Available online at www.sciencedirect.com

ScienceDirect

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

BIOMASS & BIOENERGY

A quantitative model of cookstove variability and field performance: Implications for sample size

Christian L'Orange [a,*], David Leith [b], John Volckens [a,c], Morgan DeFoort [c]

[a] Department of Environmental and Radiological Health Sciences, Colorado State University, Fort Collins, CO 80524, USA

[b] Department of Environmental Sciences and Engineering, University of North Carolina, Chapel Hill, NC 27599, USA

[c] The Energy Institute, Colorado State University, Fort Collins, CO 80524, USA

ARTICLE INFO

Article history:
Received 11 April 2014
Received in revised form
28 October 2014
Accepted 30 October 2014
Available online 5 December 2014

Keywords:
Biomass
Uncertainty
Numerical model
Sample size
Field testing
Wood combustion

ABSTRACT

Many cookstove studies conducted in the field fail to measure meaningful differences between different stove technologies. Although meaningful differences do not always exist, significant differences are often missed because of low statistical power. A numerical model has been developed to determine the minimum sample size necessary to ensure that cookstove field studies are well-designed, efficient, and have adequate statistical power to characterize the concentrations of pollutants inside homes. The numerical model uses a Monte Carlo prediction method to generate probabilistic distributions of indoor pollutant concentrations. The model is based on a series of user inputs, including emissions rate, home size, air-exchange rate, fuel-moisture content, and measurement error. Application of this model to an example situation showed that, even under optimistic measurement conditions, a substantially high number of test replicates would be required. This approach should allow organizations to select appropriate sample sizes to test cookstoves in the field and to identify factors that contribute to variability among tests.

1. Introduction

Improved biomass cookstoves are needed as emissions from traditional cookstoves often have detrimental health and climate effects [1–4]. As nearly three billion people currently use biomass cookstoves, the cookstove problem will require global initiatives and collaborations among many organizations. To help facilitate these collaborations, new testing protocols and standardized practices for sharing data have recently been adopted. One example is the International Workshop Agreement (IWA 11:2012), which was developed through the International Organization for Standardization (ISO) to help rate and compare cookstoves [5]. Discrepancies are often seen between evaluations of cookstove performance conducted in the laboratory and those evaluated in the field. Thus, there is concern that new cookstove designs are not achieving the goals of improving health and climate. As part of the IWA, a resolution was passed to prioritize research that seeks to harmonize laboratory-based evaluations with field evaluations of cookstove performance.

The goal of the IWA field-testing resolution is to ensure that improved cookstove designs, many of which have been designed in the laboratory, actually release fewer harmful emissions than traditional designs. Evaluating cookstoves in the field presents many challenges, including determining the

* *Corresponding author*. 430 N. College Ave, Fort Collins, CO 80524, USA. Tel.: +1 970 491 4793; fax: +1 970 491 4799.
E-mail address: christian.lorange@colostate.edu (C. L'Orange).
http://dx.doi.org/10.1016/j.biombioe.2014.10.031
0961-9534/

number of test replicates required. A sufficient number of replicates is needed to determine if meaningful differences exist between different stoves. However, an unnecessarily large sample size consumes program resources and may not provide additional useful knowledge [6].

Every cookstove program will have an upper limit to the number of test replicates they can collect. The limitation may be due to cost or practicality, but every project will have a limit. As sample size increases so does personnel needs, time requirements, and study cost [6]. Although what a program can afford (time wise and financially) to devote to field testing is project specific, with increasing sample size fewer and fewer programs will have the resources needed for the study.

The sample sizes used by cookstove programs are often determined by power calculations or general “rules of thumb.” Power calculations are conducted by setting a target level of precision and an estimation of the variability between test replicates to determine the sample size. Sample size “rules of thumb” are typically generated by power calculations based on assumptions regarding the level of confidence required and the variability expected for the type of testing to be conducted. The World Health Organization and the Gold Standard Foundation are examples of organizations that calculate sample sizes by “rules of thumb” [7,8].

Although “rules of thumb” are convenient for determining sample size, they are often based on assumptions that do not translate uniformly to the field. The accuracy of power calculations is also limited by an understanding of the environmental factors that affect stove performance variability. These factors vary greatly and include properties such as room size, air exchange rate, and cultural cooking practices. Although using “rule of thumb” statistical methods at first appears convenient, as demonstrated by Edwards et al. determining an appropriate sample size using this approach is far from simple. Choosing a study design and sample size requires investigators to make a number of assumptions, best-guess estimates, and hard choices [6].

The objective of this work was to develop a user-friendly method to inform the proper design and implementation of cookstove field studies. A recent study published by Wang et al. investigated sample size requirements when testing cookstoves in laboratory settings. This study clearly demonstrated the large variability in cookstove performance results even under highly controlled environments [9]. As discussed in the previously mentioned article, even highly controlled tests conducted in sophisticated labs have sources of variability which cannot be eliminated. The sources of variability only increase when testing in the field. The model developed here extends the sample size calculations by considering the effects of cultural cooking practice, home size, and cookstove design to allow a more accurate prediction of performance variability and the number of test replicates required to achieve statistical confidence when evaluating biomass cookstoves in the field.

1.1. *Basis for the model*

Many factors influence the concentrations of air pollutants produced by cookstoves [9]. Therefore, to estimate the required number of test replicates, the major sources of variability that affect stove performance must be considered.

- Stove variability: The age and condition of a cookstove can affect performance; every cookstove is unique. Minor differences in construction can affect performance as well as age of a cookstove. Therefore, two stove units of the same design may perform slightly differently, even if construction and quality control measures are standardized.
- Fuel variability: Biomass combustion is a complex process. Small differences in the condition or composition of fuel can have a large effect on emissions. Many fuel characteristics, including aspect ratio, surface area, moisture content, and fuel type or species, affect cookstove performance [10].
- User variability: The user has a large influence on the performance of a cookstove. For example, an individual who is familiar with a particular cookstove will operate it differently than a first-time user [11,12]. Large variability will also occur for the same user day to day. In the field, users often will perform multiple cooking tasks all on the same stove; often these different cooking tasks require different operating conditions.
- Situational variability: Situational variability encompasses many components related to the location at which a stove is tested. The concentration of pollutants that accumulate in a room depends on the size and shape of the room and the airflow through the room, which depends on the number of open doors and windows [13,14].
- Measurement errors: Although errors in sample measurement do not affect stove performance variability, they do contribute to the inaccuracy and imprecision of data collected in the field. Errors in measurement include systematic and random errors. An inaccurate but consistent measurement is an example of a systematic error, which leads to biased results but not necessarily increased variability between tests. Random errors (in the context of measurement) arise from the imprecision of an instrument or variable measurement readings. An increased number of replicates can reduce imprecision due to random errors [15].

These factors interact in a complex fashion that complicates field-based measurements of biomass cookstove performance. However, because these factors (and their interactions) are stochastic in nature, they may be modeled numerically. Monte Carlo is an attractive method for modeling biomass cookstove performance, as it accommodates complex interactions between various input variables. The Monte Carlo method has been applied to many fields and disciplines, such as synaptic signaling in the brain [16] and economic planning [17].

Monte Carlo simulations are typically comprised of three elements (Fig. 1). First, equations are established for basic interactions in the system, such as how emission rate affects pollutant concentrations inside a home. Second, key parameters of the model are defined as distributions, such as the range of home sizes in a particular community. Finally, a distribution of outputs is created by randomly selecting values

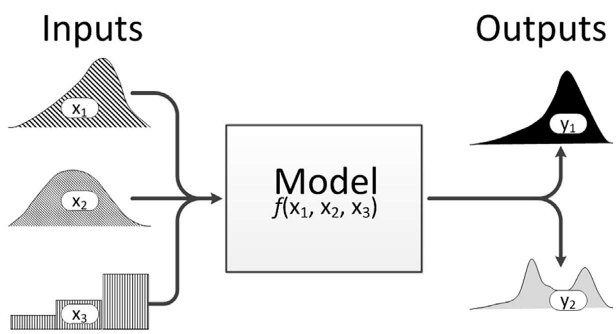

Fig. 1 – Illustrative example of Monte Carlo model interactions.

from all of the input distributions to calculate results in an iterative fashion.

A Monte Carlo method was selected, as it is suitable for highly stochastic data and has previously been shown to estimate cookstove performance accurately [14,18]. Guidelines published by the U.S. Environmental Protection Agency (EPA) were applied to help inform development of the Monte Carlo framework used here [19].

2. Methods

2.1. *Modeling overview*

A Monte Carlo model was developed to estimate the distribution of performances for Tier 3 and Tier 4 (defined by the ISO IWA 11:2012) [5] cookstoves that are going to be deployed in the field. Cookstove performance can be evaluated in many ways. We defined performance as the concentration of indoor carbon monoxide (CO) produced by the cookstove. Parameters included stove emission rates, sources of variability that affected emissions (both known and estimated), and parameters that interacted with emissions to affect CO concentrations in the room. The model was applied to estimate the number of statistical replicates required to determine meaningful differences among different cookstoves tested in the field. The modeling approach, which included various submodels, is described below.

Although the methodology presented here could be used to compare any number of cookstoves, a decision was made to illustrate the approach using Tier 3 and Tier 4 cookstoves. This decision was not made arbitrarily. The Tier 4 designation is intended to represent an aspiration goal for cookstove programs. The Tier 3/4 boundary is based upon World Health Organization (WHO) recommendations for personal exposure to carbon monoxide and particulate matter [5]. To date very few (if any) cookstoves have been shown to reliably achieve Tier 4 performance [20]. If cookstove programs are going to use the Tier 4 designation to indicate a particular design meets WHO indoor air quality recommendations (and Tier 3 does not) it will be important that the distinction between the cookstove can be backed up with field data.

The first step to establishing an estimate of indoor CO concentrations was to define the distribution of CO emission rates. The model included a set of nominal emissions rates (i.e., each rate was specific for a given type or Tier of stove) (Table 1). The model also considered three sources of variability, which were stove variability, fuel variability, and user variability (Fig. 2). Each variability term was modeled as a finite probability distribution (Table 1). These distributions were sampled (following a typical Monte Carlo approach) to produce a subsequent distribution of pollutant-emissions rates. Carbon monoxide was selected as an illustrative case and is the only metric of performance explored here; however, a similar approach could be used to evaluate particulate matter and fuel use. Fuel use was excluded as it is fairly simple and trivial calculation (in comparison to emissions). Fuel use does not require accounting for room concentration, which greatly simplifies the calculation although the framework presented could certainly be used. Evaluating particulate matter would require all the same model parameters as carbon monoxide but would typically have greater variability in the emissions values used. This larger variability would only increase the same size required, as such the sample sizes calculated in the illustrative example represent a base-case scenario.

Estimates of indoor air quality were based on these emissions rates and an air quality model that considered two additional parameters: room volume and air-exchange rate (Fig. 3). These parameters were incorporated into a single-zone, mass-balanced box model to calculate steady-state CO concentrations. Box models require a number of assumptions, including that the pollutants are derived from a single source and are perfectly and instantaneously mixed. Johnson et al. applied and validated a similar approach against experimental data [14].

dC_i/d_t = (m_dot(t)−Q(t) * C_i)/V

where:

$\dot{m}$: emissions rate of pollutant i
V: room volume
Q: air-exchange rate
C: concentration of pollutant i in room
t: time

Table 1 – Model input data. Each parameter was defined as a normal distribution with mean and standard deviation.

Model input parameter	Model inputs ± 1 s.d.
Performance variability	Tier 3: $0.49\ g * min^{-1} \pm 0.005\ g * min^{-1}$ Tier 4: $0.42\ g * min^{-1} \pm 0.004\ g * min^{-1}$
Variability between users adjustment factor	1 ± 0.2
Fuel-moisture variability	15% ± 5%
Room-size variability	$30\ m^3 \pm 2.1\ m^3$
Air-exchange-rate variability	$15\ h^{-1} \pm 3\ h^{-1}$
Measurement error	1 ± 0.1

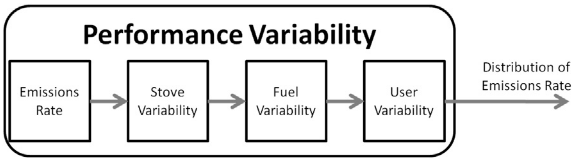

Fig. 2 – Schematic of performance variability sub-model.

2.2. *Model parameters*

Six input parameters were included in the model (Table 1). The values for each parameter (and the corresponding standard deviations) were established from published data where possible. The model assumed a Gaussian distribution for most parameters. In reality, many distributions were not Gaussian. However, because normal distributions have smaller standard deviations than other distributions, the number of required test replicates is minimized [21]. As such, this model represents a best-case scenario.

2.3. *Performance and user variability*

Cookstoves have variable performance levels. This variability can occur for many reasons, including slight differences in construction and operator-dependent factors such as cooking task. In addition, the user is often multitasking (examples could include watching children, cleaning, working on a home business, etc) which draws their attention away from the stove. These tasks often change day to day resulting in variability for each user. It is difficult to separate the relative contributions of cookstove and user variability. Our model accounted for these two sources of variability by including two input parameters. The first parameter accounted for variability associated with the physical cookstove. The results of tests conducted by Jetter et al. were used to establish an estimate of cooking variability based on the standard deviation of test replicates in a highly controlled testing environment. Data from tests in which the stoves were carefully tended were considered [20,22]. Cookstove variability was estimated from the most consistent tests performed by the EPA [20]. CO emissions rates fluctuated by more than an order of magnitude for different cookstove designs. Measurements included variations in stove performance and variations in measurements. For this study, it was assumed that the EPA data had very little measurement variability due to the highly controlled testing environment.

The second parameter for estimating cookstove variability was the difference between users. Biomass cookstove performance depends strongly on how a stove is used; each individual uses a cookstove somewhat differently. The variability between users was estimated by comparing total carbon monoxide mass emissions from Water Boiling Test [23] results conducted at different facilities. Each laboratory independently operated a traditional three-stone fire [20,24,25]. Variability between users was estimated by the standard deviation of these independent studies. Due to limitations in the data that is currently available estimates of user variability also includes instrument measurement variability and physical stove variability. The data to decouple these factors is not currently available; however, future work is planned to investigate variability between users when all tests are conducted using the same emissions hood and the same cookstove. The estimate for user variability is likely conservative. Real world variability would typically be greater than the estimates used expect in highly controlled and prescriptive style field tests.

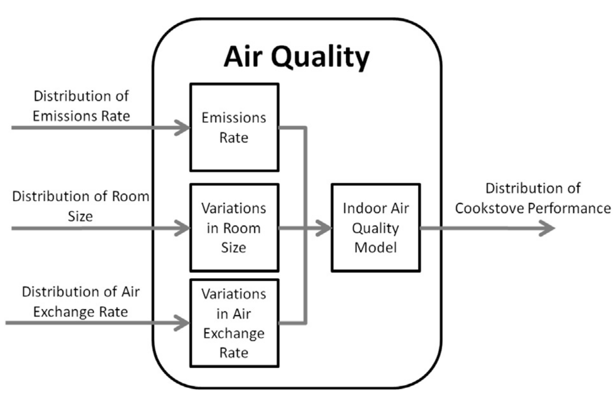

Fig. 3 – Schematic of the sub-model for indoor air quality.

2.4. *Fuel-moisture variability*

Emissions from a cookstove are strongly influenced by the type of fuel, particularly the moisture content [10,20,26]. The emissions rates will increase above or below some optimum value of fuel moisture [26,27]. Fuel moisture content is the mass fraction of water contained in fuel [28]. The content of fuel moisture depends on many factors, including how long a fuel has been drying, ambient temperature, humidity, and how the fuel has been stored. Within the same community, fuel which has been left in the rain can have substantially higher moisture content then fuel stored in a warm and dry location. A prediction of fuel-moisture content (for seasoned fuel) was established by considering a range of ambient temperatures and humidity conditions [29]. An adjustment factor for the effect of fuel-moisture content on emissions was established by considering empirical data on emissions at different moisture contents [26,27] (Fig. 4). By knowing typical fuel moisture contents seen in the field and the effect of fuel moisture on emissions, an adjustment factor was calculated.

2.5. *Room size and air-exchange rate*

The volume and air-exchange rate of a room (or home) vary with climate and location. Representative room sizes and air-exchange rates were established through a literature review [30–43]. A baseline room size and air-exchange rate was defined by considering the average room size in multiple locations around the world. Input distributions were established for room volume and air-exchange rate by considering the variability that is typically observed at a single location [33,39,42].

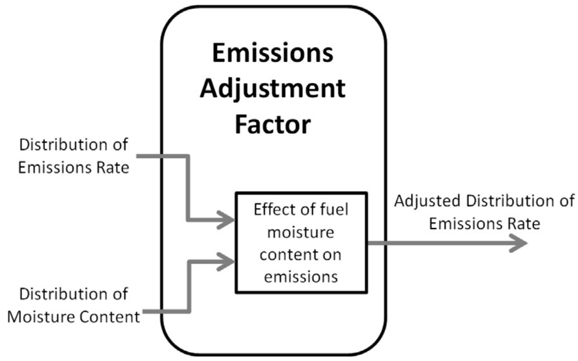

Fig. 4 – Emissions rate adjustment factor based on the effects of fuel-moisture content on emissions and the distribution of moisture contents. The emissions adjustment factor was found by weighting the effect of fuel moisture content on emissions rates [21,22] with the probability of different fuel-moisture content occurring [23].

2.6. *Measurement error*

All instruments have some uncertainty/error associated with them. Many factors, including instrument quality, how the instrument is operated, and maintenance of the instrument, can affect measurement error. The accuracy of the instrument can also depend on the magnitude of the reading. Our model included a conservative assumption that the variability associated with measurement error had a normal distribution centered at approximately 0% (i.e., no measurement error) with a relative standard deviation of 10%. Instrument error can vary widely, but 10% was selected based on anecdotal evidence.

2.7. *Calculating the required sample size*

The numerical model included the input data presented above and was applied to estimate the number of test replicates needed to distinguish the performance of a Tier 3 cookstove from a Tier 4 cookstove in the field. Two sample sizes were calculated. The first simulation determined the number of field tests needed to demonstrate that Tier 3 and Tier 4 cookstoves were statistically different based on room-level CO concentrations. The second simulation determined the number of tests needed to have confidence that a sample mean was a close approximation of the true population mean for indoor CO concentrations.

The number of replicates required to prove a statistical difference was calculated with *p*-values from Kolmogorov-Smirnov test. A detailed explanation of the Kolmogorov–Smirnov method can be found in Wang et al. [9]. *p*-values were calculated for sample sizes ranging from 3 to 300 replicates per cookstove. For each sample size, iterations were randomly selected for each cookstove, and a *p*-value was calculated. This process was repeated 100 times for each sample size. The objective was to determine the sample size that led to at least 80% of the calculated *p*-values to be less than 0.05. The selection of '80% power' was selected for this study based on convention [44, 45]; however, the statistical power required may depend on the analysis being conducted. A higher statistical power may be used if there is a need for greater confidence in the results.

A similar process was used to determine the number of samples required to estimate the average room-level CO concentration for a population households using a particular cookstove. For each distribution, the mean of the sample sizes ranging from 3 to 300 was calculated. One hundred random samples were selected at each sample size. These simulations were conducted to determine the sample size for which 80% of the sample means were within 5% of the true population mean.

3. Results and discussion

The Monte Carlo simulation revealed that Tier 3 and Tier 4 cookstoves produced very similar distributions of room-level CO concentrations despite having distinctly different emissions rates (Fig. 5). The overlap of CO concentrations was due to the compounding effect of the different sources of variation. The overlap of the concentrations has a number of important implications. First, these results suggest it would be challenging to identify differences between Tier 3 and Tier 4 cookstoves in the field. The broad distributions and the overlap in concentrations is one possible explanation for why many field studies have not found the emissions reductions expected from improved cookstoves. Although the Tier 4 cookstove resulted in lower average concentrations, statistically proving this would be challenging. Second, the broad distribution of results suggests that repeated studies of the same cookstove design could measure drastically different indoor air concentrations. If two studies were conducted on the same cookstove design but in different communities or during different seasons, their findings could be drastically different. The variability seen in studies could be partially explained by minor differences in situation variables or fuel conditions that may not initially seem important.

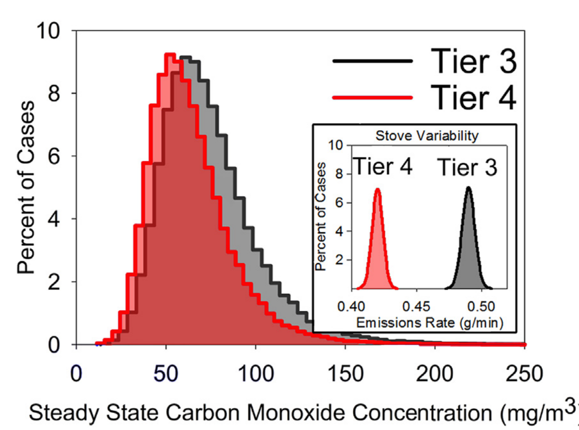

Fig. 5 – Histogram of estimated steady-state carbon monoxide concentrations produced by Tier 3 and Tier 4 cookstoves. Shown in the insert are the original emissions distributions of the Tier 3 and Tier 4 cookstoves.

3.1. *Sample size required for statistically significant results*

Due to the large overlap between distributions, many test replicates were required to differentiate the distributions statistically. The probability of finding a statistically significant difference between Tier 3 and Tier 4 stoves for different sample sizes is shown in Fig. 6. The model estimated that each cookstove would need to be tested 63 times (i.e., 126 tests for two cookstove designs) to show that they are statistically different 80% of the time. The combination of 126 tests would determine if the Tier 3 and Tier 4 cookstoves were different, but would not determine how they actually performed. If the testing were intended to quantify the average indoor concentration of CO resulting from the cookstoves, approximately 86 test replicates would be needed for each cookstove to equal 172 tests for the two stoves (Fig. 7). One hundred and seventy-two test replicates would be required for both Tier 3 and Tier 4 cookstoves to have sample means within 5% of the population mean 80% of the time. We selected 80% power based on convention; however, many cookstove evaluation programs require greater confidence. To achieve 95% power, 160 test replicates would be required per cookstove to differentiate two stoves; 180 replicates would be needed to quantify the mean concentration of CO produced by each stove.

3.2. *Identifying dominant factors affecting performance variability*

Understanding the factors that control field-testing variability can help cookstove project leaders to design studies that require fewer test replicates. Regression plots have been used to identify the major sources of variability (Fig. 8). For example, the dominant contributors to testing variability in our model were fuel-moisture content, air-exchange rate, and user variability. Thus, strategies that address these factors would have the greatest effects on reducing the number of test replicates required.

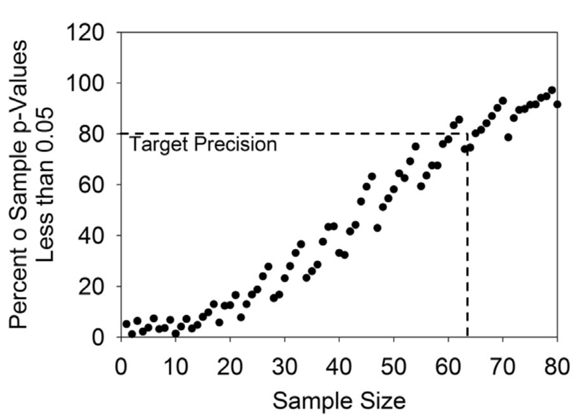

Fig. 6 — Study power as a function of sample size required to distinguish a Tier 4 from a Tier 3 cookstove in the field. Approximately 63 samples would be required for each of the two stoves (126 samples total) to achieve statistically significant differences in the measured means 80% of the time.

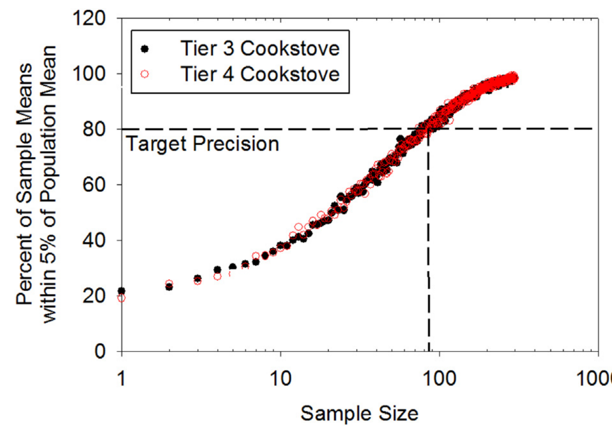

Fig. 7 — Sample size required for 80% of the sample means to be within 5% of the population mean. Approximately 86 test replicates would be required for both Tier 3 and Tier 4 cookstoves to have sample means within 5% of the population mean 80% of the time.

How testing variability is addressed depends on the specific goals of a cookstove evaluation program. To accurately measure real-world personal exposures, a study might not need to control any of the testing parameters. In contrast, a study of practical size that is designed to compare the performances of two cookstoves would probably require some of the testing parameters to be controlled.

Although there might be large variability in a specific model parameter, this variability may not have a large effect on the overall performance distribution. For example, measurement error was highly variable in our model (Table 1) but was a minor contributor to the variability in overall performance (Fig. 8f). Nevertheless, minimizing measurement error is critically important for quantifying the true population mean. In our model, we assumed that the instrument [1]: could over-measure or under-measure by the same amount, and [2] did so by the same amount at all concentrations. Neither of these assumptions is typically true. An instrument that systematically over-measures or under-measures, or one with a measurement error that is concentration-dependent, will create bias in the results. This bias is often difficult to detect.

Completely separating user variability and stove variability was not possible as different copies of stoves were used at each testing facility. To separate stove variability and user variability would require each facility to test with the same physical cookstove. Although this was not possible within the scope of this study, round-robin testing has been proposed by the Global Alliance for Clean Cookstoves [46,47]. The findings of the Alliance's work will help clarify how much variability can be attributed to individuals vs. attributed to physical stove variations.

The model was applied with a steady-state assumption. Therefore, transient factors, such as cooking practices or non-steady room air-exchange rates, were not considered. The effect of excluding time as a factor is important to interpret the results of the model correctly. Time can influence numerous aspects of a cookstove, including performance, air-

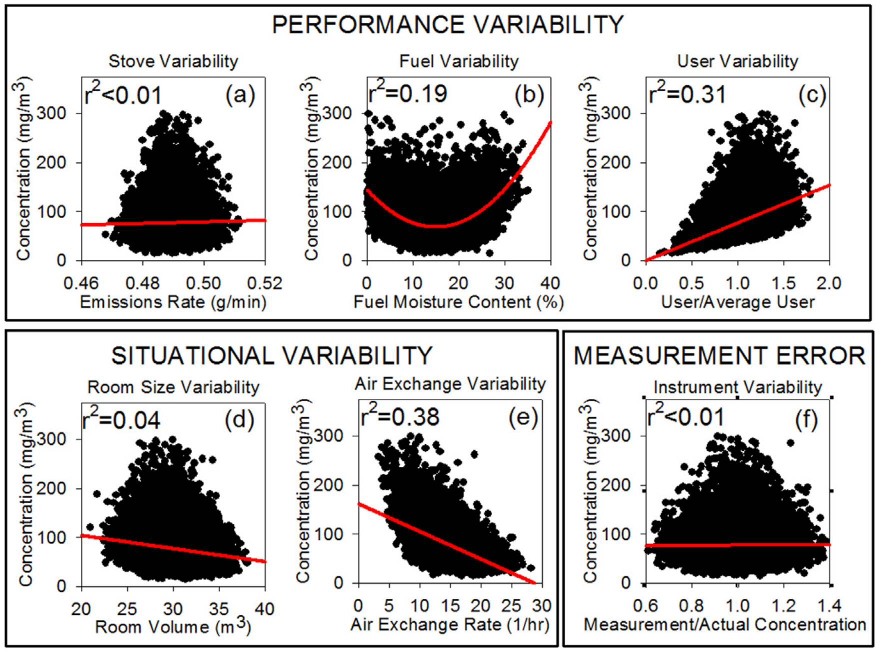

Fig. 8 – Correlations between steady-state concentrations and select sources of variability. A linear correlation was assumed in all cases except (b). A second-order polynomial was fit to (b) due to the non-linear nature of the correction factor for fuel-moisture content.

exchange rate in the home, and fuel-moisture content. Cookstove performance can change as a stove ages due to durability issues or lost components. In addition, air-exchange rates in homes often vary with seasons because families open or close windows due to changes in weather. Equilibrium of fuel-moisture content is also strongly dependent on ambient humidity. The distribution of many of the input parameters would depend on the time scale that is considered. However, studies are often conducted within a short time period, such as within one 24-h period or within one climate season. When evaluating cookstove performance within a narrow period of time, these additional factors will typically be less variable and therefore less influential.

Table 2 – Sensitivity analysis for predicting the number of test replicates required to quantify cookstove performance.

Model input parameter	Coefficient of variation used in sensitivity analysis (%)	Increase in sample size required to quantify cookstove performance
Performance variability	1.3	+2
Fuel-moisture variability	44.0	+35
User variability	26.7	+23
Room-size variability	9.3	+5
Air-exchange rate variability	26.7	+21
Measurement error	13.3	+8

3.3. *Model sensitivity to input parameters*

A sensitivity study was conducted to evaluate the effects of changes to specific input parameters on CO concentration distributions and sample size calculations. To investigate the feasibility of field testing, the range of performance (i.e., standard deviation) is more important than the mean. Input parameters were adjusted independently by increasing the standard deviation of each input distribution by 33%. Samples sizes were then recalculated.

The sensitivity analysis in Table 2 indicated that, although increased variability in model input parameters increased the required sample size in all cases, the conclusions and trends remained consistent. As expected, the dominant variables that led to testing variability were also the variables to which the model was most sensitive, i.e., fuel-moisture content, user variability, and air-exchange rate. This finding suggests that parameters should be established conservatively and broad input distributions should be included. Although the calculated sample size changed by as much as 30% in some cases, the number of test replicates required to produce statistically significant results would remain impractically large for many cookstove evaluation programs.

4. Conclusions

There will always be a need for testing cookstoves in the field. However, the sampling realities inherent to field testing results in a number of limitations. Interpreting results from field tests can be challenging due to multiple sources of variability that cannot be controlled. Many factors cause variable field measurements, restricting what conclusions can be drawn from these tests. As illustrated by the results presented above, field testing often may not be appropriate for quantifying population means or definitively proving two cookstove models result in different exposures. Although real and statistically different exposures existed in the simulated field results, few studies would have conducted enough samples to prove this difference. However, this does indicate field testing is not important; if anything it illustrates the importance of field testing. Field testing can provide (when the study is well designed) very important information such as how cookstoves are actually being used and consumer preferences among many other things. A field study which finds that the performance of two cookstoves overlap indicates that the performance of those stoves may not be substantially different, even if the differences are statistically different.

Theoretical case studies have been explored here to determine the number of test replicates needed to obtain statistically robust quantitative results from field tests. These studies showed that, even when conservative values were established for input variability, the number of replicates required to answer some questions could be problematic for many studies. Performing the number of test replicates needed to quantify cookstove performance will be difficult for many cookstove programs, especially for determining if two cookstoves produce different levels of pollutants. Cookstove programs with limited time or funding may not be able to devote the resourced required to collect the number of samples needed to answer some questions. Some of these questions may be better answered in the laboratory in combination with a numerical approach, such as the one presented here.

Designing and conducting field tests with cookstoves is an extremely complex process. The challenges and limitations need to be understood before beginning a project. Although field testing is a critical component of a cookstove program, the feasibility of performing field tests to fully quantify cookstove performance is questionable. There is a need for both laboratory and field testing; however, the two methods cannot be used interchangeably to answer some questions. To make useful connections between laboratory and field tests, more realistic views of the variability and limitations of field and laboratory testing are needed.

Acknowledgments

This work has been supported by the Colorado State University Engines and Energy Conversion Laboratory with funding from Envirofit International.

REFERENCE

[1] Ramanathan V. Role of black carbon on global and regional climate change. Testimonial to the house committee on oversight and government reform chair: the honorable Henry A Waxman. Washington DC, United States of America: House Committee on Oversight and Government Reform; 2007.

[2] Smith K, Balakrishnan K. Mitigating climate, meeting MDGs, and moderating chronic disease: the health co-benefits landscape. London, United Kingdon: Commonwealth Health Ministries; 2009. ISBN: 978-C0-85092-872-3.

[3] World Health Orginization. Indoor air pollution – killer in the kitchen. [updated 2004 Oct 14; cited 2014 Oct 17]. Available from: http://www.who.int/mediacentre/news/statements/2004/statement5/en/.

[4] World Health Orginization. Global health risks – mortality and burden of disease attributable to selected major risks. Geneva, Switzerland: World Health Orginization; 2009. ISBN: 978-92-4-156387-1.

[5] ISO. In: IWA 11:2012–guidelines for evaluating cookstove performance. ISO; 2012.

[6] Edwards R, Hubbard A, Khalakdina A, Pennise D, Smith K. Design considerations for field studies of changes in indoor air pollution due to improved stoves. Energy Sustain Dev 2007;11(2):71–81.

[7] Lwanga S, Lemeshow S. Sample size determination in health studies – a practical manual. World Health Orginization; 1991. ISBN: 92-4-154405-8.

[8] The Gold Standard Foundation. Technologies and practices to displace decentralized thermal energy consumption. [updated 2011 Nov 4; cited 2014 Oct 17]. Available from: http://www.goldstandard.org/wp-content/uploads/2011/10/GS_110411_TPDDTEC_Methodology.pdf.

[9] Wang Y, Sohn MD, Wang Y, Lask KM, Kirchstetter TW, Gadgil AJ. How many replicate tests are needed to test cookstove performance and emissions?—Three is not always adequate. Energy Sustain Dev 2014;20:21–9.

[10] Bhattacharya S, Albina DO, Myint Khaing A. Effects of selected parameters on performance and emission of biomass-fired cookstoves. Biomass Bioenergy 2002;23(5):387–95.

[11] Roden CA, Bond TC, Conway S, Pinel ABS, MacCarty N, Still D. Laboratory and field investigations of particulate and carbon monoxide emissions from traditional and improved cookstoves. Atmos Environ 2009;43(6):1170–81.

[12] Masera O, Edwards R, Arnez CA, Berrueta V, Johnson M, Bracho L, et al. Impact of patsari improved cookstoves on indoor air quality in Michoacán, Mexico. Energy Sustain Dev 2007;11(2):45–56.

[13] Zhang J, Smith KR, Uma R, Ma Y, Kishore VVN, Lata K, et al. Carbon monoxide from cookstoves in developing countries: 2. Exposure potentials. Chemosphere 1999;1(1–3):367–75.

[14] Johnson M, Lam N, Brant S, Gray C, Pennise D. Modeling indoor air pollution from cookstove emissions in developing countries using a Monte Carlo single-box model. Atmos Environ 2011;45(19):3237–43.

[15] Taylor J. In: McGuire A, editor. Introduction to error analysis, the study of uncertainties in physical measurements. 2nd ed. Sausalito, CA: University Science Books; 1997.

[16] Franks KA. Monte Carlo model of snaptic signaling [Dissertation]. San Diego (CA): University of California; 2002.

[17] Glasserman P. In: Rozovskii B, editor. Monte Carlo methods in financial engineering. New York, NY: Springer; 2004.

[18] Jeuland MA, Pattanayak SK. Benefits and costs of improved cookstoves: assessing the implications of variability in health, forest and climate impacts. PloS One 2012;7(2):e30338.

[19] U.S. Environemental Protection Agency. Guiding principles for Monte Carlo analysis. Risk assessment forum. Washington DC: U.S. Environmental Protection Agency; 1997.

[20] Jetter J, Zhao Y, Smith K, Khan B, Yelverton T, DeCarlo P, et al. Pollutant emissions and energy efficiency under controlled conditions for household biomass cookstoves and

implications for metrics useful in setting international test standards. Environ Sci Technol 2012;46(19):10827–34.
[21] NIST/SEMATECH. Standard deviations from assumed distributions. e-handbook of statistical methods. [updated 2003 June; cited 2014 Oct 7]. Available from: http://www.itl.nist.gov/div898/handbook/mpc/section5/mpc541.htm.
[22] Jetter J, Kariher P. Solid-fuel household cook stoves: characterization of performance and emissions. Biomass Bioenergy 2009;33(2):294–305.
[23] Bailis R, Ogle D, MacCarty N, Still D. The water boiling test (WBT). School of Public Health. [cited 2007 Jan; cited 2014 Oct 7]. Available from: http://ehs.sph.berkeley.edu/hem/content/WBT_Version_3.0_Jan2007a.pdf.
[24] MacCarty N, Still D, Ogle D. Fuel use and emissions performance of fifty cooking stoves in the laboratory and related benchmarks of performance. Energy Sustain Dev 2010;14(3):161–71.
[25] Hawley B, Volckens J. Proinflammatory effects of cookstove emissions on human bronchial epithelial cells. Indoor Air 2013;23(1):4–13.
[26] L'Orange C, DeFoort M, Willson B. Influence of testing parameters on biomass stove performance and development of an improved testing protocol. Energy Sustain Dev 2012;16(1):3–12.
[27] Baldwin S. Biomass stoves: engineering design, development, and dissemmination. Arlington, VA: Volunteers in Technical Assistance; 1987 [updated 2006 March; cited 2014 Oct 7]. Available from: http://www.aprovecho.org/lab/images/stories/camp08/BiomassStoves.pdf.
[28] ASTM. ASTMD4442-92: standard test methods for direct moisture content measurement of wood and wood-base materials. ASTM International; 2003.
[29] Simpson W. Equilibrium moisture content of wood in outdoor locations in the United States and worldwide. US Dept. of Agriculture, Forest Service, Forest Products Laboratory; 1998. FPL-RN-0268.
[30] Clark M, Reynolds S, Burch J, Conway S, Bachand A, Peel J. Indoor air pollution, cookstove quality, and housing characteristics in two honduran communities. Environ Res J 2010;110(1):12–8.
[31] Smith K, Aggarwal A. Air pollution and rural biomass fuels in developing countries: a pilot village study in India and implications for research and policy. Atmos Environ 1983;17(11):2343–62.
[32] Park E, Lee K. Particulate exposure and size distribution from wood burning stoves in costa rica. Indoor Air 2003;13(3):253–9.
[33] Raiyani C, Shah S, Desai N, Venkaiah K, Patel J, Parikh D, et al. Characterization and problems of indoor pollution due to cooking stove smoke. Atmos Environ 1993;27(11):1643–55.
[34] Albalak R, Bruce N, McCracken JP, Smith KR, De Gallardo T. Indoor respirable particulate matter concentrations from an open fire, improved cookstove, and LPG/open fire combination in a rural guatemalan community. Environ Sci Technol 2001;35(13):2650–5.
[35] MacCarty N, Still D, Ogle D, Drouin T. Assessing cook stove performance: field and lab studies of three rocket stoves comparing the open fire and traditional stoves in tamil nadu, india on measures of time to cook, fuel use, total emissions, and indoor air pollution. Aprovecho Research Center. [updated 2008 Jan; cited 2014 Oct 7]. Available from: http://www.aprovecho.org/lab/rad/rl/perf-stud/doc/3/raw.
[36] Menon P. Indoor spatial monitoring of combustion generated pollutants (TSP, CO, and BaP) by Indian cookstoves [Dissertation]. Honolulu (HI): University of Hawaii at Manoa; 1988.
[37] Gachanja A, Worsfold P. Monitoring of polycyclic aromatic hydrocarbon emissions from biomass combustion in Kenya using liquid chromatography with fluorescence detection. Sci Total Environ 1993;138(1):77–89.
[38] Brauer M, Bartlett K, Regalado-Pineda J, Perez-Padilla R. Assessment of particulate concentrations from domestic biomass combustion in rural Mexico. Environ Sci Technol 1995;30(1):104–9.
[39] McCracken J, Smith K. Emissions and efficiency of improved woodburning cookstoves in highland gatemala. Environ Int 1998;24(7):739–47.
[40] Gao X, Yu Q, Gu Q, Chen Y, Ding K, Zhu J, et al. Indoor air pollution from solid biomass fuels combustion in rural agricultural area of Tibet, China. Indoor Air 2009;19(3):198–205.
[41] Davidson C, Lin S, Osborn J, Pandey M, Rasmussen R, Khalil M. Indoor and outdoor air pollution in the Himalayas. Environ Sci Technol 1986;20(6):561–7.
[42] Cowlin S. Tracer decay for determining kitchen ventilation rates in San Lorenzo, Guatemala. Maxwell student projects, Max-04–4, EHS. School of Public Health, University of California: Berkeley. [updated 2005 April; cited 2014 Oct 7]. Available from: http://ehs.sph.berkeley.edu/guat/publications/Maxwell/Cowlin%2004-4.pdf.
[43] Albalak R, Keeler G, Frisancho A, Haber M. Assessment of PM10 concentrations from domestic biomass fuel combustion in two rural bolivian highland villages. Environ Sci Technol 1999;33(15):2505–9.
[44] Murphy K, Myors B, Wolach A. Statistical power analysis: a simple and general model for traditional and modern hypothesis tests. New York, NY: Routledge; 2014.
[45] Lenth R. Some practical guidelines for effective sample size determination. Am Statistician 2001;55(3):187–93.
[46] Chiang R. Standards and testing updates. Global alliance for clean cookstoves. November/December 2013 [updated 2013 Dec 20; cited 2014 Oct 7]. Available from: http://community.cleancookstoves.org/discussions/viewtopic/22/233?post_id=432.
[47] Request for applications: field studies to evaluate performance, use, and adoption of clean cookstoves and fuels. Global alliance for clean cookstoves. [updated 2012 Aug 26; cited 2014 Oct 7]. Available from: http://www.cleancookstoves.org/funding-opportunities/fieldtesting_rfa-update-2.pdf.