Supporting Information: 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] Dept of Public Health and Policy, University of Liverpool, Liverpool, UK
[4] Dept 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

Supplementary Information

S1 Site and installation photos

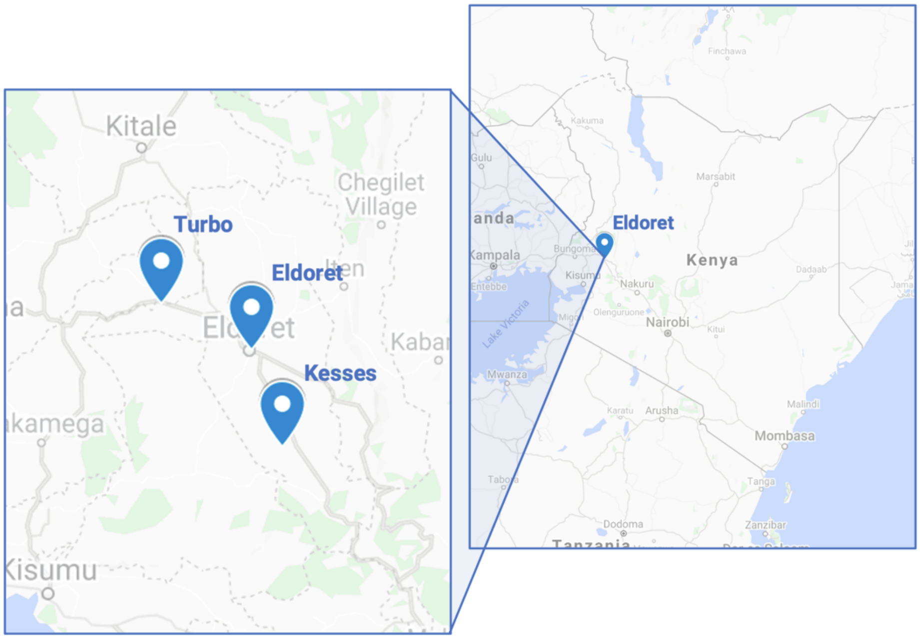

Figure S1. Map showing Turbo and Kesses, the communities to the north and south of Eldoret where the CLEAN-Air (Africa) project and subsequent household selection for this study took place (generated using Google My Maps).

Figure S2. Typical stoves encountered in the study, with the LPG and charcoal stoves shown at left (with SUMs installed), a traditional *Chepkube* stove at center, and a 4-burner LPG stove at right.

Figure S3. The emissions measurement system set-up to measure an LPG cooking event at left, and a biomass cooking event on a Chepkube stove at right. Note the stratified instruments hanging in the room as part of an intensive sample.

Figure S4. Ambient monitors and the installation site.

S2 Emissions performance

While the focus of this work was on developing and evaluating models to predict exposure to household air pollution, stove performance metrics were calculated and are presented below in Table S1. LPG had very high modified combustion efficiency (CO_2/[CO_2+CO] molar) as expected, indicating that almost all fuel carbon was being converted into CO2. Charcoal stoves had the highest CO emissions, common due the surface oxidation combustion process for the fuel. Wood stoves had the highest $PM_{2.5}$ and black carbon emission factors. Wood also had a higher $BC/PM_{2.5}$ ratio, suggesting its aerosol emissions were potentially more warming, but the climate impacts are difficult to characterize based on the limited set of point source emissions, especially as the majority of emissions for charcoal are generated during its production.

Table S1. Stove/fuel performance from measurements during cooking events. Data are presented as means ± standard deviations with the sample size in parentheses.

	LPG	Charcoal	Wood
Modified combustion efficiency (%)	99.1±0.8 (30)	80.8±0.1 (7)	94.0±2.4 (29)
Firepower (kW)	1.60±0.64 (32)	2.53±0.58 (7)	7.15±1.77 (32)
$PM_{2.5}$ emission factor (g/kg)	BDL	3.17±2.18 (7)	6.70±2.96 (29)
BC emission factor (g/kg)	BDL	0.26±0.23 (7)	0.87±0.51 (29)
CO emission factor (g/kg)	17.7±15.8 (30)	373.2±110.0 (7)	67.9±27.7 (29)
$BC/PM_{2.5}$	BDL	0.11±0.11 (7)	0.15±0.13 (29)

BDL = below detection limit

S3 Housing characteristics and socioeconomic status

Figure S5 below shows the distributions of the air exchange rates, room volumes, and cooking event durations for monitored cooking events throughout the sample. These characteristics are key inputs for the WHO and ISO physical models. The mean, standard deviation, and sample size are noted.

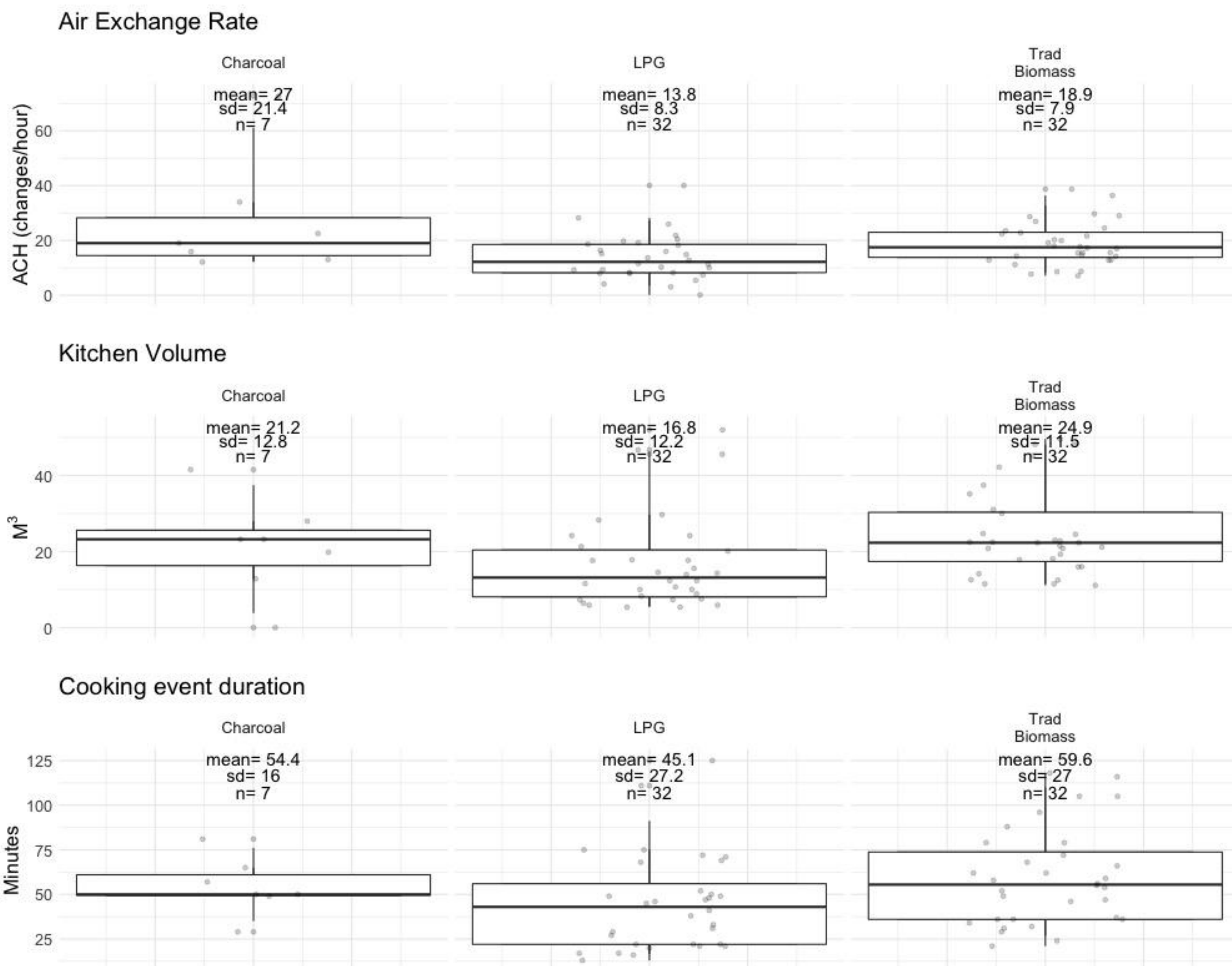

Figure S5. Distributions of kitchen volumes, air exchange rates, and cooking times (key WHO/ISO physical model inputs)

The table below shows the socioeconomic index results for the full sample. The table is split into the average fraction of homes possessing a given characteristic toward the index and the standard deviation of home responses shows the distribution of that characteristic for a given category.

Table S2. Socioeconomic index results. The average fraction of homes columns show the percentage of homes owning an asset or possessing some characteristic, grouped by the index categorization. The standard deviation columns show the distribution of the occurrences of those assets for the given category.

	Average fraction of homes					Standard deviation of home responses				
Ownership or possession	**Poorest quintile (category 1)**	**2**	**3**	**4**	**Wealthiest quintile (category 5)**	**1**	**2**	**3**	**4**	**5**
Own the land/home they live in	0.97	0.82	0.79	0.77	0.67	0.16	0.38	0.41	0.42	0.47
Animal(s)(cows, sheep, etc.)	0.87	0.78	0.75	0.73	0.60	0.34	0.42	0.43	0.44	0.49
Cellphone	0.94	0.88	0.78	0.72	0.59	0.24	0.32	0.42	0.45	0.49
Smartphone	0.04	0.18	0.46	0.77	0.87	0.19	0.38	0.50	0.42	0.33
Radio	0.60	0.69	0.74	0.84	0.84	0.49	0.46	0.44	0.37	0.36
Hi-Fi/CD-player	0.00	0.01	0.04	0.26	0.58	0.05	0.10	0.19	0.44	0.49
Solar connection	0.42	0.30	0.28	0.15	0.15	0.49	0.46	0.45	0.36	0.35
Electricity Connection	0.00	0.24	0.57	0.81	0.90	0.05	0.43	0.50	0.39	0.30
TV	0.05	0.22	0.54	0.84	0.85	0.22	0.42	0.50	0.37	0.36
Satellite TV	0.01	0.19	0.29	0.47	0.65	0.09	0.39	0.46	0.50	0.48
Refrigerator/fridge/freezer	0.00	0.00	0.00	0.02	0.41	0.00	0.00	0.05	0.14	0.49
Shower/bath within house	0.00	0.00	0.02	0.05	0.54	0.00	0.07	0.12	0.23	0.50
Land	0.93	0.79	0.77	0.78	0.72	0.26	0.41	0.42	0.41	0.45
Bicycle	0.08	0.11	0.19	0.23	0.33	0.26	0.31	0.39	0.42	0.47
Moped/Motorcycle	0.06	0.11	0.16	0.17	0.11	0.24	0.31	0.36	0.38	0.32
Pick-up truck	0.00	0.01	0.02	0.04	0.07	0.00	0.09	0.15	0.19	0.25
Car	0.00	0.01	0.03	0.10	0.43	0.00	0.11	0.17	0.30	0.50
Computer	0.00	0.00	0.01	0.03	0.21	0.00	0.07	0.08	0.17	0.41
Washing machine	0.00	0.00	0.01	0.01	0.03	0.00	0.00	0.08	0.09	0.17
Tractor	0.01	0.01	0.03	0.04	0.08	0.08	0.09	0.17	0.19	0.28
Septic or Flushing Toilet Inside	0.00	0.00	0.02	0.09	0.63	0.00	0.00	0.12	0.29	0.48
Latrine in Compound	1.00	0.97	0.99	0.99	0.86	0.00	0.17	0.10	0.11	0.35
Use LPG	0.01	0.28	0.53	0.80	0.95	0.11	0.45	0.50	0.40	0.22
pca_score	-2.17	-1.40	-0.51	0.70	3.38	0.21	0.25	0.31	0.38	1.51

S4 $PM_{2.5}$ and CO household air pollution (HAP) and personal exposure concentrations for the study population

This figure shows a typical 24-hr monitoring period time series, with all plots showing by-minute data. The top frame shows the $PM_{2.5}$ time series for the cook's personal exposure (red), and kitchen concentrations (teal) from the MicroPEM devices; the second frame shows the $PM_{2.5}$ concentration data from the PATS+ devices placed in the kitchen (directly adjacent to the kitchen MicroPEM for inter-comparability and redundancy); the third frame shows the indirect exposure estimates using three different the Beacon localization methods and the associated concentrations from the PATS+ monitors in the given rooms; the fourth frame shows the localization assignment using the three different localization approaches explained previously (color indicates room assignment); the fifth frame indicates stove usage (teal signifies the periods of cooking with the LPG stove, and red signifies the periods of not-cooking).

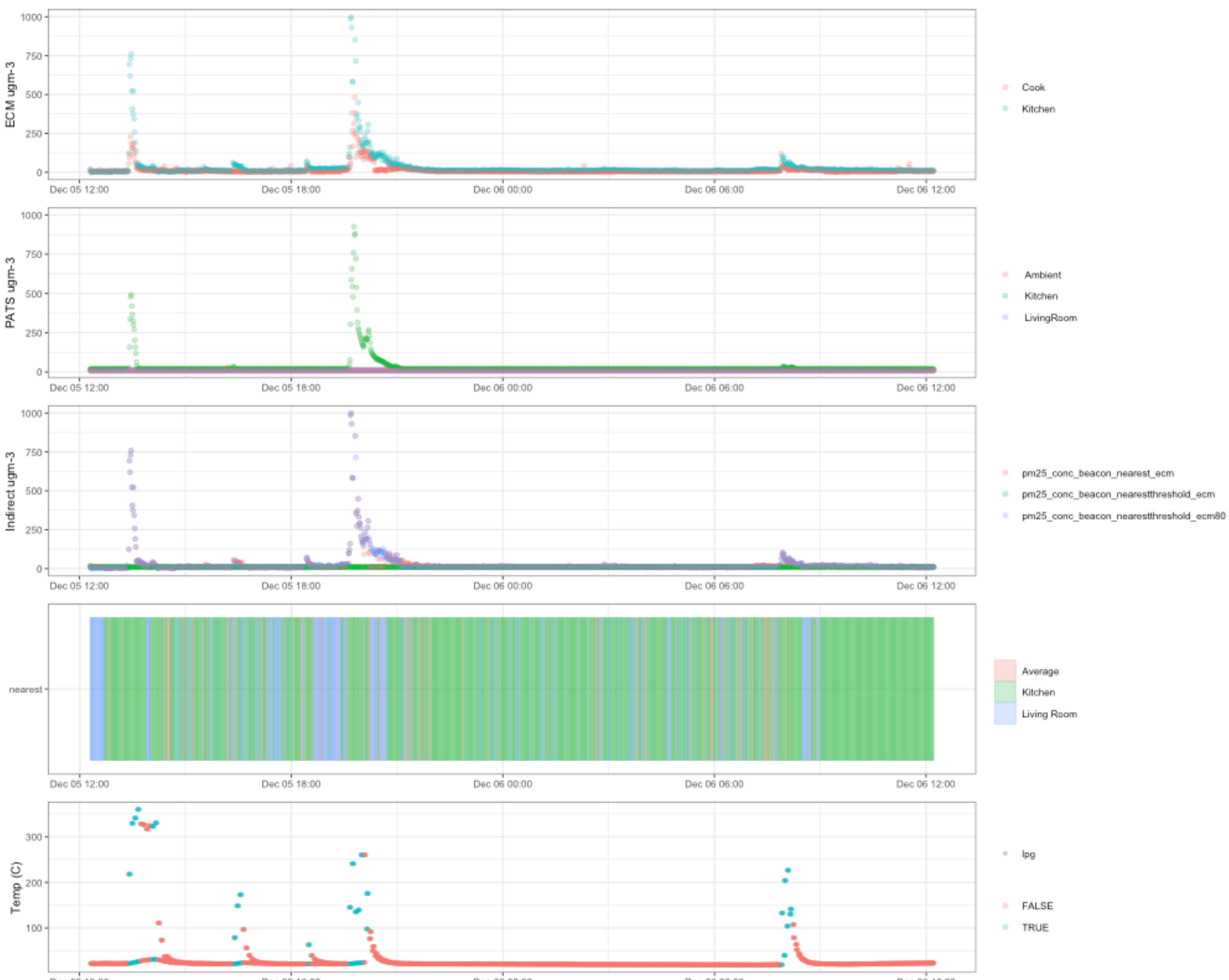

Figure S6. A typical 24-hr monitoring period time series for a single household, with all plots showing by-minute data for all instruments used.

S5 Beacon system walkthrough performance

The walkthrough results indicated that with the 'nearest logger' algorithm, the classification was correct 83.0% of the time when the equipment was in the kitchen, and was incorrect 15.5% of the time, when it classified the location as the other area in which a logger was installed (0.6% of the time, it was classified as equidistant from both loggers, and 0.9% of the time it was classified as not being near either logger, termed 'ambient'). Similarly, for the other area (typically the living room), a correct classification was made 83.2% of the time and an incorrect prediction that the equipment was in the kitchen was made 15.4% of the time, with the remaining 0.7% for both equidistant and ambient classifications.

S6 Stove usage data collection details

Two versions of the FireFinder algorithm were used, the default version for time series in which the temperature exceeded 250 C, and a sensitive version for time series in which the maximum temperature was below 250 C. The FireFinder sensitive algorithm used a primary threshold parameter of 31 C (the 95th percentile of all indoor temperature measurements collected by PATS+ monitors) and a minimum event temperature of 24 C (the 75th percentile of outdoor temperature values). As for the default FireFinder algorithm, the minimum event duration was 5 minutes, and any events within 10 minutes of each other were grouped together into a single event.

Stove usage data was collected at 91 households, for durations ranging from 48-hr to 6 months. Below, an example time series is presented for a home, showing the diurnal temperature trends typical in SUMs measurements, and the peaks produced by cooking events. A shift from LPG to charcoal use is also evident around December 10th.

Figure S7. A typical temperature trace for a single home, showing the diurnal temperature trends typical in SUMs measurements, and the peaks produced by cooking events on two different stoves. TRUE indicates points identified as cooking, while FALSE indicates not-cooking periods.

The figure below shows the thermocouples for the stove use monitoring devices deployed on several stove types. For those stoves that were stationary, the logger of the monitor was affixed above or adjacent to the stove, while the thermocouple was threaded to an appropriate distance from the combustion zone to detect cooking events in temperature traces. For those stoves that were portable, the logger of the monitor was affixed to the body of the stove, while the thermocouple was situated appropriately near the combustion zone, so that the stove and monitor were able to be moved as the participant wished with no disruption of monitoring.

Figure S8. Photos showing SUMs installation on various stove types. On the portable stoves, the logger can be observed affixed to the stove body, while on stationary stoves, the thermocouple is shown threaded to the zone of combustion.

S7 Ambient monitoring

Table S3. Ambient measurement results for $PM_{2.5}$ and CO

	$PM_{2.5}$ (μg/m³)	**CO (ppm)**
Mean	6.83	0.03
SD	4.52	0.64
Min	1.26	0
q25	3.64	0
Median	6.46	0
q75	10	0
Max	293.11	117.5
n (minutes)	55563	54321

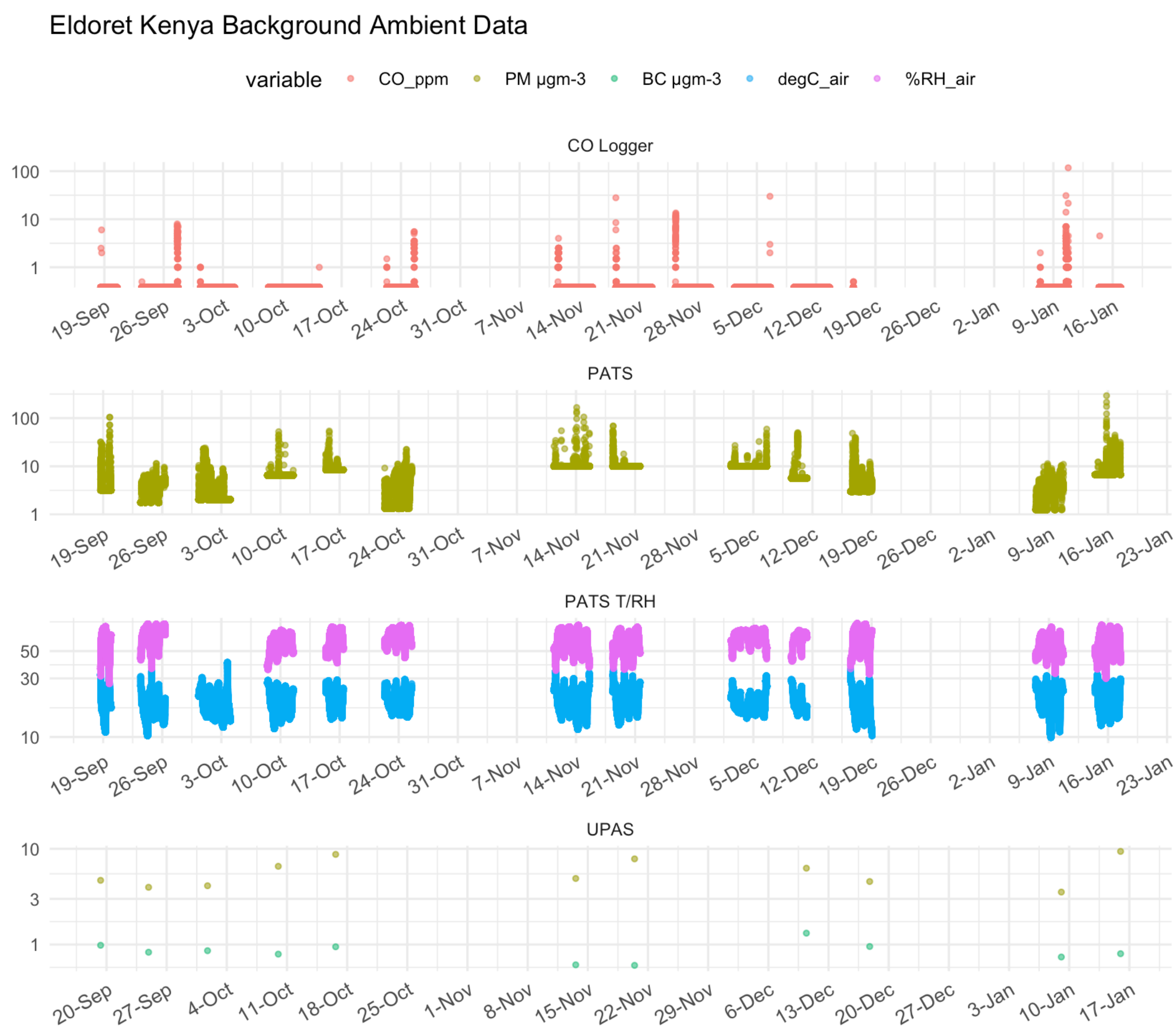

Figure S9. Ambient data, divided by monitoring instrument, over the course of the study period.

S8 Intensive monitoring

Table S4. Intensive sample summary statistics. These data will be further analyzed in future work, to assess the day-to-day variability of the household air pollution measurements, and compliance of the Beacons.

Stove group	Parameter	PM2.5 Kitchen	PM2.5 Kitchen sampling duration	PM2.5 Living Room	PM2.5 Living Room sampling duration	Kitchen CO (ppm)	CO Kitchen sampling duration	Living Room CO (ppm)	CO Living Room sampling duration
Charcoal	**Mean**	86.6	4897	54.2	3907	10.6	4177	8.5	4892
	SD	22.0	1229	42.0	1783	3.1	1471	4.6	1230
	Min	66.2	3306	11.3	2081	8.4	2772	2.6	3300
	q25	68.9	4268	30.3	2685	8.7	3167	7.3	4262
	Median	84.7	5103	47.9	3737	9.4	3930	8.9	5100
	q75	102.3	5731	71.8	4960	11.3	4939	10.1	5730
	Max	110.8	6074	109.9	6074	15.1	6074	13.7	6069
	n	4.0	4	4.0	4	4.0	4	4.0	4
Chepkube	**Mean**	1027.4	7101	49.2	7059	11.9	5056	1.5	6619
	SD	534.7	718	18.6	728	4.0	2037	2.4	92
	Min	410.0	6686	27.8	6639	9.6	3880	0.1	6513
	q25	873.0	6686	43.9	6639	9.6	3880	0.1	6593
	Median	1336.1	6686	60.0	6639	9.6	3880	0.1	6672
	q75	1336.1	7308	60.0	7270	13.0	5645	2.2	6672
	Max	1336.1	7930	60.0	7900	16.5	7409	4.2	6672
	n	3.0	3	3.0	3	3.0	3	3.0	3
LPG	**Mean**	119.2	5098	65.2	4508	8.4	4485	5.6	4719

	SD	158.9	1543	67.1	2403	7.1	1923	5.4	2051
	Min	10.7	2718	13.8	0	0.2	1767	0.3	0
	q25	29.3	4657	20.2	3944	3.6	3060	0.5	3900
	Median	58.2	4727	37.5	4726	5.7	4621	2.9	4670
	q75	111.7	5763	82.8	5762	13.8	4761	7.9	5748
	Max	543.3	8092	221.9	8043	21.6	8020	16.7	7999
	n	14.0	14	12.0	14	14.0	14	13.0	14
	Mean	630.1	4429	25.5	4429	11.1	3587	1.6	4289
	SD	463.9	3097	10.5	3096	6.4	1898	2.7	3192
	Min	131.6	2006	17.3	2006	3.4	1936	0.0	1922
	q25	263.8	2858	19.3	2857	6.4	2424	0.1	2424
Trad Biomass	**Median**	473.3	2951	21.9	2951	9.9	2925	0.3	2941
	q75	975.1	4952	26.6	4952	15.3	4233	1.7	4945
	max	1328.0	10428	47.7	10426	20.9	6934	7.2	10420
	n	7.0	7	7.0	7	7.0	7	7.0	7

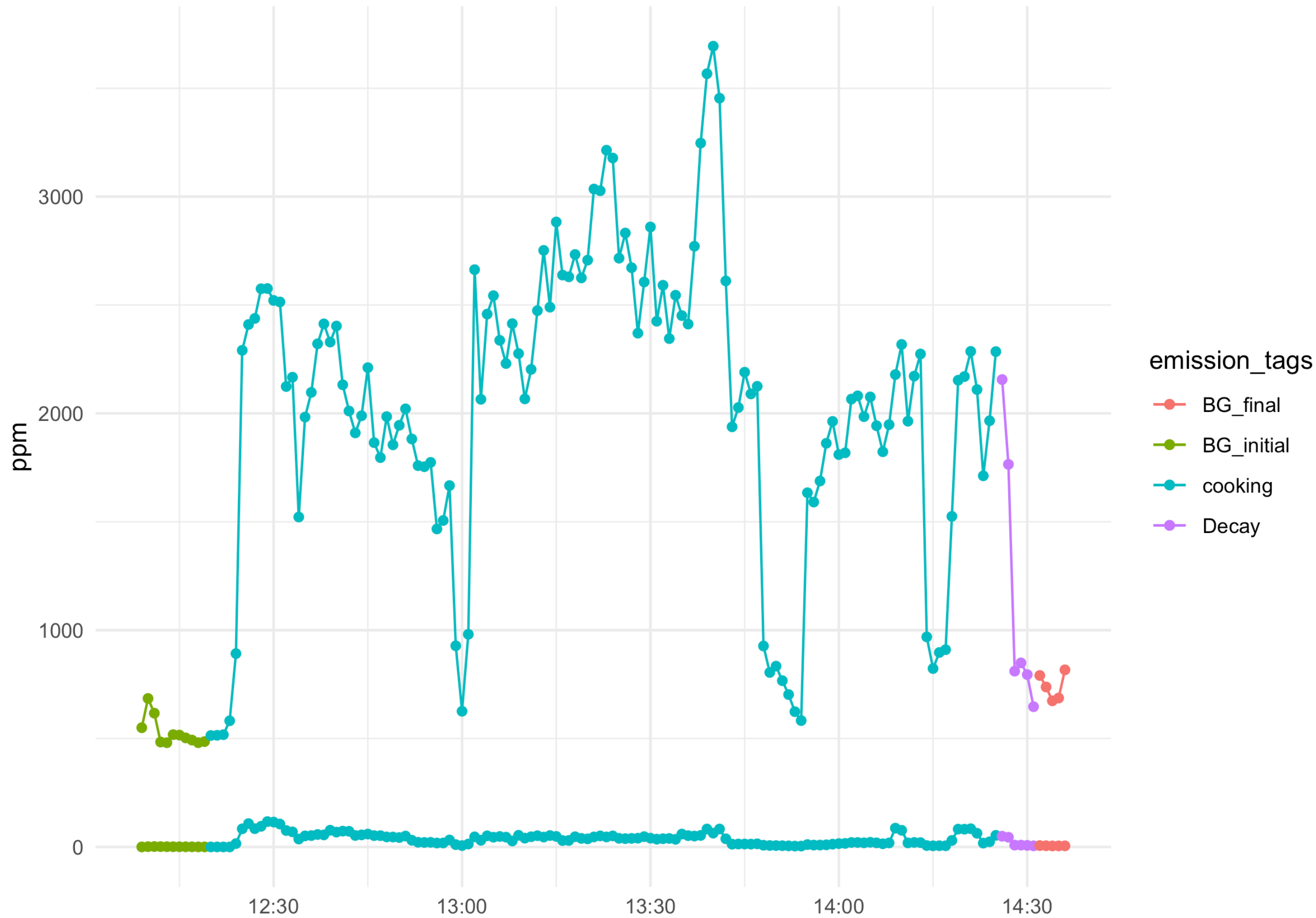

Figure S10. Typical CO and CO2 emissions time series, showing the initial background period, the cooking period, and final background period, in addition to the data points identified to be associated with the decay that can be used to calculate the kitchen ventilation rate.

Table S5. Measured KEF summary statistics by stove type

Kitchen Exposure Factor (KEF)					
Primary Stove	Mean	SD	Min	Median	Max
Traditional Biomass	0.32	0.22	0.07	0.21	0.86
Charcoal	0.80	0.86	0.07	0.58	2.14
LPG	1.02	0.85	0.16	0.82	3.55
Overall	0.7	0.73	0.07	0.46	3.55

Table S6. Pollutant concentrations during cooking events for the stratified samples that had monitors placed at 1.0, 1.5, and 2.0 meters from the floor. Stoves types are grouped by categories of LPG vs. traditional due to low sample sizes. The traditional group includes charcoal, traditional biomass, and Chepkube biomass stoves.

Pollutant (units)	Stove category	Height from floor (meters)	Mean	Median	Standard deviation	n
$PM_{2.5}$ (μg/m^3)	LPG	1.0	61.0	15.2	70.9	9
$PM_{2.5}$ (μg/m^3)	LPG	1.5	126.1	33.7	143.0	9
$PM_{2.5}$ (μg/m^3)	LPG	2.0	109.1	69.7	151.8	9
CO (ppm)	LPG	1.0	2.6	0.0	4.6	9
CO (ppm)	LPG	1.5	4.7	3.8	4.9	9
CO (ppm)	LPG	2.0	9.3	4.5	16.9	9
$PM_{2.5}$ (μg/m^3)	Traditional	1.0	295.1	200.9	285.6	6
$PM_{2.5}$ (μg/m^3)	Traditional	1.5	1009.9	625.5	1073.0	6
$PM_{2.5}$ (μg/m^3)	Traditional	2.0	1668.8	658.8	2514.6	4
CO (ppm)	Traditional	1.0	29.3	27.8	20.3	6
CO (ppm)	Traditional	1.5	42.3	38.7	39.9	6
CO (ppm)	Traditional	2.0	57.2	36.5	58.4	6

Table S7. Intraclass Correlation Coefficients (ICC) for model parameters by primary stove type. The ICC is the proportion of variability explained by between group differences. In the table below, ICCs were estimated using the *R* package *ICC* (described in Wolak et al.[1]), A low ICC indicates high variability within stove groups; a high ICC indicates high variability between stove groups.

Variable	Within-group variance	Between-group variance	ICC
Cook's PM2.5 Exposure (μg/m3)	16000	11000	0.39
Kitchen PM2.5 concentrations (μg/m3)	290000	270000	0.48
Kitchen volume (m3)	160	9.1	0.05
Door and window area (m2)	1.5	0.29	0.16
score	2.8	9.4	0.77
Air changes per hour	60	11	0.16

References:

1. Wolak ME, Fairbairn DJ, Paulsen YR. Guidelines for estimating repeatability. *Methods in Ecology and Evolution*. 2012;3:129–137.