occurs during lunch and dinner preparation, with *matooke* and beans the most common and most time consuming foods cooked. *Matooke*, a main food for lunch and dinner, is unripe plantain eaten after steaming for 3–5 h. Beans, another common dish, is prepared by boiling and simmering for generally 2–4 h. Thus for the main meals, it is common cooking practice to simmer and/or steam foods for several hours in a row.

The study tracked stove usage (before and after the purchase of a fuel efficient stove) amongst twelve households in each of fourteen rural parishes in Mbarara (168 total households). Upon arriving in a new parish, staff displayed the new Envirofit and offered it for sale to anyone who wanted to purchase at USD $16 (local currency converted at rate of 2500 UGX USD^{-1} on March 31, 2012, see Ref. [30] for an overview of the sales contract). Consumers who wanted to buy the stove were randomly assigned into two groups (early buyers, late buyers). The project asked both early buyers and late buyers if they would agree to have SUMs placed on their traditional stoves immediately. Then approximately two weeks later the early buyers group received their first Envirofit stove, and approximately four to five weeks after that the late buyers received their first Envirofit stove. Households were eligible to participate in the study if they mainly used wood as a fuel source, regularly cooked for eight or fewer persons, someone was generally home every day, and cooking was largely in an enclosed kitchen. In each parish, more than twelve households met these criteria and agreed to join the study; therefore among those that agreed, we randomly selected twelve households per parish for the usage study with the SUMs. Stove temperatures were tracked for approximately six months (April–September 2012).

Approximately six weeks after late buyers received their Envirofits, both groups were surprised with a second Envirofit stove. Because common cooking practices in the area require two simultaneous cooking pots (for example rice and beans, or *matooke* and some type of sauce), and the Envirofit is sized for one cooking pot, we gave a second Envirofit to mimic normal cooking behavior as much as possible. Each household had as many as two three stone fires and two Envirofit stoves monitored with SUMs throughout the study. Household temperature data from both Envirofits and up to two three stone fires are used in the study. The results of stove use and the substitution patterns between stove types will be presented in a subsequent analysis.

4. Results

4.1. *Descriptive statistics*

The data consist of approximately 1.7 million temperature readings from more than 34,000 stove-days of use recorded across 580 stoves at 168 households. The sample sizes by stove type are in Table 1. Among the 168 households, 154 initially had two three stone fires present in their home and all 168 households accepted the second Envirofit stove when it was offered.

SUMs must be placed close enough to the heat source to capture changes in temperatures, but not so close that they exceed 85 °C, the maximum temperature the SUMs used in this study can record before they overheat and malfunction. We do not need to recover the exact temperature of the hottest part of the fire to learn about cooking behaviors. Even with SUMs that are reading temperatures 20–30 cm from the center of the fire, as long as the temperature readings for times when stoves are in use are largely different than times when stoves are not used the logistic regression will be able to predict a probability of usage. SUMs for three stone fires were placed in a SUM holder (Fig. 4) and then placed under one of the stones in the three stone fire (left panel, Fig. 5). The SUMs for Envirofits were attached using duct tape and wire and

Table 1 – Temperature readings, days, stoves observed by stove type.

Stove type	Readings	Days	Stoves
Envirofit 1st	432,033	8945	153
Envirofit 2nd	137,413	2974	107
Three stone fire 1st	708,048	14,081	168
Three stone fire 2nd	417,149	8204	152
Totals	1,694,643	34,204	580

Note: data collected March–September 2012.

Fig. 4 – SUM holder designed to encase the stove use monitor to protect it from malfunctions when exceeding temperatures of 85 °C.

placed at the base of the stove behind the intake location for the firewood (right panel, Fig. 5). Over multiple weeks users likely move the locations of the stones in their three stone fires depending on their cooking needs, pushing the stones closer together for a smaller fire and wider to fit more fuel wood for a larger fire. Therefore we expect the tracking of three stone fire temperatures to be more difficult with SUMs than the tracking of Envirofit stove usage.

Fig. 6 shows an example of SUMs temperature data for a household across about three weeks. The left panel shows the temperatures registered in a three stone fire versus the ambient temperature also recorded with SUMs in this household, while the right panel compares the temperature of the Envirofit to the ambient temperature reading. These two figures illustrate a point that is largely consistent throughout our wider data set. SUMs readings of the three stone fires are less responsive than Envirofit SUMs readings to temperature changes. It seems that residual heat stays in the cooking area of a three stone fire preventing temperatures from dropping all the way to the ambient temperature. The design of the Envirofit, which is engineered to be much more heat efficient, shows up in these SUMs graphs as less residual heat when a stove is not in use by the temperatures dropping to room temperature or below.

A final difficulty is found July 2–4 (left panel, Fig. 6) where temperatures for the three stone fire remain above 40 °C for about a 48-h period. While it could be that the fire was on for this entire period, it could also be that the household was banking embers (keeping coals warm, but not cooking) and the SUM was close to those embers. This figure highlights some of the potential challenges of using temperatures as a predictive mechanism for cooking metrics, especially with three stone fires.

Ambient temperature sensors were placed in one household in each of the fourteen parishes. Fig. 7 is a box plot of the ambient temperatures by hour of the day (averaged over each day of the experiment).

There are 1668 visual observations for three stone fires, and 782 for the Envirofit stove (see the web enabled Supplementary Materials for a detailed description of the process to match observations to temperature data). The summary statistics for the entire sample and for the trimmed sample (dropping observations that were implausible given the ambient temperature) are in Table 2. Note the spread of mean temperatures (“off” vs. “on”) is wider for observations for Envirofits (26.2 °C vs. 41.5 °C in full sample, and 25.3 °C vs. 42.0 °C in trimmed sample) as compared to the three stone fires (30.2 °C vs. 38.3 °C in full sample, and 28.2 °C vs. 39.0 °C in trimmed sample).

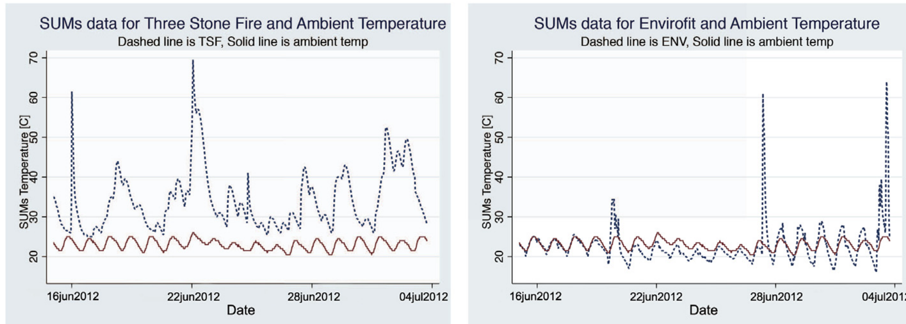
4.2. *Comparison of fit: full vs. trimmed sample*

We first examine one regression with two leads and lags on both the full and trimmed samples. Table 3 shows the logistic regression predicting the observation that the stove was recorded as lit. The percent correctly classified (observations were classified as “on” if the predicted probability was $\geq$0.5) is good (83.1%), but the pseudo-R^2 is a modest 0.17. Fig. 8 shows the predicted probability a stove is observed in use at each SUM temperature. Contrary to our priors, the predicted

(a) Three Stone Fire

(b) Envirofit

Fig. 5 – Approximate placement of SUMs on three stone fire and Envirofit in this study.



(a) Three Stone Fire (b) Envirofit

Fig. 6 – Temperature readings on a three stone fire, Envirofit, and ambient air in one household over 20 days.

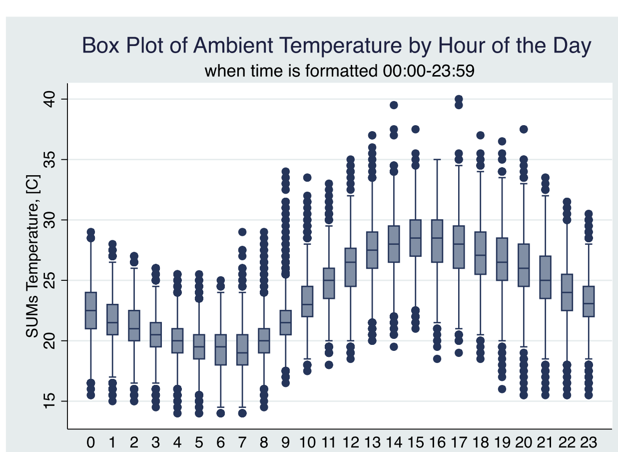
probability is fairly linear for the full sample (especially for the three stone fire).

We originally hoped to let the dataset "speak for itself." However, the predicted probability a three stone fire with a temperature reading of 50 °C is "lit" is only about 50%. Because 50 °C is much higher than the highest ambient temperature we recorded (40 °C), it should have a predicted probability of cooking much higher than 50%.

The predictive power rises substantially when we discard outliers, defined as observations coded "lit" and more than two degrees below the hourly mean air temperatures (Fig. 7), or "not lit" and more than two degrees above the maximum air temperature we observed that hour of the day across the duration of the experiment (Fig. 7). This increase in fit holds for the three stone fires (pseudo-R^2 increases from 0.17 to 0.40, percentage correctly classified increases from 83.1% to 89.3%) and Envirofits (pseudo-R^2 increases from 0.45 to 0.70, percentage correctly classified increases from 90.7% to 93.8%). (Note that trimming these outliers mechanically improves the fit and percent correctly classified.) When we drop outliers the predicted probability now has the expected logistic S-shape (Fig. 8).

4.3. *Regression specifications*

We test ten regression specifications on the trimmed sample for both the three stone fires and the Envirofit (Table 4). Initially, we include only the temperature at time *t* (specification 1); it has a pseudo-R^2 of 37% and 60%, and correctly classifies 87.5% and 96.1% of the three stone fire and Envirofit sample, respectively. Adding a lead and lag of one period (spec. 2) increases the pseudo R^2 to 41% and 68%, and increases the percentage correctly classified to 89.2% for three stone fires, but decreases the percentage correctly classified to 94.9% in the Envirofits. Adding a second (spec. 4), third (spec. 6) or fourth (spec. 8) lead and lag does not improve fit by much. Adding a slope-squared term when the slope is positive for the time step from *t-1* to *t* (coding slope-squared as zero otherwise) measures steep positive increases in slope, as would be the case when a fire is lit. However adding this slope-squared term (col. 3, 5, and 7) minimally changes the results. Only including leads (spec. 9) or only including lags (spec. 10) reduces pseudo-R^2, but increases percentage correctly classified to 89.4% and 95.3% (spec. 10) for three stone fires and Envirofits, respectively.



Box and whisker plot: The bottom and top of each box represent the first and the third quartiles of the hourly ambient temperature distribution, the horizontal band inside each box represents the median hourly ambient temperature. The end of the whiskers marked with a horizontal line represent the lowest datum within 1.5 times the interquartile range of the lower quartile and the highest datum within 1.5 times the interquartile range of the upper quartile. The dots represent individual observations that are outliers outside of this range.

Fig. 7 – Distribution of hourly ambient temperatures.

Table 2 – Summary statistics: observations of stove usage.

	Mean	Std. Dev.	Min.	Max.	N
Observations					
Temp [C] when three stone fire observed as off	30.17	7.27	19.5	85	1339
Temp [C] when three stone fire observed as on	38.30	10.63	20	85	329
Temp [C] when Envirofit observed as off	26.17	6.34	18.5	76	704
Temp [C] when Envirofit observed as on	41.52	11.7	20	69.5	78
Observations–trimmed sample[a]					
Temp [C] when three stone fire observed as off	28.21	4.60	19.5	41.5	1169
Temp [C] when three stone fire observed as on	38.97	10.36	22	85	315
Temp [C] when Envirofit observed as off	25.27	3.86	18.5	41	678
Temp [C] when Envirofit observed as on	42.04	11.39	20	69.5	76

[a] Note: In the trimmed sample, observations are dropped if 'lit' and more than two degrees below hourly mean air temperature, or if observed as 'not lit' but more than two degrees above the hourly maximum air temperature. The trimmed sample removes 11.0% (184 out of 1668) and 3.6% (28 out of 782) of the observation samples for the three stone fire and Envirofits, respectively.

In order to determine which of the models is most appropriate we test the ten specifications with the Akaike information criterion (AIC) [31,32]. The AIC trades off goodness of fit of the model with the complexity of the model to guard against over fitting. Given a set of candidate models for a set of data, the preferred model is the one with the lowest AIC. The AIC is appropriate in cases where the number of observations is many times larger than the square of the number of predictors [33], as in this case.

Among both the three stone fire and Envirofit sample, the AIC is a much higher value for the specifications with temperature at t (spec. 1), temperature at t and four leads (spec. 9), temperature at t and 4 lags (spec. 10) than for the other specifications. The remaining seven specifications all have similar AICs. The specification that includes two leads and two lags (spec. 4) has the lowest AIC among the three stone fire sample. The lowest AIC among the Envirofit sample is the specification with four leads and four lags (spec. 8). For simplicity, we use the same specification on both the three stone fire sample and the Envirofit sample; therefore, we rely on the specification with two leads and lags (spec. 4) for both samples.

Table 3 – Probability (odds ratio) that three stone fire (TSF) or Envirofit (ENV) is lit: full versus trimmed samples.

	(1) TSF	(2) TSF trimmed	(3) ENV	(4) ENV trimmed
SUMs temperature [C], t	1.13***	1.37***	1.08	1.46***
	(0.03)	(0.06)	(0.06)	(0.13)
SUMs temperature [C], t − 1	0.99	0.95	0.99	1.09
	(0.06)	(0.07)	(0.06)	(0.08)
SUMs temperature [C], t + 1	0.98	0.98	1.16***	1.24***
	(0.03)	(0.04)	(0.05)	(0.07)
SUMs temperature [C], t − 2	0.95	1.03	1.05	0.94
	(0.05)	(0.07)	(0.05)	(0.05)
SUMs temperature [C], t + 2	1.06**	1.07**	0.96	0.90**
	(0.03)	(0.03)	(0.03)	(0.05)
Control: hour of day	Yes	Yes	Yes	Yes
Observations	1184	1052	366	356
Pseudo R-squared	0.167	0.403	0.445	0.696
Correctly classified	83.11%	89.26%	90.71%	93.82%

Robust standard error exponentiated form in parentheses ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.
Note: Columns (1) and (3) use the full observation sample. Columns (2) and (4) are trimmed for outliers, meaning observations are dropped if 'lit' and more than two degrees below hourly mean air temperature, or if observed as 'not lit' but more than two degrees above the hourly maximum air temperature.
Note: Standard errors clustered at stove level.

4.4. *Logistic regression specification*

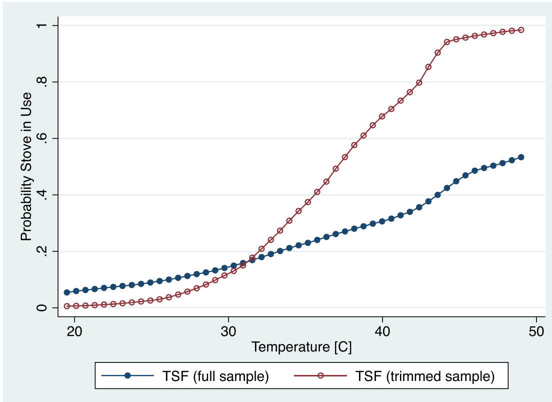
According to the AIC for three stone fires, the preferred specification is:

$$O_{it} = F(\gamma_1 T_{it} + \gamma_2 T_{it-1} + \gamma_3 T_{it+1} + \gamma_4 T_{it-2} + \gamma_5 T_{it+2} + \gamma_6 H_{it} + e_{it}) \tag{1}$$

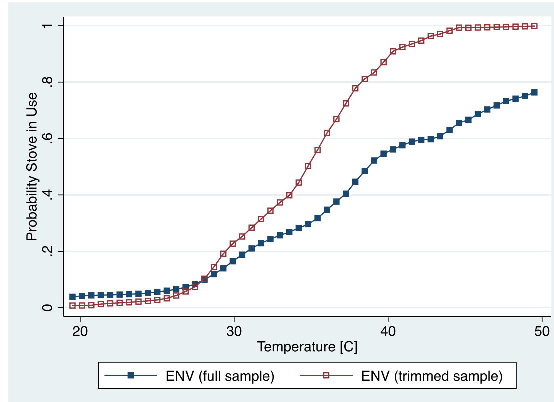
where $F(\cdot)$ is the logistic functional form, O_{it} is a dummy variable for observations for stove i at time t. T_{it} is the SUMs temperature reading for stove i at time t. $T_{it+\tau}$ is the SUMs temperature reading for stove i at time $t + \tau$, for $\tau = -2, -1, 1,$ and 2 (that is 30 and 60 min prior to the observation, and 30 and 60 min after). H_{it} is a dummy variable for the hour of the day for stove i at time t. The terms $\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \gamma_6$ are the parameters and e_{it} is an error term.

Estimated odds ratios of the logistic regression for both the three stone fires and the Envirofit are presented in Table 3 (columns 2 and 4). An increase of 1 °C at time t is associated with an odds ratio of 1.37 and 1.46 (both statistically significant at the 1% level) for the three stone fire and Envirofit, respectively. This means, holding all else constant, that an increase of 1 °C at time t means the odds of a stove being "on" are approximately 37% higher for three stone fires, and 46% higher for Envirofits than the odds of the stove being "on" at the initial time t temperature. The three stone fire temperature reading at time $t + 2$ is associated with an odds ratio of 1.07 (significant at 5% level), meaning the odds of a three stone fire being on at time t are 7% higher in the presence of a one degree increase in temperature at time $t + 2$ holding all else constant. The Envirofit temperature reading at time $t + 1$ has an odds ratio of 1.24 (significant at 1% level) and at time $t + 2$ an odds ratio of 0.90 (significant at 5% level). This means the odds of an Envirofit being on at time t are 24% higher in the presence of a one degree increase in temperature at time $t + 1$, and 10% lower in the presence of a one degree increase in temperature at time $t + 2$, holding all else constant.

We also tested probit and the linear probability model (LPM) in addition to the logistic regression. Predictions of the probability of stove usage from the probit and logistic



(a) Three Stone Fire



(b) Envirofit

Note: The trimmed sample dropped observations when observed as 'lit' and more than two degrees below hourly mean air temperature, or if observed as 'not lit' but more than two degrees above the hourly maximum air temperature.

Fig. 8 – Comparison of predicted probability of stove in use: full verses trimmed sample.

regression were correlated at levels higher than 0.99 for both the three stone fire and Envirofit sample. Among the predictions of probability of stove use from the linear probability model, 28.8% and 36.2% of the prediction fell outside of the range [0 to 1] for the three stone fire and Envirofit, respectively. As the proportion of linear probability model predicted probabilities that fall outside of the unit interval increases, the potential bias of using LPM also increases [34]. Because we have a high proportion of LPM predictions that fall outside of this range, using LPM would likely result in biases.

As a robustness check Table 5 shows the in sample and out of sample predictive accuracy of the preferred specification. We randomly split the sample of observations of each stove in half and ran the logistic regression on half of the observations. We then predict the probability of use to both halves of the sample. For the three stone fires the in-sample predictive accuracy is 89.8% while out-of-sample predictive accuracy is 87.8%. For the Envirofits the in-sample predictive accuracy is 95.5% while out–of-sample predictive accuracy is 91.1%. Because the in-sample and out-of-sample predictive accuracy is very similar for each stove type, it is unlikely that we are over-fitting the data.

A classification table for the entire sample (Table 6) shows how well the regression predicts the entire observation sample for the three stone fires and for the Envirofit. The sensitivity (probability predicted "lit" when observation observed as "lit") is only 56.5% for the three stone fires and 77.6% for the Envirofits, showing that the technique struggles when only predicting the sample that was observed as "lit." The specificity (probability predicted "not lit" when the observation observed as "not lit") is much better, 97.6% for the three stone fire, and 97.0% for the Envirofits. This technique generally holds more predictive power with Envirofits than three stone fires as evidenced by the Envirofit sample having a higher percentage of overall correctly classified observations (93.8% vs. 89.2%).

Table 4 – Candidate regression specifications: all specifications include temperature at time [t] and control for hour of the day.

Spec	Leads	Lags	Slope^2 if slope > 0	Three stone fire				Envirofit			
				Pseudo R^2	AIC	% Correctly predicted	N	Pseudo R^2	AIC	% Correctly predicted	N
(1)	0	0	–	0.37	984.01	87.49	1479	0.60	212.05	96.05	735
(2)	1	1	–	0.41	649.04	89.16	1052	0.68	120.29	94.94	356
(3)	1	1	Yes	0.42	649.44	89.07	1052	0.69	121.55	95.22	356
(4)	2	2	–	0.42	648.29	89.26	1052	0.70	120.14	93.82	356
(5)	2	2	Yes	0.42	648.49	89.35	1052	0.70	121.61	94.10	356
(6)	3	3	–	0.42	650.87	89.15	1051	0.72	117.33	94.93	355
(7)	3	3	Yes	0.42	651.19	89.34	1051	0.72	118.49	94.93	355
(8)	4	4	–	0.42	653.48	89.05	1050	0.73	116.33	94.93	355
(9)	4	0	–	0.39	865.75	86.98	1260	0.69	149.97	94.57	516
(10)	0	4	–	0.40	772.18	89.36	1269	0.61	183.79	95.30	574

Note: Classified as correctly predicted if predicted Pr(D) $\geq$0.5: True D defined when stove use observed as 'on'.
Note: Control for hour of the day is a dummy variable for each hour of the day stoves were observed in use.
Akaike information criterion: defined as AIC $= 2k - 2\ln(L)$, where k is number of parameters and L is the maximized value of the likelihood function. Given a set of candidate models for a set of data, the preferred model is the one with the minimum AIC.

Table 5 – Out of sample predictive accuracy: three stone fire and envirofit classification tables for in-sample predictions and out-of-sample predictions.

	Three stone fire		Envirofit	
	In-sample	Out-sample	In-sample	Out-sample
Sensitivity: classified as 'on' when observed as 'on'	62.04	49.07	80.00	67.86
Specificity: classified as 'off' when observed as 'off'	97.09	97.45	98.65	95.09
Positive predictive value: observed 'on' when classified 'on'	84.81	82.81	92.31	70.37
Negative predictive value: observed 'off' when classified 'off'	90.70	88.45	96.05	94.51
False positive rate given observed as 'off'	2.91	2.55	1.35	4.91
False negative rate given observed as 'on'	37.96	50.93	20.00	32.14
False positive rate given classified as 'on'	15.19	17.19	7.69	29.63
False negative rate given classified as 'off'	9.30	11.55	3.95	5.49
Overall correctly classified (percentage)	**89.81**	**87.78**	**95.51**	**91.10**
Sample size	520	540	178	191

Note: Unless noted cell contents to be read as probabilities out of one hundred.
Note: Classified as 'on' if predicted probability ≥0.5.
Note: For this exercise, the samples of observations are randomly split into two groups, the logistic regression is run on one of the two groups, then based on the coefficients of that regression, the prediction is made to the other (out-of-sample) group. Running a classification table on the in-sample and out-of-sample groups shows very little difference between the two samples, suggesting the technique is robust out-of-sample. When the samples were randomly split in half the groups were equally sized (three stone fire groups sized 742/742 and Envirofit groups sized 377/377), however the sample size of the groups presented in this table is different because various observations drop out of the regression if lacking temperature data for any of the leads and lags. The observation sample is trimmed for outliers, observations are dropped if 'lit' and more than two degrees below hourly mean air temperature, or if observed as 'not lit' but more than two degrees above the hourly maximum air temperature.

4.5. *Applying published SUMs algorithms to our data*

We replicate the algorithm described in Refs. [25,26] to convert temperature readings to hours of usage, henceforth called the RMM algorithm (Ruiz-Mercado-Mukhopadhyay et al. algorithm). We first determined a threshold increase and decrease in stove temperature equal to the 99th and 1st percentiles of ambient temperature slopes. In our sample of approximately 56,000 ambient temperature readings the 99th percentile of slopes is 83.3 mK min^{-1} (triggering slope of an increase of 2.5 °C in 30 min), and the 1st percentile of slopes is −50.0 mK min^{-1} (exit slope of a decrease of 1.5 °C in 30 min).

The RMM algorithm counts a stove as turning "on" when the slope is larger than a triggering slope, then it continues to count each subsequent temperature reading as "on" (as long as temperatures are higher than the ambient temperature) until it reaches a negative slope that is steeper than the exit slope. We calculate the on/off status of our entire data set using this algorithm. We apply this algorithm to both the traditional three stone fires and the Envirofit stoves; however [25,26] designed their algorithms for nontraditional stoves and do not endorse this technique for three stone fires.

We observe some irregularities when the RMM algorithm is applied to our data, most notably a very hot stove could be marked as "off" too quickly if the temperature falls steeper than a decrease of 1.5 °C in 30 min even if the absolute temperature is still hot enough to likely indicate cooking. Take for example the following series of temperature data {26, 51, 52, 47, 49, 46, 41, 37, 33, 31 °C} with each reading 30 min apart. The algorithm would return {off, on, on, off, off, off, off, off, off, off}. When the temperature falls from 52 °C to 47 °C between the third and fourth reading the slope is steep enough and the algorithm switches to off. However, in reality the subsequent temperature readings above 40 °C are likely to indicate stove usage (recall no ambient air temperatures were ever that high). Another irregularity with the RMM algorithm is that stoves can cool too slowly to trigger an exit slope so the algorithm returns an excessively long period of cooking (at times multiple days) despite temperatures low enough that cooking is unlikely. For example the data series {26, 33, 32, 31, 30, 29, 28, 27, 26, 25, 24, 23, 22, 21, 20 °C} would return {off, on, on, on, on, on, on, on, on, on, on, on, on, on, on} despite 20 °C being below the mean ambient temperature

Table 6 – How well do temperature readings generated from stove use monitors predict observations of stove use.

	Three stone fire	Envirofit
Sensitivity: classified as 'on' when observed as 'on'	56.54	77.59
Specificity: classified as 'off' when observed as 'off'	97.61	96.98
Positive predictive value: observed 'on' when classified 'on'	85.82	83.33
Negative predictive value: observed 'off' when classified 'off'	89.79	95.70
False positive rate given observed as 'off'	2.39	3.02
False negative rate given observed as 'on'	43.46	22.41
False positive rate given classified as 'on'	14.18	16.67
False negative rate given classified as 'off'	10.21	4.30
Overall correctly classified (percentage)	**89.26**	**93.82**
Sample size	1052	356

Note: Unless noted cell contents to be read as probabilities out of one hundred.
Note: Classified as 'on' if predicted probability ≥0.5.
Note: The observation sample is trimmed for outliers, observations are dropped if 'lit' and more than two degrees below hourly mean air temperature, or if observed as 'not lit' but more than two degrees above the hourly maximum air temperature.

in our period, and almost certainly not hot enough to indicate stove usage.

We created an adjusted RMM algorithm to account for these two cases. The first adjustment is that once the algorithm has a triggering slope marking the stove as "on" it does not consider an exit slope to trigger as "off" until the absolute temperature is less than 40 °C. The second adjustment is that once the algorithm has a triggering slope marking the stove as "on" it will switch to "off" even without a sufficiently steep exit slope if it has negative slope readings for an hour straight and the absolute temperature is below 35 °C.

4.6. Comparison of fit

Next we test which algorithm correctly predicts the highest percentage of observations. In addition to testing the RMM algorithm, and the adjusted RMM algorithm, we also test various cutoff points in relation to mean ambient temperature. For example considering all data points above temperature thresholds such as 32 °C, 34 °C, 36 °C, and 38 °C as "on" (these correspond to approximately 8 °C, 10 °C, 12 °C, and 14 °C above mean ambient temperature, respectively).

Table 7 reports the classification table of how well the algorithms predict the observations of three stone fire use. The strict threshold cutoff of 36 °C (correctly classified 85.0% of the observations. This was the best performing strict threshold, correctly predicting more than the thresholds of 32 °C (76.8%), 34 °C (81.9%), and 38 °C (84.8%). The adjusted RMM algorithm (77.7%) correctly predicted better than the RMM algorithm (58.9%), but not as well as the strict threshold of 36 °C (85.0%).

The logistic regression with temperature readings at times $t-2, t-1, t, t+1, t+2$ (that is at time t, and 30 and 60 min prior to, and 30 and 60 min after) and hour of the day predicts the most observations of three stone fire use, predicting 89.3% of observations correctly. The pseudo R-squared of the logistic regression explains 41.8% of the variation in observations, whereas each of the temperature thresholds has pseudo R-squared in the range of 10.3%–18.4%. The RMM algorithm and the adjusted RMM algorithm have a pseudo R-squared of 0.1% and 7.7%, respectively.

Table 8 shows how well the algorithms predict the observations of Envirofit use. Among Envirofits, the strict threshold cutoff of 36 °C correctly classified 94.7% of the observations. This was the best performing strict threshold, correctly predicting more than the thresholds of 32 °C (91.6%), 34 °C (93.9%), and 38 °C (94.3%). The strict threshold of 36 °C also correctly classified more observations (94.7%) than both the RMM algorithm (84.4%), and the adjusted RMM algorithm (85.9%).

The logistic regression with temperature readings at times $t-2, t-1, t, t+1, t+2$ (that is at time t, and 30 and 60 min prior to, and 30 and 60 min after) and hour of the day correctly predicts slightly fewer (93.8%) than the best performing temperature threshold of 36 °C (94.7%). However, considering the pseudo R-squared the logistic regression explains 67.1% of the variation in trimmed observations, whereas each of the temperature thresholds has a pseudo R-squared in the range of 27.0%–36.7%. The RMM algorithm and the adjusted RMM algorithm have a pseudo R-squared of 14.4% and 19.6%, respectively. Because the performance of a strict temperature threshold would vary based on the local climate the logistic regression is a more robust algorithm.

Table 7 – How well do temperature readings generated from stove use monitors predict observations of three stone fires in use: classification tables comparing algorithms.

	Three stone fire						
	Over 32	Over 34	Over 36	Over 38	RMM	adj RMM	Logit
Sensitivity: classified as 'on' when obs as 'on'	53.0	44.1	36.8	31.4	40.0	41.3	56.5
Specificity: classified as 'off' when obs as 'off'	83.2	92.0	98.0	99.1	64.0	87.5	97.6
Positive predictive: obs 'on' when classified 'on'	46.0	59.9	83.5	90.8	23.0	47.1	85.8
Negative predictive: obs 'off' when classified 'off'	86.8	85.9	85.2	84.3	79.8	84.7	89.8
False positive rate given observed as 'off'	16.8	8.0	2.0	0.9	36.0	12.5	2.4
False negative rate given observed as 'on'	47.0	55.9	63.2	68.6	60.0	58.7	43.5
False positive rate given classified as 'on'	54.0	40.1	16.5	9.2	77.0	52.9	14.2
False negative rate given classified as 'off'	13.2	14.1	14.8	15.7	20.2	15.3	10.2
Overall correctly classified (percentage)	**76.8**	**81.9**	**85.0**	**84.8**	**58.9**	**77.7**	**89.3**
Pseudo R-squared	10.3	13.4	18.4	17.7	0.1	7.7	41.8

Note: Each column represents the results of a logistic regression with the trimmed sample of observations as the dependent variable and the column header as the independent variable, cell contents are classified as 'on' if predicted probability $\geq$0.5, the psuedo R-squared is reported for each regression. The first four columns represent a simple (non-statistical) decision rule that assigns the stove status 'off' if the temperature is below the threshold, and 'on' if the temperature exceeds the threshold. The column titled 'RMM' uses the 'on'/'off' status as determined by the Ruiz-Mercado-Mukhopadhyay et al. algorithm as the independent variable for the logistic regression. The column titled 'adj RMM' uses the 'on'/'off' status as determined by the adjusted Ruiz-Mercado-Mukhopadhyay et al. algorithm (allowing the stove to remain 'on' if temperatures are very hot even if an exit slope has triggered, and switching 'off' if cooling and approaching ambient temperature even if exit slope not triggered) as the independent variable. The column titled 'Logit' uses the logistic regression developed in this paper where the observations of stove use are regressed against temperature at time t, and two lead and lag temperature readings, and the hour of day. In cases where the regression does not run due to insufficient variability or collinearity, the classification tables are calculated manually, and the pseudo R-squared is reported as n/a. All cells are to be read as probabilities out of one hundred.

Note: The observation sample is trimmed for outliers, observations are dropped if 'lit' and more than two degrees below hourly mean air temperature, or if observed as 'not lit' but more than two degrees above the hourly maximum air temperature.

Table 8 – How well do temperature readings generated from stove use monitors predict visual observations of Envirofit use: classification tables comparing algorithms.

	Envirofit						
	Over 32	Over 34	Over 36	Over 38	RMM	adj RMM	Logit
Sensitivity: classified as 'on' when obs as 'on'	57.9	52.6	48.7	43.4	56.6	63.2	77.6
Specificity: classified as 'off' when obs as 'off'	95.4	98.5	99.9	100.0	87.5	88.5	97.0
Positive predictive: obs 'on' when classified 'on'	58.7	80.0	97.4	100.0	33.6	38.1	83.3
Negative predictive: obs 'off' when classified 'off'	95.3	94.9	94.6	94.0	94.7	95.5	95.7
False positive rate given observed as 'off'	4.6	1.5	0.1	0.0	12.5	11.5	3.0
False negative rate given observed as 'on'	42.1	47.4	51.3	56.6	43.4	36.8	22.4
False positive rate given classified as 'on'	41.3	20.0	2.6	0.0	66.4	61.9	16.7
False negative rate given classified as 'off'	4.7	5.1	5.4	6.0	5.3	4.5	4.3
Overall correctly classified (percentage)	**91.6**	**93.9**	**94.7**	**94.3**	**84.4**	**85.9**	**93.8**
Pseudo R-squared	27.0	32.2	36.7	n/a	14.4	19.6	69.6

Note: Each column represents the results of a logistic regression with the trimmed sample of observations as the dependent variable and the column header as the independent variable, cell contents are classified as 'on' if predicted probability $\geq$0.5, the psuedo R-squared is reported for each regression. The first four columns represent a simple (non-statistical) decision rule that assigns the stove status 'off' if the temperature is below the threshold, and 'on' if the temperature exceeds the threshold. The column titled 'RMM' uses the 'on'/'off' status as determined by the Ruiz-Mercado-Mukhopadhyay et al. algorithm as the independent variable for the logistic regression. The column titled 'adj RMM' uses the 'on'/'off' status as determined by the adjusted Ruiz-Mercado-Mukhopadhyay et al. algorithm (allowing the stove to remain 'on' if temperatures are very hot even if an exit slope has triggered, and switching 'off' if cooling and approaching ambient temperature even if exit slope not triggered) as the independent variable. The column titled 'Logit' uses the logistic regression developed in this paper where the observations of stove use are regressed against temperature at time t, and two lead and lag temperature readings, and the hour of day. In cases where the regression does not run due to insufficient variability or colinearity, the classification tables are calculated manually, and the pseudo R-squared is reported as n/a. All cells are to be read as probabilities out of one hundred.

Note: The observation sample is trimmed for outliers, observations are dropped if 'lit' and more than two degrees below hourly mean air temperature, or if observed as 'not lit' but more than two degrees above the hourly maximum air temperature.

5. Conclusion

We describe an algorithm that combines observations of stove use with continuous temperature data from a stove usage monitor to predict the on/off status of a stove. We test various regression specifications and use the Akaike Information Criterion for choosing a logistic regression where observations are regressed against temperature at times $t-2$, $t-1$, t, $t+1$, $t+2$ and a dummy variable for hour of the day. This regression specification correctly predicts 89.3% of three stone fire observations and 93.8% of Envirofit observations of stove usage. The predictive accuracy is almost the same in and out-of-sample.

We continue by comparing the performance of this algorithm to previously published algorithms [25,26] and strict temperature threshold cutoffs. Our logistic regression correctly classifies more observations, with a higher pseudo R-squared, than any other algorithm for three stone fires. Among the Envirofit observations our logistic regression correctly predicts a higher percentage of observations than both the RMM algorithm and the adjusted RMM algorithm. The strict 36 °C threshold predicts slightly more observations (0.4%) than our logistic regression, however the logistic regression has a much higher pseudo R-squared than all other algorithms. We argue that a predictive logistic regression with observations of actual use and temperature readings at times $t-2$, $t-1$, t, $t+1$, $t+2$ and hour of the day is the best algorithm for both traditional and nontraditional stove types. In considering the external validity of this algorithm, we argue that it is transferable to many climatic zones (both hot and cold) whereas strict temperature thresholds and algorithms based on entry and exit slopes need adjustment depending on the local climate and local cooking practices. Additionally, while we have demonstrated the approach can be used with traditional cooking technologies such as the three stone fire, we anticipate the logistic regression approach to have a higher fit for stoves with a lower thermal mass (e.g. the Berkley Darfur stove) that heat up and cool down quickly.

Some challenges to the approach are the placement and location of SUMs on three stone fires and training enumerators to differentiate between a stove that is in use versus one that is not in use. We use SUM holders to protect the SUMs from overheating. While necessary to prevent overheating, this insulates the sensor from heat (likely slowing down the rate of heating up and cool down for three stone fires). Additionally, the SUM holder is placed beneath one of the stones of the three stone fire, however this location potentially moves around over multiple weeks as cooks move the location of the stones as cooking needs dictate. Moving the sensor can add noise to the cooking signal and likely contributes to why this technique performs less well on three stone fires. Lastly, we had to remove some of the original visual observations (11.0% of the three stone fire sample, and 3.6% of the Envirofit sample) because they did not make logical sense (too hot to be coded as 'off' or too cold to be coded as 'on'). For future studies more time should be spent in the up front training of enumerators to differentiate the 'on'/'off' status of a stove, and/or other higher-cost automatic devices could be installed instead (i.e., video camera, thermocouple data logger, infrared thermometer and data logger, etc.) to reduce the visual classification errors we encountered in our study that ultimately had to be removed from the sample.

By combining relatively few observations of whether a household stove is in use and quantitative SUMs data, researchers can measure stove adoption in a cost-effective fashion. Our algorithm performs better with Envirofit stoves than three stone fires. At the same time, this technique advances the frontier for calculating usage with SUMs temperature data for traditional stoves such as three stone fires. Measuring usage of both traditional and manufactured stoves is crucial due to "stove stacking" (using both old and new stoves at the same time). As such, our algorithm can help measure how new stoves affect the risks of household air pollution and fuel use (and, thus eligibility for carbon credits).

Acknowledgments

This study was funded by the United States Agency for International Development under Translating Research into Action, Cooperative Agreement No. GHS-A-00-09-00015-00. The recipient of the grant was Impact Carbon who co-funded and managed the project. Juliet Kyaesimira and Stephen Harrell expertly oversaw field operations and Amy Gu provided excellent research support. We thank Impact Carbon partners Matt Evans, Evan Haigler, Jimmy Tran, Caitlyn Toombs, and Johanna Young; Berkeley Air project partners including Dana Charron, Dr. David Pennise, Dr. Michael Johnson, and Erin Milner, the USAID TRAction Technical Advisory Group, and seminar participants at Cornell University for valuable comments. Special thanks to Kirk Smith, Ilse Ruiz-Mercado, Ajay Pillarisetti and colleagues from the U.C. Berkeley Household Energy, Climate, and Health Research Group for the development of the Stove Use Monitoring system (SUMs) concept and consultations on optimal SUMs placement and holder design for this project. Data collection was carried out by the Center for Integrated Research and Community Development (CIRCODU), and the project's success relied on expert oversight by CIRCODU's Director General Joseph Ndemere Arineitwe and field supervisors Moreen Akankunda, Innocent Byaruhanga, Fred Isabirye, Noah Kirabo, and Michael Mukembo. This study was made possible by the support of the American people through the United States Agency for International Development (USAID). We thank the Atkinson Center for a Sustainable Future at Cornell University and the Cornell Population Center for additional funding of related expenses. The findings of this study are the sole responsibility of the authors, and do not necessarily reflect the views of their respective institutions, nor USAID or the United States Government.

Appendix A. Supplementary data

Supplementary data related to this article can be found online at http://dx.doi.org/10.1016/j.biombioe.2014.08.008.

REFERENCES

[1] Lim SS, Vos T, Flaxman AD, Danaei G, Shibuya K, Adair-Rohani H, et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012;380:2224–60.

[2] Kammen DM, Bailis R, Herzog A. Clean energy for development and economic growth: biomass and other renewable energy options to meet energy and development needs in poor nations. New York: UNDP and Government of Morocco; 2002.

[3] Bailis R, Ezzati M, Kammen DM. Mortality and greenhouse gas impacts of biomass and petroleum energy futures in Africa. Science 2005;308:98–103.

[4] Bond T, Venkataraman C, Masera O. Global atmospheric impacts of residential fuels. Energy Sustain Dev 2004;8:20–32.

[5] Ramanathan V, Carmichael G. Global and regional climate changes due to black carbon. Nat Geosci 2008;1:221–7.

[6] Simons AM, Beltramo T, Blalock G, Levine DI. Using unobtrustive sensors to quantify and minimize hawthorne effects. Ithaca, NY: Evidence from Cookstoves; 2014.

[7] Thomas EA, Barstow CK, Rosa G, Majorin F, Clasen T. Use of remotely reporting electronic sensors for assessing use of water filters and cookstoves in Rwanda. Environ Sci Technol 2013;47:13602–10.

[8] Martin II WJ, Glass RI, Balbus JM, Collins FS. A major environmental cause of death. Science 2011;334:180–1.

[9] McCracken J, Schwartz J, Mittleman M, Ryan L, Artiga AD, Smith KR. Biomass smoke exposure and acute lower respiratory infections among Guatemalan children. Epidemiology 2007;18:S185.

[10] Gold Standard Foundation. The gold standard: simplified methodology for efficient cookstoves. Geneva-Cointrin, Switzerland; 2013.

[11] GIZ. Carbon markets for improved cooking stoves: a GIZ guide for project operators. Poverty-oriented basic energy services. Eschborn, Germany; 2011.

[12] Freeman OE, Zerriffi H. Carbon credits for cookstoves: trade-offs in climate and health benefits. Forest Chron 2012;88:600–8.

[13] Masera OR, Saatkamp BD, Kammen DM. From linear fuel switching to multiple cooking strategies: a critique and alternative to the energy ladder model. World Dev 2000;28:2083–103.

[14] Ruiz-Mercado I, Canuz E, Walker JL, Smith KR. Quantitative metrics of stove adoption using stove use monitors (SUMs). Biomass Bioenergy 2013;57:136–48.

[15] Masera OR, Navia J. Fuel switching or multiple cooking fuels? Understanding inter-fuel substitution patterns in rural Mexican households. Biomass Bioenergy 1997;12:347–61.

[16] Ezzati M, Saleh H, Kammen DM. The contributions of emissions and spatial microenvironments to exposure to indoor air pollution from biomass combustion in Kenya. Environ Health Perspect 2000;108:833–9.

[17] Ezzati M, Kammen DM. Evaluating the health benefits of transitions in household energy technologies in Kenya. Energy Policy 2002;30:815–26.

[18] Ezzati M, Kammen DM. Indoor air pollution from biomass combustion and acute respiratory infections in Kenya: an exposure-response study. Lancet 2001;358:619–24.

[19] Smith KR, McCracken JP, Thompson L, Edwards R, Shields KN, Canuz E, 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 2010;20:406–16.

[20] Smith KR, McCracken JP, Weber MW, Hubbard A, Jenny A, Thompson LM, et al. Effect of reduction in household air pollution on childhood pneumonia in Guatemala (RESPIRE): a randomised controlled trial. Lancet 2011;378:1717–26.

[21] Smith KR, Bruce N, Arana B, Jenny A, Khalakdina A, McCracken J, et al. Conducting the first randomized control trial of acute lower respiratory infections and indoor air pollution: description of process and methods. Epidemiology 2006;17:S44.
[22] Smith-Sivertsen T, Díaz E, Pope D, Lie RT, Díaz A, McCracken J, et al. Effect of reducing indoor air pollution on women's respiratory symptoms and lung function: the RESPIRE Randomized Trial, Guatemala. Am J Epidemiol 2009;170:211–20.
[23] Ruiz-Mercado I, Lam NL, Canuz E, Davila G, Smith KR. Low-cost temperature loggers as stove use monitors (SUMs). Boil Point 2008;55:16–8.
[24] Ruiz-Mercado I, Masera O, Zamora H, Smith KR. Adoption and sustained use of improved cookstoves. Energy Policy 2011;39:7557–66.
[25] Ruiz-Mercado I, Canuz E, Smith KR. Temperature dataloggers as stove use monitors (SUMs): field methods and signal analysis. Biomass Bioenergy 2012;47:459–68.
[26] Mukhopadhyay R, Sambandam S, Pillarisetti A, Jack D, Mukhopadhyay K, Balakrishnan K, et al. Cooking practices, air quality, and the acceptability of advanced cookstoves in Haryana, India: an exploratory study to inform large-scale interventions. Glob Health Action 2012;5:19016.
[27] Bailis R, Smith KR, Edwards R. Kitchen performance test (KPT). Berkeley, CA: Univerisity of California; 2007.
[28] Granderson J, Sandhu JS, Vasquez D, Ramirez E, Smith KR. Fuel use and design analysis of improved woodburning cookstoves in the Guatemalan Highlands. Biomass Bioenergy 2009;33:306–15.
[29] Harrell S, Beltramo T, Blalock G, Levine DI, Simons AM. What is a meal?: comparative methods of auditing carbon offset compliance for fuel efficient cookstoves. Ithaca, NY; 2013.
[30] Levine DI, Beltramo T, Blalock G, Cotterman C. What impedes efficient adoption of products? evidence from randomized variation in sales offers for improved cookstoves in Uganda. Berkeley, CA; 2013.
[31] Akaike HA. New look at the statistical model identification. Autom Control IEEE Trans 1974;19:716–23.
[32] Akaike H. Likelihood of a model and information criteria. J Econom 1981;16:3–14.
[33] Burnham KP, Anderson DR. Model selection and multimodel inference: a practical information-theoretic approach. 2nd ed. New York, NY: Springer-Verlag; 2002.
[34] Horrace WC, Oaxaca RL. Results on the bias and inconsistency of ordinary least squares for the linear probability model. Econ Lett 2006;90:321–7.