Received: 21 December 2020 | Accepted: 24 December 2020
DOI: 10.1111/ina.12790

ORIGINAL ARTICLE

WILEY

Modeling approaches and performance for estimating personal exposure to household air pollution: A case study in Kenya

Michael Johnson[1] | Ricardo Piedrahita[1] | Ajay Pillarisetti[2] | Matthew Shupler[3] | Diana Menya[4] | Madeleine Rossanese[1] | Samantha Delapeña[1] | Neeraja Penumetcha[5] | Ryan Chartier[6] | Elisa Puzzolo[3,7] | Daniel Pope[3]

[1]Berkeley Air Monitoring Group, Berkeley, California, USA
[2]Gangarosa Department of Environmental Health, Rollins School of Public Health, Emory University, Atlanta, Georgia, USA
[3]Department of Public Health and Policy, University of Liverpool, Liverpool, UK
[4]Department of Epidemiology and Medical Statistics, School of Public Health, College of Health Sciences, Moi University, Eldoret, Kenya
[5]Clean Cooking Alliance, Washington, DC, USA
[6]RTI International, Research Triangle Park, North Carolina, USA
[7]Global LPG Partnership, London, UK

Correspondence
Michael Johnson, Berkeley Air Monitoring Group, Berkeley, California, USA.
Email: mjohnson@berkeleyair.com

Funding information
United Nations Office for Project Services; Climate and Clean Air Coalition; Clean Cooking Alliance; National Institute for Health Research; HERCULES Center

Abstract

This study assessed the performance of modeling approaches to estimate personal exposure in Kenyan homes where cooking fuel combustion contributes substantially to household air pollution (HAP). We measured emissions ($PM_{2.5}$, black carbon, CO); household air pollution ($PM_{2.5}$, CO); personal exposure ($PM_{2.5}$, CO); stove use; and behavioral, socioeconomic, and household environmental characteristics (eg, ventilation and kitchen volume). We then applied various modeling approaches: a single-zone model; indirect exposure models, which combine person-location and area-level measurements; and predictive statistical models, including standard linear regression and ensemble machine learning approaches based on a set of predictors such as fuel type, room volume, and others. The single-zone model was reasonably well-correlated with measured kitchen concentrations of $PM_{2.5}$ (R^2 = 0.45) and CO (R^2 = 0.45), but lacked precision. The best performing regression model used a combination of survey-based data and physical measurements (R^2 = 0.76) and a root mean-squared error of 85 μg/m^3, and the survey-only-based regression model was able to predict PM2.5 exposures with an R^2 of 0.51. Of the machine learning algorithms evaluated, extreme gradient boosting performed best, with an R^2 of 0.57 and RMSE of 98 μg/m^3.

1 | INTRODUCTION

Globally, nearly 3 billion people burn solid fuels (eg, wood, dung, charcoal) in inefficient and poorly vented combustion devices (ie, open fires, traditional stoves) to meet daily cooking needs[1] and exposure to the resulting household air pollution (HAP) is a leading risk factor for global morbidity and mortality.[2] There are a relatively limited number of studies on personal exposures to HAP[3,4] given these health implications as well as the variety of interventions and contexts for which exposure data is needed. The lack of personal exposure data has in turn limited our understanding of health implications and the potential benefits of transitioning to cleaner cooking technologies and fuels. Field studies of personal exposure to $PM_{2.5}$ (particulate matter <2.5 microns in aerodynamic diameter) are some of the most difficult to conduct, as they require costly equipment, highly trained technicians, and participant compliance with monitors. These characteristics make them difficult to execute in low-income settings, especially as study households are often spread out across large areas.

Currently, the available HAP exposure models, such as those used for global burden of disease estimates, have relied on simple proxies, such as primary fuel type combined with relationships to the kitchen area concentrations,[5,6] and those which have made use of survey data have shown varying degrees of predictive power.[7–9] Recent efforts by Sanchez et al.[7] (peri-urban South India) and Hill et al.[8] found survey and machine learning models and survey-based regression models were able to predict HAP exposure with R^2 values of up to 0.25 and 0.3, respectively. Shupler et al.[5] modeled global HAP exposures using kitchen exposure ratios and reported R^2 values of 0.22. Microenvironmental models, which weight concentrations based on time spent in a given environment, have performed better (eg, Liao et al.[10] [Spearman $\rho = 0.83$] and Balakrishnan et al.[11] [$R^2 = 0.50$]); however, this approach still requires specialized equipment and technical capacity.

This study aimed to build on those efforts by using a combination of new tools (eg, proximity sensors and machine learning models) and more comprehensive physical measurements (eg, emissions, stove usage, and air change rates) to help understand which approaches may improve upon the predictive capacity of past efforts. Specifically, we sought to: (i) adapt and refine current physical modeling approaches used by the World Health Organization (WHO)[12] and the International Organization for Standardization (ISO)[13] (for deriving air quality-based stove performance targets) to predict personal exposures to $PM_{2.5}$; and (ii) develop multivariable and machine learning models that predict personal $PM_{2.5}$ exposures based on physical and behavioral parameters. All of these modeling approaches are potentially important for estimating personal exposures to HAP. They may be used to support characterizing exposures being estimated as part of health research efforts (eg, exposure response and/or intention-to-treat studies for various health endpoints), or efforts to characterize population health impacts (eg, burden of disease estimates). The single-zone model also has practical implications for linking stove emission performance with indoor air quality and exposures, and thus characterizing its performance may aid future WHO and ISO efforts to apply the model for deriving performance targets.

We also sought to conduct the work in a location where the modeling approaches could have substantive applicability, and the data collected for the inputs could help fill knowledge gaps. In Kenya, traditional cookstoves and open fires are estimated to cause nearly 16 600 premature deaths every year, of which 4900 are children.[14] While 67% of Kenyan households rely primarily on biomass for their cooking needs,[15] the country also has a relatively strong market for modern biomass cookstoves and clean cooking energy sources. The Kenyan clean cooking market is supported by innovative consumer finance programs such as pay-as-you go systems and microfinance.[16] A national charcoal production and sale ban enacted in 2019 may be accelerating household energy transitions. Kenya has also lowered import taxes on cookstoves and on liquefied petroleum gas (LPG), which is currently exempted from value-added tax.[17] There are limited data on stove emissions and HAP exposure in Kenya. To our knowledge, the only published field emission performance studies on stove interventions focused on charcoal and kerosene.[18,19] The data on personal exposures are also limited, with available studies reporting only carbon monoxide[20] or associated with only wood-fueled stoves.[21]

Practical Implications

Reliable and accurate models for estimating HAP exposure are valuable tools for researchers and program evaluators, and these models suggest a promising step forward. However, the substantial exposure contrast between the LPG and biomass user groups was largely responsible for the relatively good performance of the simple model, with fuel type being the most important predictor. This caveat implies that the model performance may rely on those large exposure contrasts, which are not always evident. Testing and further developing the modeling approaches in different contexts (fuels, geographies, stove use patterns, household environmental characteristics) would help characterize how robustly they operate and/or the degree to which they may be applied universally or tuned to specific conditions.

Here we present results from our model development and assessment in Kenya to predict HAP exposures. While not a substitute for direct field studies of exposure, these models could help guide programmatic decisions toward the most effective household energy solutions and enhance future HAP-related risk assessments.

2 | METHODS

2.1 | Study design and field site

Data collection for this modeling effort was conducted as a subsample of research being conducted by the NIHR CLEAN-Air(Africa) Global Health Research Group (GHRG) in Kenya, a project led by the University of Liverpool and Moi University in Eldoret, Kenya. The CLEAN-Air (Africa) GHRG works directly with government ministries in Cameroon, Ghana, and Kenya to help understand and estimate the potential public health and climate impacts of scaling adoption of clean cooking in the form of LPG to achieve announced policy targets.[22] Research by CLEAN-Air(Africa) in Kenya included rapid survey data on fuel usage and household characteristics from 2248 homes from Turbo and Kesses, rural/peri-urban communities in Western Kenya near Eldoret. The most common traditional stoves in the study area were traditional three-stone-fires, *jiko*-style charcoal stoves, and handmade mud stoves (*Chepkube* stoves; see Figure S2). None of the stoves had chimneys or other venting mechanisms. In-depth surveys were conducted in a subset of ~400 homes (approximately 200 exclusive biomass- and 200 primary LPG-using households) participating in the rapid survey. A subset of 100 of those homes (50 from each fuel use group) received personal exposure and household-level $PM_{2.5}$ and CO measurements along with monitoring of all their stoves. A further subset of 69 of the 100 households ($n = 57$ after removal of samples which had a damaged filter, flow

TABLE 1 Measurement and model overviews

Model	Description		
Single-zone (SZ)	Estimates kitchen and personal concentrations using physical, mass-balance approach		
Linear regression (LR)	Predicts personal exposure using multivariate linear models		
Kitchen exposure factor (KEF)	Uses the ratio of kitchen concentrations to personal exposure to predict personal exposures in homes where only kitchen concentrations were measured		
Machine learning (ML)	Predicts personal exposure using ensemble supervised learning techniques		
	Sampling period and instrument measures		
Measurement	**Emission measurements during a cooking event**	**24-h measurements**	**Intensive measurements (1–4 days)**
Personal exposure		RTI MicroPEM ($PM_{2.5}$) Lascar CO (CO) Berkeley Air Beacons (participant location) SZ, LR, KEF, ML	Berkeley Air Beacons (participant location)
Kitchen HAP		RTI MicroPEM ($PM_{2.5}$) Lascar CO (CO) SZ, LR, KEF, ML	Berkeley Air PATS+ ($PM_{2.5}$) Lascar CO (CO)
Secondary HAP		Berkeley Air PATS+ ($PM_{2.5}$) Lascar CO (CO) LR	Berkeley Air PATS+ ($PM_{2.5}$) Lascar CO (CO)
PM2.5 Emissions	UPAS ($PM_{2.5}$) Berkeley Air PATS+ ($PM_{2.5}$, placed at 1 m, 1.5 m, 2 m) SZ	-	-
Gas emissions (CO, CO2)	TSI (CO) Lascar CO (CO, placed at 1 m, 1.5 m, 2 m) SZ	-	-
Participant Behavior	Observation	Berkeley Air Beacon Loggers (participant location) Survey (time activity, fuel use, socioeconomic status) LR, ML	Berkeley Air Beacon Loggers (participant location) Survey (time activity, fuel use, socioeconomic status)
Stove use	Geocene SUMS (temperature loggers) SZ, LR, KEF, ML		
Environment	Kitchen volume, air exchange rate, housing characteristics SZ, LR, KEF, ML		
Ambient	UPAS ($PM_{2.5}$), PATS+ ($PM_{2.5}$), Lascar CO (CO) SZ, KEF		

A brief description of the various models is provided followed by the measures used by those models, which are cross-referenced with the applicable model's initials. Measurements were conducted at multiple time scales, from the measurement of single cooking events to multi-day household air pollution monitoring.

error, or runtime error, in accordance with those criteria described in Chartier et al.[23]), primarily using LPG (n = 32), wood (n = 32), or charcoal (n = 7), were selected for this study to receive additional measurements of emissions, room/stove proximity, and living area $PM_{2.5}$ and CO measurements. A summary of measurements, cross-referenced with the applicable models, is presented in Table 1.

Household selection and assignment to the study groups were informed by results from the rapid and in-depth surveys. Field measurements for this work were conducted from October 2019 to January 2020, encompassing the dry season. Although the climate throughout the year is generally temperate, and cooking occurs primarily indoors, it is conceivable that ventilation and fuel usage patterns may change over the seasons, influencing model performance.

Ambient sampling was conducted concurrently with household sampling in most cases. A subset of 28 homes were selected for more intensive sampling that included stove usage monitoring (SUM) for up to four months and HAP monitoring for up to four days (Table S4). During emission sampling in these homes, stratified kitchen CO and $PM_{2.5}$ concentrations were sampled at heights of 1 and 2 m, in addition to the typical 1.5 m height for kitchen concentrations. A map of the study area and installation photographs can be found in Appendix S1.

2.2 | Data collection and analysis methods

2.2.1 | Emission measurements

Emission samples were collected during uncontrolled cooking tests in participants' homes, where the cooks were instructed to prepare a meal as they normally would, without altering stove operation or cooking techniques. The emission sample was collected with a multi-port probe suspended in the smoke plume (Figure S3), and the sample stream was drawn through a Teflon filter after a $PM_{2.5}$ size cut cyclone to determine $PM_{2.5}$ mass deposition. Carbon dioxide (CO_2) and CO were measured with real-time instrumentation (TSI IAQ CALC 7545). Background concentrations of CO_2, CO, and $PM_{2.5}$ were measured in the kitchens for at least 5 minutes immediately before and at least 10 minutes immediately after each sampling event (see Figure S10) and subtracted from those measured in the emission plume. Samples with more than 15% of the CO or CO_2 readings above the instrument maximum measurable value were removed from analysis. If real-time measurements of $PM_{2.5}$ concentrations exceeded 50 µg/m^3 in the home before the emission test, testing was delayed until a lower background concentration was observed. Filter analyses were performed at Colorado State University (Fort Collins, CO, USA) using an electronic microbalance (Mettler Toledo, USA) with 0.1 µg resolution in a temperature and humidity-controlled chamber. Mass depositions were gravimetrically determined by weighing the filters before and after sampling and correcting for handling effects by using the median mass deposition of collected blank filters (n = 20, 5.8 µg [2.5% of the average deposition]). Limit of detection (LoD) was calculated as three times the standard deviation of field blanks (7.8 µg [3.6% of the average deposition]) of the mass deposition on the blank filters.[24] Black carbon was optically quantified by transmittance (before and after sampling) using a SootScan OT21 analyzer (Magee Scientific, Berkeley, CA, USA), and adjusted using calibration factors as reported in Garland et al.[18]

Emission factors were determined using the carbon balance approach, as has been done in previous studies of stove emissions and as is described in the ISO protocol for stove emission testing.[25–27] Emission rates were calculated by dividing the total emissions during a sampled stove use event by the amount of time the event lasted. Observations and measurements of operational conditions, which may affect emission performance, were also recorded for the duration of each cooking event, such as lighting techniques, pot types, and fuel conditions.

2.2.2 | Ventilation rate determination

Ventilation was measured via tracer gas method, according to the standard WHO protocols specifically designed for the single-zone box model. Briefly, CO levels were elevated in the cooking area due to the emission source, and the natural log of the rate at which the gas decreased at the end of the cooking event was converted into air changes per hour (Figure S10).[28] The air change rates were calculated using data from the CO monitor placed at 1.5 m height, and we also assessed it using data from the monitors at 1 and 2 m for homes that had intensive sampling. In cases where the kitchen monitor did not provide valid data, the data collected by the emission monitoring system were used to estimate the ventilation rate.

2.2.3 | Stove usage monitoring

Stove usage was directly measured at 5-min intervals on all stoves in the study homes using Geocene stove use monitors[29] (SUMs) (Geocene, Vallejo, CA, USA) and participant surveys. To ensure proper SUM placement (~15 cm from center of flame), data were examined from weekly automated emails to check for temperatures exceeding the acceptable range (100–500°C) and immediate corrections were applied as needed. Placement of the SUMs was piloted on all stove types to ensure successful data collection (see Figure S8). We conducted one week to six months of SUMs monitoring, depending on feasibility given study logistics.

To generate cooking events from the raw SUMS temperature time series, this project used two versions of the Geocene FireFinder algorithm[28] (see Section S6 in Appendix S1). An example time series is shown in Figure S7. The time cooked with each stove type per day was a direct input into the single-zone model and was used to predict kitchen-level concentrations, as well as personal exposure.

2.2.4 | Ambient monitoring

Ambient monitoring of gravimetric and nephelometric $PM_{2.5}$ and real-time CO was carried out during most emission sample collection periods. An ambient monitoring station was designated in a rural background location in the Kesses region (0°25'07.7"N 35°19'24.4"E). Gravimetric and black carbon $PM_{2.5}$ measurements were collected (UPAS, Access Sensor Technologies, Fort Collins, CO, USA), alongside real-time $PM_{2.5}$ (PATS+, Berkeley Air Monitoring Group, Berkeley, CA, USA) and CO (Lascar EL-USB-300, Lascar Electronics, UK). Instrument inlets were placed at a height of 6 m, and away from trees, buildings, or other obstructions (see Figure S4), and there were no substantial nearby air pollution sources except for one biomass burning kitchen ~50 m from the site.

2.2.5 | Personal exposure monitoring

Personal exposure was measured for the primary cook using the RTI MicroPEM or RTI Enhanced Children's MicroPEM monitors (ECM, RTI International, Research Triangle Park, NC, USA), combined gravimetric and real-time nephelometric $PM_{2.5}$ monitors. MicroPEM (0.40 L/min, 50% duty cycle, 25 mm PTFE filter) and ECM (0.30 L/

min, 50% duty cycle, 15 mm PTFE filter) filter samples were pre- and post-weighed at RTI International and the field blank-corrected mass concentrations were used to post-correct the nephelometer readings, which are sensitive to aerosol physical properties (eg, size distribution, real and imaginary refractive indices) that may vary by source type and atmospheric conditions. The filters were equilibrated in a temperature and humidity-controlled weighing facility (21°C, 35% RH) prior to pre- and post-weighing on an ultramicrobalance (Mettler Toledo UMX2, 0.1 µg readability). Clean filter handling and MicroPEM calibration protocols were applied to ensure high data quality throughout the study. The built-in accelerometers were used to assess personal monitor wearing compliance via a rolling standard deviation method, and a 20-min rolling average was applied to the magnitude of the composite acceleration. Movement was recorded during 42% of the 24-h monitoring period (Table 4). Considering only daytime hours (5 a.m. to 9 p.m.) for the charcoal, Chepkube, LPG, and traditional stove groups, movement was recorded for 49%, 73%, 52%, and 63% of the time, respectively. These values are reasonably high given that participants did not wear the equipment while sleeping, bathing, or other activities when it is not practical. All data were retained regardless of compliance fraction.

Personal carbon monoxide exposure monitoring was performed using Lascar EL-USB-CO300 monitors. All CO sensors used in the study underwent two-point calibrations before the study with certified calibration standards in Berkeley Air Monitoring Group's California laboratory. Linear calibration corrections were applied to the data. The data were then manually reviewed as a sense check and filtered in cases of clear instrument malfunctions.

2.2.6 | Household air pollution monitoring

Household air pollution concentrations were also measured with the MicroPEM as part of the Kenyan CLEAN-Air (Africa) research. These data were supplemented with additional real-time PM and CO instrumentation (PATS+ and Lascar CO). In a subset of households, PATS+CO (Berkeley Air Monitoring Group, Berkeley, CA, USA) were used rather than the standard PATS+, as its integrated high-precision CO electrochemical sensor allowed for comparison with the Lascars. A sampling pack containing a MicroPEM, PATS+, and Lascar CO monitor was installed in the kitchen area for 24 h, including during the emission sampling event. PATS+ and Lascar CO monitors were also placed in a separate room where the participants reported spending the most time (generally the living area). In a subset of 28 households, a set of three evenly spaced Lascar CO monitors was also hung between the ceiling and floor (at 1, 1.5, and 2 m in height) in a stratified sampling configuration to capture the spatial variability of pollutants in the kitchen space, which informs the variability of the ventilation conditions within the room, and in turn the variability of kitchen to personal exposure estimation methods. In the 28 intensive homes, PATS+ and Lascar CO monitors in the kitchen and living areas were left installed for up to five days to assess day-to-day variability in HAP concentrations.

2.2.7 | Beacon-based time-activity monitoring

The Berkeley Air Beacon Logger System is a time-activity monitoring system specially designed for household energy applications.[10,30] The system is composed of two components, a coin-sized Bluetooth Beacon, which safely emits a unique ID multiple times a second over Bluetooth Low Energy, and a Beacon Logger, which records the address and the strength of the Beacon's emitted signal. The system components are low in cost, power consumption, and maintenance efforts, especially in comparison with personal exposure monitors.

Beacon Loggers were installed in all kitchens and living areas, and Beacons were given to the primary cooks to wear during the 24-h monitoring period from which emission data were captured. Users generally wore the two Beacons on a necklace or in the pocket alongside the exposure monitors. A minute-wise time series of presence in each microenvironment was generated for the user by associating the signal strength of their Beacons with the fixed locations of the Beacon Loggers (the primary kitchen and the living area where they spent most of their time during the day). Presence in a given location was assigned to the location with highest signal strength, on a minute-by-minute basis. If neither logger recorded a signal, the location was classified as "ambient."

A performance check of the system, called a walkthrough, was carried out at each home before the start of the deployment, to assess system performance. This entailed leaving the equipment in each area for a five-minute period, to determine whether the classification was correct (see S5). In the subset of intensive households, the Beacons and Beacon Loggers remained in place alongside PATS+ and Lascar CO microenvironment monitors for a period of up to five days. This longer-term period allowed us to assess participant acceptability of protocols, model performance, and compare day-to-day within-person variability to between-person variability.

2.2.8 | Behavioral factors

Time activity, fuel use, and cooking behavior data were collected using standard questionnaires that have been used both in this part of Africa and in other countries, including India, Mongolia, Laos, and Cambodia. Additional questions were added based on their potential utility to contribute explanatory power to statistical models and included parameters such as trash-burning, animal fodder preparation, and smoking habits. Additional information on socioeconomic status and educational status was evaluated. Socioeconomic status was assessed using principal component analysis (PCA) on asset ownership and home characteristic variables as per Vyas & Kumaranayake[31] to generate a five-category index. Households were assigned to a category using the prediction from the first principal component of the analysis. The first index categorized was associated with low ownership of assets (such as cars, smart phones, and computers), lower use of LPG, and outdoor sanitation facilities, while the fifth was characterized by high ownership rates of those assets, indoor sanitation facilities, and access to water indoors (Table S2). Only the socioeconomic status, basic kitchen characteristics, and primary fuel category were used in models. The remaining survey data are planned for use in subsequent analyses. Survey data were collected with Mobenzi Researcher (Cape Town, South Africa), a tablet-based data entry system that has been used extensively in similar studies[32] and minimizes the likelihood of transcription errors and data loss.

2.3 | Modeling approaches

The four primary modeling approaches are outlined below. Note that our primary metric of performance was R-squared, as it provides an indication of how well the model explains variability in the predicted measure, and also provides a basis of comparison with previously reported models. Given that the best performing previous models have reached R^2 values of approximately 0.5, this level provides a reasonable benchmark for comparison.

2.3.1 | Single-zone model

We used the single-zone, mass-based model employed by ISO and WHO to estimate kitchen $PM_{2.5}$ concentrations and derive emission performance targets.[12,13] The model predicts room concentrations of pollutants using input distributions of emission rates and usage times of the sources (in this case, stoves); a room's ventilation rate and volume; the fraction of emissions from the sources that enter the room (important for chimney stoves); and the background/ambient concentration. The mathematical description is provided below:

$$C(t) = \frac{q_1 f_1 + q_2 f_2 + q_3 f_3 + \ldots q_n f_n}{\alpha V}\left(1 - e^{-\alpha t}\right) + C_o\left(e^{-\alpha t}\right) + C_b \quad (1)$$

where

$C(t)$ = pollutant concentration for a given time point

q_x = emission rate for source x (mass/min)

f_x = fraction of emissions from source x that enters the kitchen environment

α = first-order loss rate (nominal air exchange rate) (changes/min)

V = kitchen volume (m^3)

t = time interval (1 min)

C_o = concentration from preceding time interval (unit mass/m^3)

C_b = background concentration

The model produces 24 h of minute-by-minute concentration estimates, where the emission rates for the respective sources are inputs for three discrete, evenly spaced cooking times. The sum of these periods is the device usage time, which is also a model input. To calculate the predicted 24-h mean concentration in the kitchen (C_k), the concentrations for each time point are summed over the day and divided by the number of minutes in a day. To estimate exposure, the 24-h mean concentration C_k was multiplied by a Kitchen Exposure Factor (KEF), as shown in Equation 2 below. The exposure ratios are ideally location-specific (as possible here), though global averages have been applied such as those from the Global Burden of Disease Study (0.742 for women, 0.628 for young children, 0.450 for men).[6] The personal exposure based on this ratio (E_r) is defined as follows:

$$E_r = C_k * \text{KEF} \quad (2)$$

Each set of household input data was run independently, so kitchen exposure ratios (*r* above) could be calculated for each sample and by stove-fuel group and compared against the respective kitchen and personal exposure concentrations. The model was then tuned based on the relationships between modeled and measured concentration during cooking events, and run through Monte Carlo simulations (Risk Analyzer software package, Add-ins LLC, DE, USA), to predict distributions of personal exposures for the different fuel use groups.

2.3.2 | Prediction using statistical models

In addition to the single-zone modeling of the first approach, $PM_{2.5}$ exposures were predicted using linear statistical models and ensemble or stacked modeling techniques. Covariates for both modeling approaches included sensor-based (indoor location, stove usage patterns, kitchen pollutant measurements) and survey (characteristics of the home, kitchen, fuel, etc.) data.

Linear regression

We developed models with different sets of covariates, beginning with those that are easiest to collect and most crude (survey data) and continuing with increasingly complex data streams. Previous work has shown that some statistical variation in exposure can be explained through data easily obtained via questionnaire.[33–35] We also compared models using both sensor-based measurements and modeled estimates of kitchen concentrations.

We used linear regression (*R* 3.6.2 and 4.0.2) to model $PM_{2.5}$ exposures among primary cooks. 50 households provided data that passed quality checks for use in predictive modeling. We imputed missing covariate data by using the stove-group specific median values. The dependent variable—the cook's measured exposure to $PM_{2.5}$—was log-transformed to meet normality requirements for linear models.

Univariate models assessed the relationship between exposure proxies and measured concentrations and exposures. The use of kitchen concentrations and ratios of concentrations to exposure were assessed with measures of correlation (Spearman's ρ). Multivariable models were used to assess the relationship between personal exposure and sociodemographic characteristics, stove-fuel energy use patterns, household characteristics, and other physical measurements (time activity, stove use, etc.) in the home. Variable selection occurred using multiple modalities. First, we used an automatic variable selection algorithm (from the “leaps” r package) to pick parameters that optimized between model comparison parameters, including adjusted R^2, Bayesian information criterion (BIC), and Malloy's C_p (shortened to C_p). Models identified using the automatic variable selection algorithm were further screened using 10-fold cross-validation. The model form that minimized RMSE during 10-fold cross-validation was selected for further evaluation. Second, based on our prior knowledge and a review of the literature, we evaluated sets of predictors that we anticipated would be easier to collect in the field using surveys, less intrusive monitoring devices placed in a kitchen, or less intrusive personal monitors.

Kitchen exposure factor

We investigated the utility of KEFs as potential exposure proxies. We evaluated the predictive power of three literature-based KEFs: the single estimate from the Global Burden of Disease and two estimates derived from Shupler et al.[5] We also evaluated the utility of subsets of our measured KEFs to predict "true" exposures. For each stove type, we sampled subsets (of size ranging from 1 to 20) of the total set of measured KEFs. For each size subset, 10 random samples were generated. We calculated the mean stove-specific KEF for each subset and estimated exposures by multiplying this mean by the corresponding kitchen concentration. We then calculated the RMSE, comparing the KEF-subset derived exposures to the measured personal exposures.

Machine learning

Finally, we evaluated machine learning models using SuperLearner (SL)[36] as a proof-of-concept, following the example of Hill et al.[8] SL uses cross-validation to estimate the performance of multiple types of machine learning models or parameterizations of the same type of model. The same samples and imputation for mission data were used as for the regression modeling, though exposure estimates were was not log-transformed as a normal distribution is not a required assumption for these machine learning models. SL as implemented created a cross-validated ensemble—a weighted average of these machine learning models—for prediction. Individual machine learning models and the ensemble performance were evaluated. All machine learning models utilized in SL were run with default parameters and 10-fold cross-validation. The ensemble learner was externally cross-validated (20-fold). Linear and SL model performance was compared using adjusted R^2 and RMSE.

3 | RESULTS

Results of the four primary modeling approaches are outlined below. Note that our primary metric of performance was *R*-squared, as it provides an indication of how well the model explains variability in the predicted measure, and also provides a basis of comparison with previously reported models. Given that the best performing previous models have reached R^2 values of approximately 0.5, this level provides a reasonable benchmark for comparison. Root mean standard error (RMSE) is also used as a measure of model precision.

3.1 | Single-zone model

Table 2 presents the summary statistics for the input parameters measured during cooking events for the single-zone model, as well as the corresponding kitchen concentrations. Ventilation (mean = 17.8 air changes per hour [ACH]; range = 6–73 [ACH]), kitchen volumes (mean = 21.1 m^3; range = 5–52 m^3), and cooking event durations (mean = 51; range = 7–125 min/event) were generally in line with those used by ISO and WHO,[12,13] as well as other similar modeling exercises.[37–39] Kitchen event concentrations (mean = 886 μg/m^3; range of 10–16 161 μg/m^3 for $PM_{2.5,}$ and mean = 28.6 ppm; range = 0–196 ppm for CO) and 24-h exposures for PM2.5 (mean = 135 μg/m^3; range = 14–686 μg/m^3) were reasonable given previous studies' ranges of 24-h exposures for these fuel user groups.[3,40] Emission rates were also similar to estimates for wood and charcoal stoves in the region.[19,27] Our estimated LoD for $PM_{2.5}$ emission rates was approximately 5 mg/min, which was greater than what we measured for LPG. We therefore have used the $PM_{2.5}$ emission rates for LPG reported by Weyant et al.[41] and Johnson et al.,[19] who were able to measure them in the field. Overall, the emission performance and pollutant concentrations were highly variable, ranging from very clean (LPG) to highly polluting (wood stoves), with charcoal in between. Ambient measurements of $PM_{2.5}$ and CO were also made (see Table S3) and showed consistently low levels (6.8 ± 5.4 μg/m^3 for $PM_{2.5}$, 0.9 ± 2.7 ppm for CO). Note that stove/fuel performance metrics not directly used in the modeling efforts, including combustion efficiency, emission factors ($PM_{2.5}$, CO, and black carbon), and firepower, can be found in Table S1.

The relationship between modeled and measured estimates of kitchen concentrations for $PM_{2.5}$ and CO is shown in Figure 1. There are clear positive correlations, following the anticipated trend of lower pollutant concentrations during LPG use, and higher during biomass stove use; however, there is considerable scatter ($PM_{2.5}$ model RMSE = 767 μg/m^3; CO model RSME: 30 ppm), with the model explaining 45% of the variability in the measured event concentrations of both $PM_{2.5}$ and CO ($PM_{2.5}$: R^2 = 0.45, p < 0.01; CO: R^2 = 0.45, p < 0.01). The model also overestimates the measured kitchen concentrations (~10-fold for $PM_{2.5}$) and (~6-fold for CO). This bias is similar to what was reported by Piedrahita et al.[42] and Johnson et al.,[43] who both found the model to overestimate measured concentrations in the kitchen. There are several reasons for this potential bias, the most likely being due to the model assumptions that all emitted pollutants instantaneously and perfectly mix throughout the room. Indeed, measurements of the stratified air pollution concentrations shown in Table S6 indicate substantial pooling of PM2.5 and CO higher in the kitchen. It is likely that a substantial fraction of emissions escape through windows, eaves, or other openings before mixing throughout the room, and mixing is incomplete, with higher concentrations pooling higher in the room. The variability in mixing and stratification of pollutants also likely contributes to the amount of scatter in the plots. It is also evident that modeled $PM_{2.5}$ emissions are clustered near the y-axis, potentially due to setting the LPG emissions rate to 1 mg/min (LoD). This potential artifact may be

TABLE 2 Summary statistics for cooking event measurements used in the single-zone model

	Mean	Median	SD	Range	*n*
Kitchen $PM_{2.5}$ (μg/m^3)					
LPG	135	92	130	10–531	28
Charcoal	855	642	970	28–2442	6
Wood	2048	1580	2870	71–16 161	29
Kitchen CO (ppm)					
LPG	4.8	2.6	5.1	0.0–15.6	27
Charcoal	70.4	63.8	73.9	5.9–150.2	4
Wood	44.3	38.2	39.8	0.4–196.1	28
$PM_{2.5}$ emission rate (mg/min)					
LPG	1[a]	NA	0.5[a]	0.1[a]–2.5[a]	NA
Charcoal	15	15	10	3–30	7
Wood	159	147	65	69–343	29
CO emission rate (g/min)					
LPG	0.04	0.05	0.02	0.001–0.20	30
Charcoal	1.87	1.41	0.87	1.1–3.1	7
Wood	1.68	1.50	0.79	0.50–3.6	29
Ventilation (air changes/h)					
LPG	14.3	12.2	8.3	5.5–40.0	30
Charcoal	27.0	19.0	21.4	12.1–72.6	7
Wood	18.3	17.5	7.9	7.1–38.7	32
Kitchen volume (m^3)					
LPG	16.5	13.2	12.2	5.4–51.9	32
Charcoal	24.8	23.2	9.6	12.8–41.5	6
Wood	25.0	22.3	11.5	11.1–49.6	32
Event duration (min)					
LPG	45	45	28	7–125	31
Charcoal	54	50	16	29–81	7
Wood	58	54	26	21–116	30

[a]Assumed from Weyant et al. 2019 and Johnson et al. 2019.

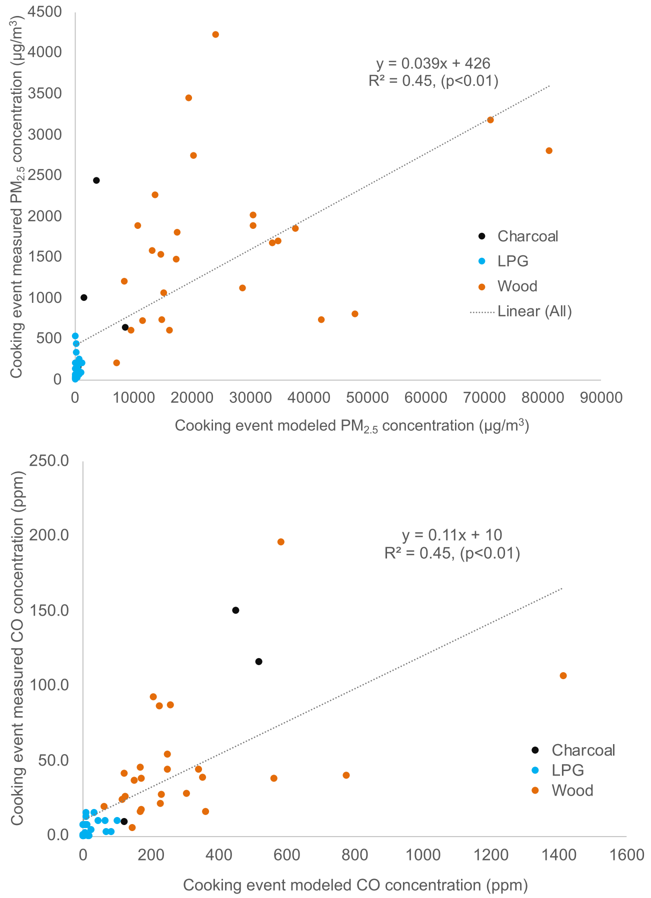

FIGURE 1 Relationships between the single-zone modeled and measured kitchen concentrations (PM top, CO below) during the sampled cooking events

shifting the overall relationship between the modeled and measured concentrations, resulting in a high y-intercept.

Our stratified CO samples show mean concentrations sequentially increased from 14 ppm at 1 m above the ground, to 20 ppm at 1.5 m above the ground (HAP standard protocol height), to 28 ppm at 2 m (presented by fuel type in Table S6). Stratified samples of CO by Johnson et al.[43] suggested a similar pattern, and MacCarty et al.[44] did a systematic investigation of the model's performance in a test kitchen, showing that $PM_{2.5}$ concentrations increased in an S-shaped curve, pooling at the ceiling. These two studies also found that a height of ~1.5 m was likely the best proxy height to capture the average room concentration and/or exposure of a standing adult.

To model 24-h $PM_{2.5}$ exposures, we applied simple correction factors (ratio of measured to modeled means: 0.07, 0.24, and 0.50 and for wood, charcoal, and LPG, respectively) to normalize model response to the measured kitchen concentrations and then applied the measured main cook KEFs (0.69, 0.85, and 0.83 for wood, charcoal, and LPG, respectively). The single-zone model was run through a Monte Carlo simulation (10 000 iterations) for each of these fuel user groups, defined by the greatest amount of a given fuel used during the day exposure was measured. All stoves used within the house for the given day were included.

The modeled distributions generally compared well with the measured 24-h $PM_{2.5}$ exposures. Figure 2 shows the fitted distributions, illustrating the overlap between the modeled and measured estimates (LPG in blue, charcoal in gray/black, wood in red/pink). The LPG and wood distributions compare most favorably, with their interquartile ranges overlapping substantially, and mean and medians reasonably close (see Table 3). The modeled exposures for charcoal were lower than that of the measured concentrations (mean and median values were ~70% and 40% of the measured estimates, respectively), though this comparison is the most tenuous as only seven samples were available for analysis.

Overall, this approach shows promise that the model can be applied to estimate distributions of $PM_{2.5}$ exposures, though care needs to be taken to ensure that inputs for normalizing the model account for bias. Given the scatter in the relationship of individual estimates, it is also not recommended to use the model for predicting specific households, but rather as a tool for understanding how group-level exposures may be impacted by changes in stove use, stove performance, environmental conditions or other parameters that may change over time or as a result of an intervention.

3.2 | Statistical models

3.2.1 | Linear regression

We selected parsimonious models with between 6 and 7 predictors out of the >20 evaluated to offer an optimal compromise between root mean-squared error (RMSE) and adjusted R^2. As with the single-zone model, we focused on predicting $PM_{2.5}$ due to the importance of its association with health impacts. We first present the summary statistics for the variables included in model selection in Table 4. Note that these values are slightly different from those from the single-zone modeling exercise, as the data completeness changed with the exclusion of the direct emission-related measurements. Summary findings for the regression models are shown in Table 5.

For the entire dataset, survey data alone—consisting of a measure of kitchen volume, the primary stove type, and a socioeconomic index comprised of assets, housing characteristics, and other variables (Table S2)—had an adjusted R^2 of 0.51 and a root mean-squared error of 130. The most predictive model (Figure 3) for the overall dataset included stove type, measures of CO (personal, kitchen, and living room), kitchen volume and ventilation, and a microenvironmental PM estimate. The R^2 of this model was 0.76, and the RMSE was 86 (RMSE/mean of measured exposures [nRMSE] =0.66). Of note, and unsurprisingly, all predictive models poorly predicted overall dispersion in biomass households, a phenomenon consistent with the literature.

Although the microenvironmental estimate using the beacon system (model 4) significantly improved model performance over just kitchen $PM_{2.5}$ data from model 2, it did not significantly improve the model fit over the survey data (model 1), likely due to the strong predictive ability of the survey data in this dataset. Similarly, adding kitchen $PM_{2.5}$ to the survey data (as in model 5) did not significantly improve model fit over the survey-only model.

3.2.2 | Kitchen exposure factors (KEFs)

For the current set of measurements, kitchen $PM_{2.5}$ concentrations adjusted by KEF values from the literature were a poor predictor of exposure (Table 6). The exposures predicted by the global burden of disease KEF were significantly different from the measured exposures (Wilcoxon rank sum test, $p < 0.001$) and exhibited wide variability within and between primary stove types. The KEFs derived from Shupler et al.[5] performed better, with lower RMSEs and no significant differences in the distributions between measured and modeled estimates. Mean measured KEFs in this study were 0.32, 0.80, and 1.02 for the LPG, traditional biomass, and charcoal groups, respectively. Additional summary statistics for measured KEFs by stove type are in Table S5, and the distributions by stove type are shown in Figure 4. The relatively high number of KEFs exceeding 1 in LPG households indicates that exposure may have been driven by sources outside of the kitchen, consistent with what one would expect for households that rely on LPG as a cooking fuel.

A random sampling procedure was used to determine the average RMSE between KEFs and personal exposure, with a varying number of households (Figure 5). After 7–10 days, we note no value of additional measurements for this dataset. We also note that we are bounded overall by the poor predictive power of KEF in this dataset.

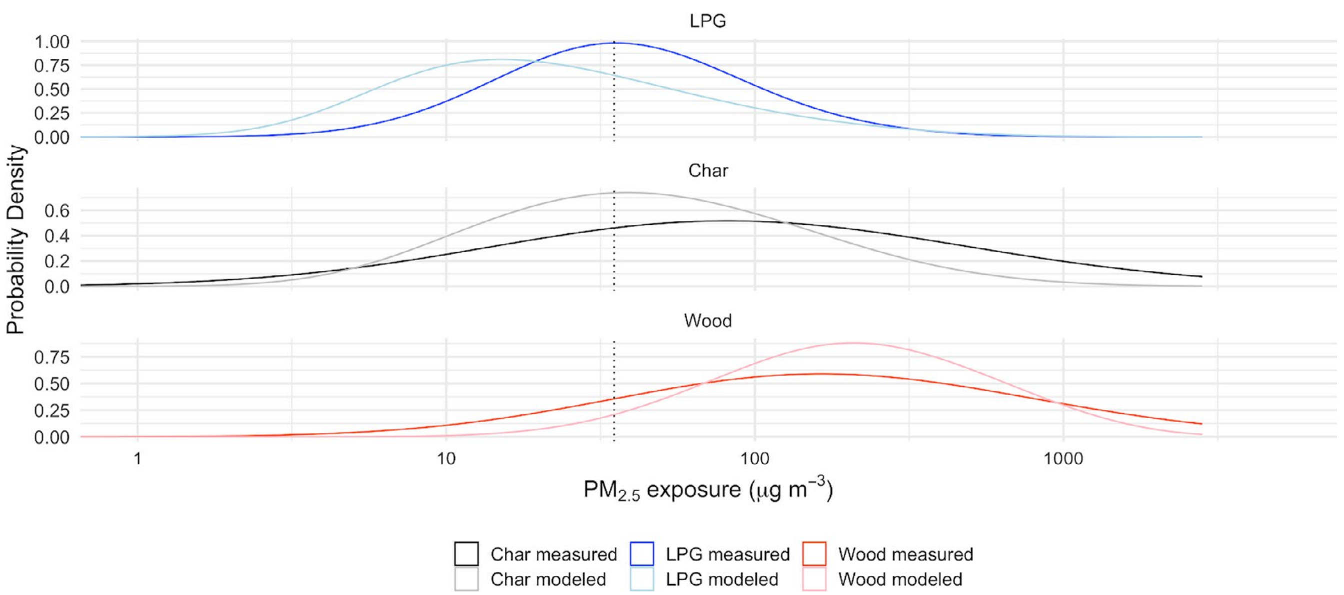

FIGURE 2 Modeled and measured probability distributions (fitted) of $PM_{2.5}$ exposures. The dashed vertical line at 35 μg/m^3 represents the WHO interim target-1 annual guideline

TABLE 3 Comparison of modeled and measured 24-h $PM_{2.5}$ exposures (μg/m^3)

	Modeled	Measured
LPG		
Mean	41	43
Median	19	29
25th–75th percentile	10–41	27–46
n	10 000	19
Charcoal		
Mean	80	115
Median	43	110
25th–75th percentile	21–92	50–121
n	10 000	7
Wood		
Mean	296	225
Median	207	182
25th–75th percentile	113–386	104–292
n	10 000	21

3.2.3 | Machine learning

We utilized a limited set of potential learners available in the SuperLearner R package to estimate personal exposure with the same set of predictors as in the linear regression modeling procedure: random forest (SL.randomForest), extreme gradient boosting (SL.xgboost), support vector machines (SL.ksvm), recursive partitioning and regression trees (SL.rpartPrune), and generalized linear modeling with regularization (SL.glmnet). The independent variable (personal $PM_{2.5}$ exposure) in this case was not log-transformed, an important difference from the linear regressions. 20-fold external cross-validation was performed to select a discrete learner from the algorithms selected. The best performing—based on a mean square error loss function—was extreme gradient boosting (selected in 85% of cross-validation runs), followed by generalized linear modeling (10%), and random forest (5%). The cross-validated ensemble model weights followed this trend (not shown). Comparisons of individual models and the ensemble model are in Figure 6 and Table 7.

Cross-validated machine learning models and the ensemble model performed reasonably well in predicting exposures and warrant further investigation—including creation of additional independent variables using real-time data streams, combinations of existing variables, and additional questionnaire data.

4 | DISCUSSION

4.1 | Model application considerations

Collecting personal $PM_{2.5}$ exposure data is challenging due to cost, technical requirements, and logistical considerations, making predictive models important tools for assessing the health impacts of HAP exposure. To date, HAP exposure models have not been widely applied and performance has varied. Several past works have found that survey data have provided statistically significant predictive power for exposure estimation,[7,34,35] and others have reported comparisons of kitchen and personal $PM_{2.5}$ with varying success.[9,10,45] Additional studies have compared CO exposures (an easier measurement) to $PM_{2.5}$ exposures but found CO to be an unreliable surrogate in many settings.[46,47] A small number of studies have also evaluated the impact of collecting multiple days of measurements in homes or among participants to establish more stable estimates of $PM_{2.5}$ concentrations.[48–50]

Dioniso et al.[51] assessed modeling performance of exposure to CO in children using survey-type data, and though not directly comparable to our work, found a model RMSE of 0.86 ppm (nRMSE = 0.74). Dionisio et al.[52] also assessed the performance of a model in predicting child $PM_{2.5}$ exposure from personal CO, survey data, and kitchen $PM_{2.5}$ concentrations, but did not find strong relationships in any of the model permutations ($R^2 < 0.01$). Hill et al.[8] applied regression models and machine learning models to estimate $PM_{2.5}$ exposure ($n = 36$) in a rural area of Laos, but reported adjusted R^2 values below 0.3, and RMSE of 40.0 μg/m^3 (nRMSE = 0.39). Sanchez et al.[7] used stepwise regression models with survey-based inputs to predict $PM_{2.5}$ exposures in peri-urban South India with R^2 values ranging from 0.09 to 0.25.

Here, our models had R^2 values ranging from 0.23 to 0.76, indicating that overall performance was generally favorable compared that which has been previously reported. Our application of machine learning models was able to explain almost twice as much variability in exposure estimates (R^2 0.23–0.57) than that by Hill et al.[8] and the linear regression models explained up to three times (R^2 = 0.32–0.76) that of Sanchez et al,[7] though there continues to be room for improvement in predictive power. The better performance may be the result of varying environmental factors between different study locations (eg, higher exposure contrasts), or recent improvements in measurement techniques. The survey-based model was moderately predictive of $PM_{2.5}$ exposures (R^2 = 0.51), which has significant implications being that surveys are a much simpler data collection method (compared with direct measurement of exposures or HAP). The best performing model used a combination of survey-based data and measurement, resulting in an R^2 of 0.76 and a root mean-squared error of 85 μg/m^3. Should the predictive capacities be robust, these modeling approaches provide substantial value in mitigating the need or extent of costly and complicated exposure studies.

Our use of SuperLearner was exploratory to show the potential utility of this type of predictive modeling. We likely did not take full advantage of algorithms implemented in SuperLearner, which may have benefitted from the creation of additional predictor variables, including features derived from the minute-to-minute real-time data (from the PM and CO monitors and Beacon system, eg). Future work with SuperLearner should define a robust set of potential predictors to maximize predictions. Such predictors could be derived strictly from survey data and combinations of survey-based variables and would provide an approach to apply machine learning without

TABLE 4 Summary statistics for the 24-h datasets included in the regression models (mean, standard deviation, minimum, 25th percentile, median, 75th percentile, maximum, and number of valid samples)

Variable	Mean	SD	Min	25th %-tile	Median	75th %-tile	Max	*n*
Cook's personal $PM_{2.5}$ exposure (μg/m^3)	139	151	14	43	86	156	687	50
Compliance (fraction of 24-h period monitors in motion)	0.42	0.19	0.03	0.28	0.42	0.62	0.79	50
Kitchen $PM_{2.5}$ (μg/m^3)	492	673	26	50	192	695	3819	50
Secondary Area $PM_{2.5}$ (μg/m^3)	60	118	10	14	22	31	689	47
Ambient $PM_{2.5}$ (μg/m^3)	7	3	3	4	7	9	10	30
Cook's CO exposure (ppm)	4.7	5.7	0.0	1.5	2.7	5.0	32	44
Kitchen CO (ppm)	16.5	22.6	0.0	2.9	8.9	20.3	131	48
Secondary area CO (ppm)	5.0	7.7	0.0	0.0	1.1	8.7	38	49
Ambient CO (ppm)	1.2	2.9	0.0	0.0	0.0	0.2	10	38
Charcoal stove (min)	330	274	45	125	215	575	881	12
Biomass stove (min)	320	188	0	205	370	463	570	13
LPG (min)	126	92	15	73	99	180	410	21
Total cooking time using all stoves (min)	214	219	0	22	115	361	881	50
Beacon PM2.5 indirect exposure estimate (μg/m^3)	255	514	7	37	73	401	3483	50
Beacon CO indirect exposure estimate (ppm)	5.1	5.6	0	0.6	2.9	8.0	24.0	50
Number of walls in the kitchen with open eaves	0.22	0.65	0	0	0	0	3	50
Kitchen volume (m^3)	22.5	12.8	5.4	13.1	20.5	27.2	52.0	50
Open door area (m^2)	2.1	1.3	0	1.7	1.9	2.6	6.0	50
Socioeconomic status index	1.56	2.9	−2.35	−1.05	0.93	4.19	7	50
Air exchange rate (1/h)	17.4	8.2	0.1	11.4	17.1	20.8	40.0	48

TABLE 5 Linear model performance and fit statistics for the personal PM2.5 exposure estimation models

				Overall (*n* = 50)		
	Model	Adj R^2	RMSE (μg/m^3)	Mean (μg/m^3)	SD (μg/m^3)	p[b]
1	Survey-type data[a]	0.51	130	110	65	0.2
2	Kitchen $PM_{2.5}$	0.32	206	115	180	0.47
3	Kitchen $PM_{2.5}$ + CO	0.44	139	118	139	0.45
4	1 + Microenvironmental $PM_{2.5}$ Estimate	0.53	135	114	89	0.30
5	1 + 2	0.52	131	112	80	0.27
6	1 + 3	0.59	112	118	106	0.42
7	Stove Type +Personal, Kitchen, Living room CO +Kitchen volume +Microenvironmental PM Estimate[c]	0.76	86	124	105	0.55
	Measured			139	151	

[a]Includes primary stove type, a socioeconomic index, and kitchen volume.

[b]Comparing predicted and measured values.

[c]Identical to the model selected by the variable selection algorithm.

necessarily collecting hard-to-gather data, like personal exposure to pollutants or microenvironmental location. We additionally note that the generalized linear model used in SuperLearner performed less well than the linear model we used to estimate exposures. We identify two potential realizations for this difference: (1) the algorithms use different solvers to estimate the models (penalized maximum likelihood versus least squares) and (2) in the linear models, the outcome variable was log-transformed, whereas it was not transformed for SuperLearner.

The single-zone model used by WHO and ISO for setting emission targets did reasonably well predicting kitchen concentrations and exposures, after adjusting for its systematic overestimation. The combination of the single-zone model's correlation with measured kitchen concentrations and systematic overestimation suggests it is a reasonable tool for setting performance targets as it provides a conservative approach for linking emission performance with indoor air quality. Given that the model overestimates kitchen concentrations, by potentially up to an order of magnitude, there may be room to adjust the modeling approach such that it provides more reasonable estimates of kitchen concentrations while still erring on the side of conservativeness. Better characterizing the specific factors which contribute to the bias could help in calibrating the model for specific contexts, which would justify potential adjustments for its use in setting stove performance targets.

Previous work has shown that a kitchen exposure factor (KEF) alone has poor predictive power in some contexts.[5] We found substantial variation in our measured KEFs by stove type, indicating that the common method of applying a single ratio globally likely misestimates exposure depending on the type of stove in use in households. Our biomass KEF (0.30) is close to the modeled estimate from Shupler et al.[5] for users of wood-burning stoves (0.327). Our LPG KEF, contrastingly, is much higher (1.03) than the estimate reported by Shupler et al.[5] (0.13). Given that exposures within the home are lower when clean-burning fuels are used and considering the relative contributions to exposure from outside the home (both from conducting activities outside the home and outdoor air penetrating indoors), our measured KEFs in LPG households seem reasonable.

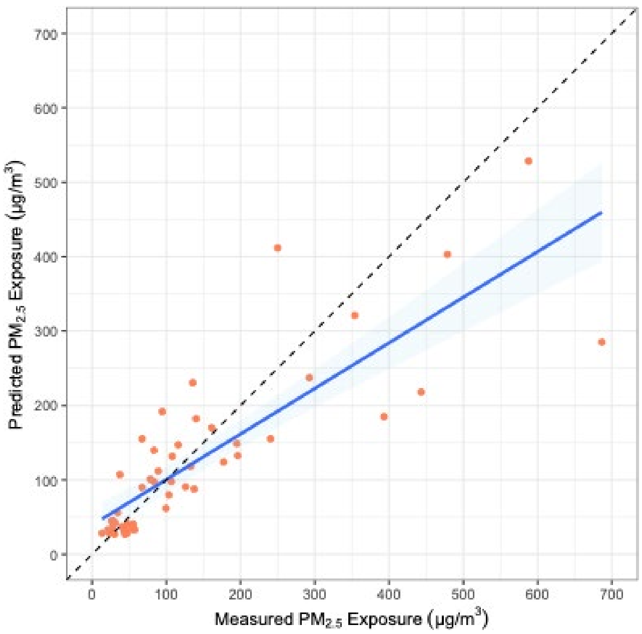

FIGURE 3 The relationship between predicted and measured personal exposure to $PM_{2.5}$. The dotted line is a 1:1 line; dots represent individual data points; the blue line is a linear model including terms for personal, living room, and kitchen CO levels; kitchen volume and ventilation; and an estimate of PM concentrations derived from a microenvironmental location monitoring system

For campaigns measuring mainly kitchen $PM_{2.5}$ levels and monitoring only a subset of personal exposures, our findings indicate that increasing the number of personal measurements can help provide better, less dispersed, and more predictive KEF estimates. Although no clear cutpoint for any stove type was evident, we noticed consistent decreases in RMSE and variability as the number of samples included in KEF-based exposure estimates increased. Future measurement and modeling work should continue to disentangle KEFs in other contexts, as factors such as kitchen layout, background contributions, and behavioral patterns are likely to vary.

4.2 | Limitations

There is potential to reduce the burden of data collection on participants for large-scale projects, as the performance of the exposure estimation regression models that used survey data or less intrusive concentration measurements moderately explained exposure variability. With the ability to collect survey, stove usage, and household air pollution data over multiple days, there are scenarios where this approach should be considered, including instances where long-term trends are of interest. While the models from this work generally performed higher than past efforts, it is not clear that they are repeatable in different contexts, including other seasons as the measurements were conducted solely during the dry season.

We opted to use our resources for conducting a relatively comprehensive set of measurements on potential predictors, which resulted in somewhat small sample sizes. This trade-off seemed appropriate given our goal of exploring new modeling approaches for their potential utility; however, the smaller sample sizes limited our power to evaluate with more certainty which predictors and approaches were strongest.

TABLE 6 Literature KEF exposure prediction performance and fit statistics in comparison with personal $PM_{2.5}$ exposure measurements

	Model	R^2	RMSE (µg/m³)	Mean (µg/m³)	SD (µg/m³)	p
1	Global Burden of Disease KEF (0.742)[6]	0.21	500	365	500	0.003
2	Shupler et al.[5] (stove-specific)[a]	0.22	200	160	220	0.56
3	Shupler et al.[5] (0.327)	0.22	200	153	224	0.72

[a]0.327 for charcoal and biomass stoves; and 0.133 for LPG stoves.

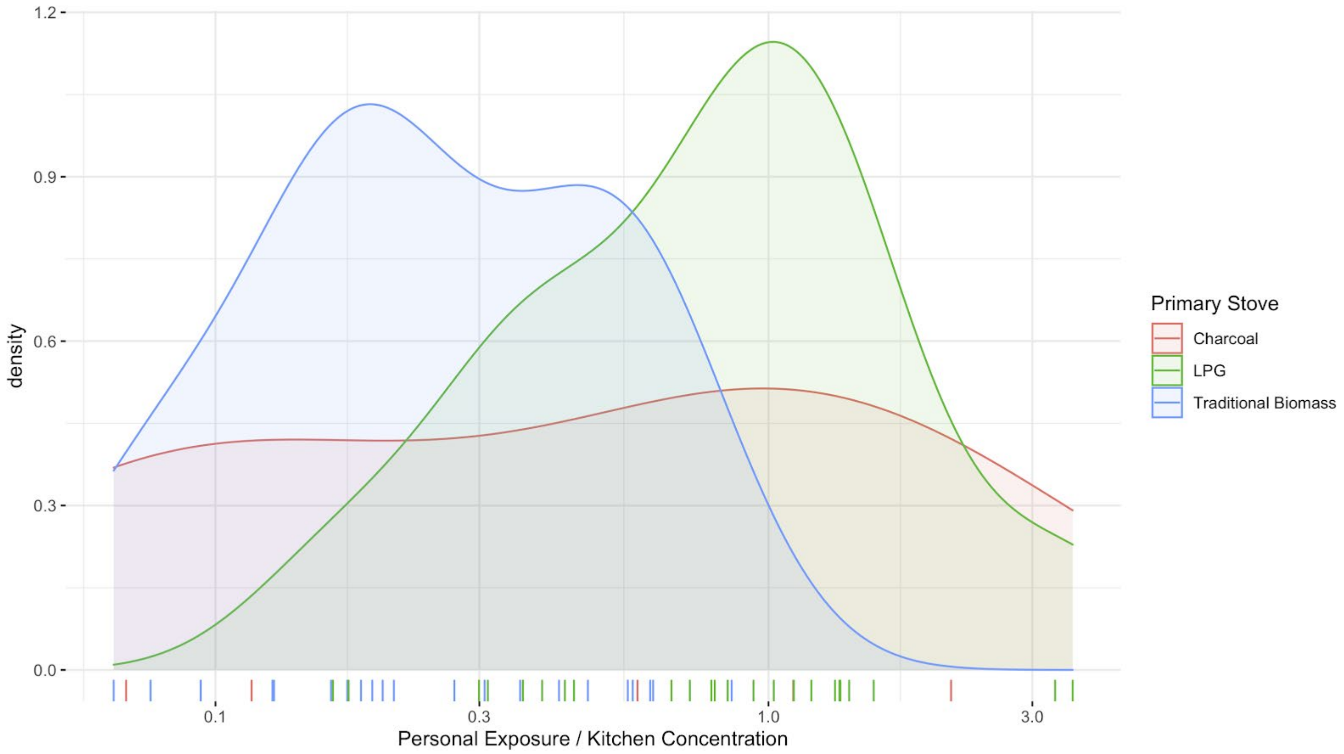

FIGURE 4 Distribution of the kitchen exposure factors (KEF)—the ratio of personal exposure to kitchen concentration—in the current study. A KEF <1 indicates that personal exposure is less than kitchen concentration, while KEF ≥1 indicates that personal exposures exceed kitchen concentrations. In the current study, mean LPG user KEFs were greater than 1, indicating their exposure may be derived outside of the kitchen

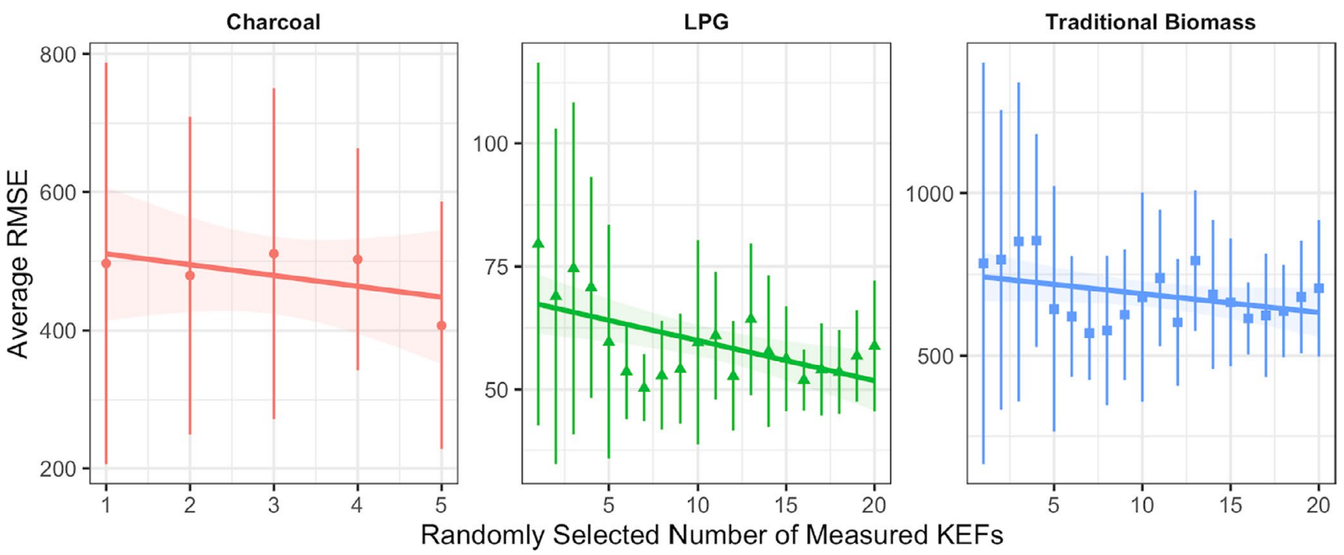

FIGURE 5 Changes in the error of prediction using KEFs based on measurements from varying numbers of households. The x-axis is the number of samples randomly selected from the total stove-specific set of measured KEFs. Random samples of each size were drawn ten times. The y-axis is the mean RMSE of the average of these samples of various sizes. Error bars are the average plus and minus the SD across the ten sets of random sampling. The line is a linear model (RMSE ~N of Samples); the shaded area is the standard error. Note that there were only five households with a primary charcoal stove

Although the sample sizes were relatively small, data management for this study still presented a substantial challenge given the number and types of measurements involved and could likewise be problematic if several data streams are required for various model inputs. Finding the right balance in terms of expected predictive ability and data collection cost and analysis complexity is difficult, and likely will differ with the technical capacity of the group conducting the work. Of course, as improved methods and equipment

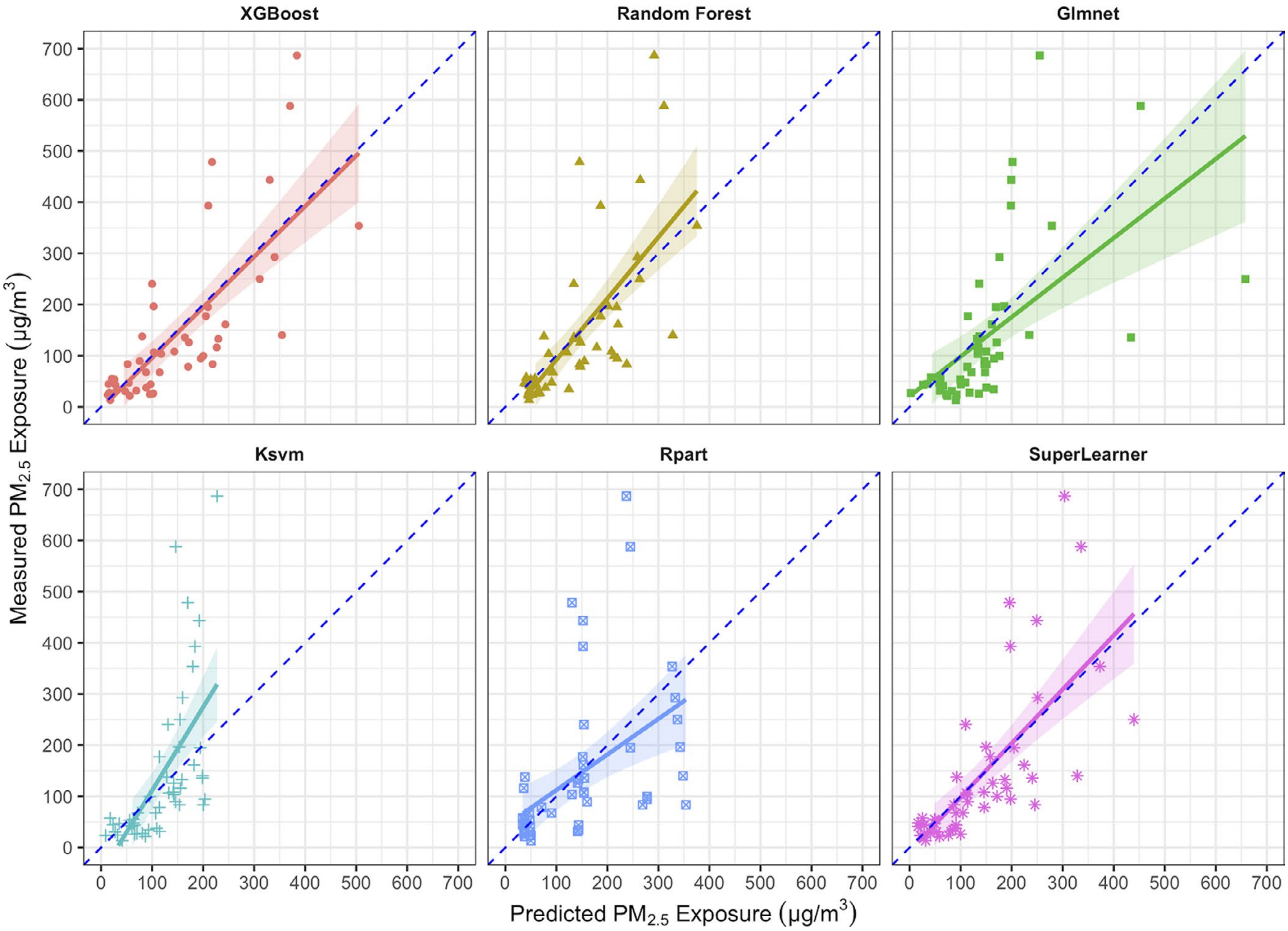

FIGURE 6 Measured vs predicted output from cross-validated machine learning models. The "SuperLearner" panel is the ensemble, weighted model. The best performing models (as assessed by minimization of mean-squared error) were extreme gradient boosting (XGBoost), followed by generalized linear regression (Glmnet) and random forest. KSvm is support vector machines; Rpart is recursive partitioning and regression trees. The dotted line is the 1:1; the solid lines are linear models comparing measured and predicted exposures. Shaded areas are standard errors

TABLE 7 Machine learning model performance and fit statistics

				Overall (n = 50)		
	Model	Adj R^2	RMSE (μg/m^3)	Mean (μg/m^3)	SD (μg/m^3)	p
1	XGBoost	0.57	98	139	121	0.98
2	Random Forest	0.50	107	139	89	0.99
3	Glmnet	0.32	126	153	111	0.60
4	Ksvm	0.40	123	117	59	0.34
5	Rpart	0.23	134	139	104	0.99
6	SuperLearner (ensemble)	0.50	105	138	101	0.99
	Measured			139	151	

continue to reach a wider audience, future analyses can become more streamlined.

It should be noted that any use of these types of models must be considered exploratory unless model validation is performed. There are many idiosyncrasies related to specific contexts that can affect predictive models, such as the intensity of neighbors cooking with dirty fuels, community-level use of polluting fuels, ambient air pollution, temperature, and housing characteristics. In this study, the low housing density and low ambient air pollution levels (Table S3) provided relatively low variability in environmental conditions, allowing the majority of air pollution exposure to be assumed as a result of cooking. Variability of other parameters, grouped by primary stove type, is presented in Table S7. Thus, fuel type indicators may not perform as strongly in other contexts.

4.3 | Recommendations

We recommend testing these models in different geographies or fuel use scenarios as they were developed from a single study community. Continuing to build and test HAP exposure models in different contexts (cooking fuels, geographies, stove use patterns, housing characteristics) would enable more robust evaluation of how they can be extended to other contexts. For clean cooking standards, additional characterization of the single-zone model's bias (more regions, fuel types, housing types, etc.) would help support potential modifications to the model's application for deriving performance targets. Future modeling efforts would also benefit from machine learning approaches, including both supervised and unsupervised methods. These approaches, including those presented here, have shown some promise in generating reasonable predictive power, especially when combined with traditional statistical modeling approaches.[53] De-aggregated real-time data may enhance machine learning predictive power, by identifying data features that may be predictive of mean $PM_{2.5}$ exposures. We also recommend exploration of additional predictors that may be relevant for exposure prediction and considering inclusion of the single-zone model-based exposure estimates in machine learning models as predictors.

ACKNOWLEDGEMENTS

We would foremost like to thank the kind and gracious study participants from the Eldoret area for taking part in the study, without whom it would not have been possible. This work was made possible by the NIHR CLEAN-Air(Africa) Global Health Research Group, led in Kenya by Dr. Diana Menya at Moi University, including Dr. Judy Mangeni, Edna Sang, Gilbert Nyauke, Mary Lydia Kiano, Bernard Bosire, Menya Anabwani, Noelle Sutton, Seth Owiti, Sharon Cherono, Rachel Samoei, Ruth Jepchirchir, Joan Chepng'eno, Joseck Erambo, Zipporah Mageto, and Mondesta Malemo. Thank you to the entire University of Liverpool team of CLEAN-Air(Africa), who supported the fieldwork training, project management, and data preparation, especially Iva Cukic, Rachel Anderson de Cuevas, Sara Ronzi, and Emily Nix. Thank you to Dr. Daniel Wilson at Geocene for assistance in processing the stove usage monitor and exposure data. Thank you to Parker Alexandra Matthews for her assistance in managing data. This work was funded by the United Nations Office for Project Services (UNOPS; RFP/2017/2592), a subsidiary organ of the United Nations, through the Climate and Clean Air Coalition, and managed by the Clean Cooking Alliance. Funding for a substantial portion of the measurements and fieldwork that were complementary to this work was provided by the CLEAN-Air(Africa) Global Health Research Group. CLEAN-AIR(Africa) was funded by the National Institute for Health Research (NIHR) (ref: 17/63/155) using UK aid from the UK government to support global health research. The views expressed in this publication are those of the author(s) and not necessarily those of the NIHR, UK Department of Health and Social Care, UNOPS, the Climate and Clean Air Coalition, or the Clean Cooking Alliance. Ajay Pillarisetti was partially supported by HERCULES Center P30ES019776. The Moi University Institutional Research and Ethics Committee reviewed and granted approval of this study (#0003298).

AUTHORS' CONTRIBUTIONS

MJ: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Methodology, Project administration, Software, Supervision, Visualization, Writing – original draft, Writing – review & editing. RP: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Project administration, Software, Supervision, Visualization, Writing – original draft, Writing – review & editing. AP: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. MS: Data curation, Investigation, Methodology, Project administration, Writing – original draft, Writing – review & editing. DM: Funding acquisition, Investigation, Methodology, Project administration, Supervision, Writing – review & editing. MR: Data curation, Writing – review & editing. SD: Data curation, Writing – review & editing. NP: Funding acquisition, Writing – review & editing. RC: Resources, Writing – review & editing. EP: Conceptualization, Funding acquisition, Methodology, Writing – review & editing. DP: Conceptualization, Funding acquisition, Methodology, Supervision, Writing – review & editing.

DATA AVAILABILITY STATEMENT

Data may be made available through the Climate and Clean Air Coalition and/or the Clean Cooking Alliance.

ORCID

Michael Johnson https://orcid.org/0000-0002-4886-1534
Ajay Pillarisetti https://orcid.org/0000-0003-0518-2934

REFERENCES

1. Bonjour S, Adair-Rohani H, Wolf J, et al. Solid fuel use for household cooking: country and regional estimates for 1980–2010. *Environ Health Perspect*. 2013;121:784-790.
2. Stanaway JD, Afshin A, Gakidou E, et al. Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet*. 2018;392:1923-1994.
3. Pope D, Bruce N, Dherani M, Jagoe K, Rehfuess E. Real-life effectiveness of 'improved' stoves and clean fuels in reducing PM2.5 and CO: systematic review and meta-analysis. *Environ Int*. 2017;101:7-18.
4. Quansah R, Semple S, Ochieng CA, et al. Effectiveness of interventions to reduce household air pollution and/or improve health in homes using solid fuel in low-and-middle income countries: a systematic review and meta-analysis. *Environ Int*. 2017;103:73-90.
5. Shupler M, Godwin W, Frostad J, Gustafson P, Arku RE, Brauer M. Global estimation of exposure to fine particulate matter (PM2.5) from household air pollution. *Environ Int*. 2018;120:354-363.

6. Smith KR, Bruce N, Balakrishnan K, et al. Millions dead: how do we know and what does it mean? methods used in the comparative risk assessment of household air pollution. *Annu Rev Public Health*. 2014;35:185-206.
7. Sanchez M, Milà C, Sreekanth V, et al. Personal exposure to particulate matter in peri-urban India: predictors and association with ambient concentration at residence. *J Eposure Sci Environ Epidemiol*. 2020;30:596-605.
8. Hill LD, Pillarisetti A, Delapena S, et al. Machine-learned modeling of PM2.5 exposures in rural Lao PDR. *Sci Total Environ*. 2019;676:811-822.
9. Piedrahita R, Kanyomse E, Coffey E, et al. Exposures to and origins of carbonaceous PM2.5 in a cookstove intervention in Northern Ghana. *Sci Total Environ*. 2017;576:178-192.
10. Liao J, McCracken JP, Piedrahita R, et al. The use of bluetooth low energy Beacon systems to estimate indirect personal exposure to household air pollution. *J Expo Sci Environ Epidemiol*. 2019;1-11.
11. Balakrishnan K, Ghosh S, Thangavel G, et al. Exposures to fine particulate matter (PM2.5) and birthweight in a rural-urban, mother-child cohort in Tamil Nadu, India. *Environ Res*. 2018;161:524-531.
12. Johnson M, Smith K, Edwards R, Morawska L, Nicas M. WHO guidelines for indoor air quality: household fuel combustion - model for linking household energy use with indoor air quality. 2014. https://www.who.int/airpollution/guidelines/household-fuel-combustion/Review_3.pdf?ua=1
13. ISO. *Technical Report 19867-3: Clean Cookstoves and Clean Cooking Solutions — Harmonized Laboratory Test Protocols — part 3: Voluntary Performance Targets For Cookstoves Based On Laboratory Testing*. Geneva: International Organization for Standardization; 2018. https://www.iso.org/standard/73935.html
14. HEI, IHME. Explore the Data. State of Global Air. 2018. https://www.stateofglobalair.org/data. Accessed May 23, 2018.
15. KNBS. *Survey on Socio Economic Impact Of Covid-19 On Households Report*. Nairobi: Kenya National Bureau of Statistics; 2020.
16. Shupler M, O'Keefe M, Puzzolo E, et al. Pay-as-you-go LPG supports sustainable clean cooking in Kenyan informal urban settlement, including during a period of COVID-19 lockdown. *medRxiv*. 2020. https://doi.org/10.1101/2020.11.20.20235978
17. GLPGP. *National Feasibility Study: LPG for Clean Cooking in Kenya*. New York: Global LPG Partnership; 2019. http://glpgp.org/s/GLPGP-Clean-Cooking-for-Africa-Kenya-National-Assessment-2019.pdf
18. Garland C, Delapena S, Prasad R, L'Orange C, Alexander D, Johnson M. Black carbon cookstove emissions: a field assessment of 19 stove/fuel combinations. *Atmos Environ*. 2017;169:140-149.
19. Johnson MA, Garland CR, Jagoe K, et al. In-home emissions performance of cookstoves in Asia and Africa. *Atmosphere*. 2019;10:290.
20. Ochieng CA, Vardoulakis S, Tonne C. Are rocket mud stoves associated with lower indoor carbon monoxide and personal exposure in rural Kenya? *Indoor Air*. 2013;23:14-24.
21. Yip F, Christensen B, Sircar K, et al. Assessment of traditional and improved stove use on household air pollution and personal exposures in rural western Kenya. *Environ Int*. 2017;99:185-191.
22. Puzzolo E, Menya D, Asante KP, et al. Clean Energy Access for the Prevention of Non-Communicable Disease in Africa. *Environmental Epidemiology*. 2019;3:319. http://dx.doi.org/10.1097/01.ee9.0000609492.94114.c9
23. Chartier R, Phillips M, Mosquin P, et al. A comparative study of human exposures to household air pollution from commonly used cookstoves in Sri Lanka. *Indoor Air*. 2017;27:147-159.
24. Shrivastava A, Gupta V. Methods for the determination of limit of detection and limit of quantitation of the analytical methods. *Chron Young Sci*. 2011;2:21.
25. ISO. *Standard 19867–1: Clean Cookstoves And Clean Cooking Solutions – Harmonized Laboratory Test Protocols – part 1: Standard Test Sequence For Emissions And Performance, Safety And Durability*. Geneva: International Organization for Standardization; 2018.
26. Johnson M, Edwards R, Alatorre Frenk C, Masera O. In-field greenhouse gas emissions from cookstoves in rural Mexican households. *Atmos Environ*. 2008;42:1206-1222.
27. Champion WM, Grieshop AP. Pellet-fed gasifier stoves approach gas-stove like performance during in-home use in Rwanda. *Environ Sci Technol*. 2019;53:6570-6579.
28. Cowlin SC. Tracer decay for determining kitchen ventilation rates in San Lorenzo, Guatemala. *Maxwell Student Projects*, *Max-04-4* (Vol. 1, P. 2). Berkeley: EHS, School of Public Health, University of California; 2005.
29. Wilson DL, Williams KN, Pillarisetti A. An integrated sensor data logging, survey, and analytics platform for field research and its application in HAPIN, a multi-center household energy intervention trial. *Sustainability*. 2020;12:1805.
30. Piedrahita R, Coffey ER, Hagar Y, et al. Attributing air pollutant exposure to emission sources with proximity sensing. *Atmosphere*. 2019;10:395.
31. Vyas S, Kumaranayake L. *Constructing Socio-Economic Status Indices: How To Use Principal Components Analysis*. Oxford: Oxford University Press; 2006.
32. Singh Y, Jackson D, Bhardwaj S, Titus N, Goga A. National surveillance using mobile systems for health monitoring: complexity, functionality and feasibility. *BMC Infect Dis*. 2019;19:786.
33. Balakrishnan K, Ghosh S, Ganguli B, et al. Modeling national average household concentrations of PM2.5 from solid cookfuel use for the global burden of disease -2010 assessment: results from cross-sectional assessments in India. *Environmental Health*. 2013;12:77.
34. Baumgartner J, Schauer JJ, Ezzati M, et al. Patterns and predictors of personal exposure to indoor air pollution from biomass combustion among women and children in rural China. *Indoor Air*. 2011;21:479-488.
35. Clark ML, Reynolds SJ, Burch JB, Conway S, Bachand AM, Peel JL. Indoor air pollution, cookstove quality, and housing characteristics in two Honduran communities. *Environ Res*. 2010;110:12-18.
36. van der Laan MJ, Polley EC, Hubbard AE. Super learner. *Stat Appl Genet Mol Biol*. 2007;6:Article25.
37. L'Orange C, Leith D, Volckens J, DeFoort M. A quantitative model of cookstove variability and field performance: implications for sample size. *Biomass Bioenerg*. 2015;72:233-241.
38. Johnson M, Piedrahita R, Bilsback K, et al. Modeling kitchen air pollution concentrations from emissions and stove usage data in Tamil Nadu, India. 2018.
39. Ruiz-García VM, Edwards RD, Ghasemian M, et al. Fugitive emissions and health implications of plancha-type stoves. *Environ Sci Technol*. 2018;52:10848-10855.
40. Balakrishnan K, Mehta S, Ghosh S, et al. WHO guidelines for indoor air quality: household fuel combustion - population levels of household air pollution and exposures. 2014. http://www.who.int/airpollution/guidelines/household-fuel-combustion/Review_5.pdf?ua=1
41. Weyant CL, Thompson R, Lam NL, et al. In-field emission measurements from biogas and Liquified Petroleum Gas (LPG) stoves. *Atmosphere*. 2019;10:729.
42. Piedrahita R, Johnson M, Bilsback KR, et al. Comparing regional stove-usage patterns and using those patterns to model indoor air quality impacts. *Indoor Air*. 2020;30(3):521-533.
43. 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;45:3237-3243.
44. MacCarty N, Bentson S, Cushman K, et al. Stratification of particulate matter in a kitchen: a comparison of empirical to predicted concentrations and implications for cookstove emissions targets. *Energy Sustain Dev*. 2020;54:14-24.

45. Balakrishnan K, Sambandam S, Ramaswamy P, Mehta S, Smith KR. Exposure assessment for respirable particulates associated with household fuel use in rural districts of Andhra Pradesh, India. *J Expo Anal Environ Epidemiol*. 2004;14:S14-S25.
46. Carter E, Norris C, Dionisio KL, et al. Assessing exposure to household air pollution: a systematic review and pooled analysis of carbon monoxide as a surrogate measure of particulate matter. *Environ Health Persp*. 2017;125:076002.
47. Dionisio KL, Howie SRC, Dominici F, et al. The exposure of infants and children to carbon monoxide from biomass fuels in The Gambia: a measurement and modeling study. *J Expo Sci Environ Epidemiol*. 2012;22:173-181.
48. Cynthia AA, Edwards RD, Johnson M, et al. Reduction in personal exposures to particulate matter and carbon monoxide as a result of the installation of a Patsari improved cook stove in Michoacan Mexico. *Indoor Air*. 2008;18:93-105.
49. McCracken JP, Schwartz J, Bruce N, Mittleman M, Ryan LM, Smith KR. Combining individual- and group-level exposure information: child carbon monoxide in the Guatemala woodstove randomized control trial. *Epidemiology*. 2009;20:127-136.
50. Pillarisetti A.Inspecting what you expect: applying modern tools and techniques to evaluate the effectiveness of household energy interventions. 2016. https://escholarship.org/uc/item/7hw5z2w2. Accessed November 20, 2020.
51. Dionisio KL, Howie S, Fornace KM, Chimah O, Adegbola RA, Ezzati M. Measuring the exposure of infants and children to indoor air pollution from biomass fuels in The Gambia. *Indoor Air*. 2008;18:317-327.
52. Dionisio KL, Howie SRC, Dominici F, et al. Household concentrations and exposure of children to particulate matter from biomass fuels in the Gambia. *Environ Sci Technol*. 2012;46(6):3519–3527.
53. Yuchi W, Gombojav E, Boldbaatar B, et al. Evaluation of random forest regression and multiple linear regression for predicting indoor fine particulate matter concentrations in a highly polluted city. *Environ Pollut*. 2019;245:746-753.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Johnson M, Piedrahita R, Pillarisetti A, et al. Modeling approaches and performance for estimating personal exposure to household air pollution: A case study in Kenya. *Indoor Air*. 2021;31:1441–1457. https://doi.org/10.1111/ina.12790