Supplementary Materials:

Ricardo Piedrahita[1], Evan R. Coffey[1,*], Yolanda Hagar[2], Ernest Kanyomse[3], Katelin Verploeg[1], Christine Wiedinmyer[4,6] , Katherine L. Dickinson[5], Abraham Oduro[3] and Michael P. Hannigan[1]

1. G-Pod cooking area micro-environment monitoring

The G-Pod air quality monitor sampling inlets were placed 1-meter away from the cookstove of interest, at 1-meter height, with BLE Beacons bolted to the outside of the cases. CO was measured with Alphasense CO-B4 electrochemical sensors. CO_2 was measured with NDIR sensors (S200, ELT Corp.). Temperature, humidity, and barometric pressure were also measured in the G-Pod, and on a subset of samples, total VOCs were measured with PID sensors (pID-Tech Plus Silver, Baseline-Mocon Inc.). Integrated $PM_{2.5}$ was collected and analyzed as described elsewhere [1]. From December 2013 through November 2014, only the most-used cooking area was monitored, but from November 2014 – January 2016, the two most-used cooking areas were monitored.

2. Cooking area microenvironment measurements calibration and data processing

We employed a multi-step protocol to ensure data quality over the duration of the study. CO and CO_2 sensors underwent lab calibrations at the University of Colorado before and after each sampling period (November 2013, October 2014, May 2015, October 2015, and February 2016). An exponential calibration model controlling for temperature was used for the Alphasense CO sensor [2], while a first order linear model was used for the ELT CO_2 sensor [3]. Span checks were performed at the NHRC in March 2015 after receiving cylinders of span gases. Calibrations were very consistent over time for these sensors, as has been previously shown [3].

The G-Pods were configured to sample at 15-second intervals, and 1-minute medians were used for further analysis. CO and CO_2 data were baseline-adjusted to the 5^{th} percentile of the ambient background concentration, to mitigate baseline sensor drift over time. We found evidence of uniform background levels of these pollutants (data not yet published), and since ventilation rates are very high in the measured microenvironments due to building styles, we considered this to be a reasonable approach since we could not perform full calibrations as often as desired.

3. Beacon distance calibration

The iBeacon protocol includes a calibration constant to normalize the RSSI-to-distance conversion, but Android devices do not use this method, which made it necessary to perform a calibration with our specific hardware (Figure 1). Calibration was performed on an open sports field free of extraneous objects using an equation of the form $distance = 10^{(p1 * RSSI)/p2)}$, as has been used commonly in beacon work [4]. Stationary data was collected with two phones and two beacons at distances of 2, 5, 10, 20, and 40m, for durations of three minutes at each distance.

The data from each phone/beacon combination was fit individually to but in the end a single calibration using all the data was kept for further analysis because we determined that the inter-phone and inter-beacon differences were due to random experimental error, like orientation, rather than systemic differences in hardware. Additionally, not all phones were available at the time of calibration, so a bulk approach was deemed prudent. An R-squared value of 0.72 was obtained using all the data, with evenly distributed residuals.

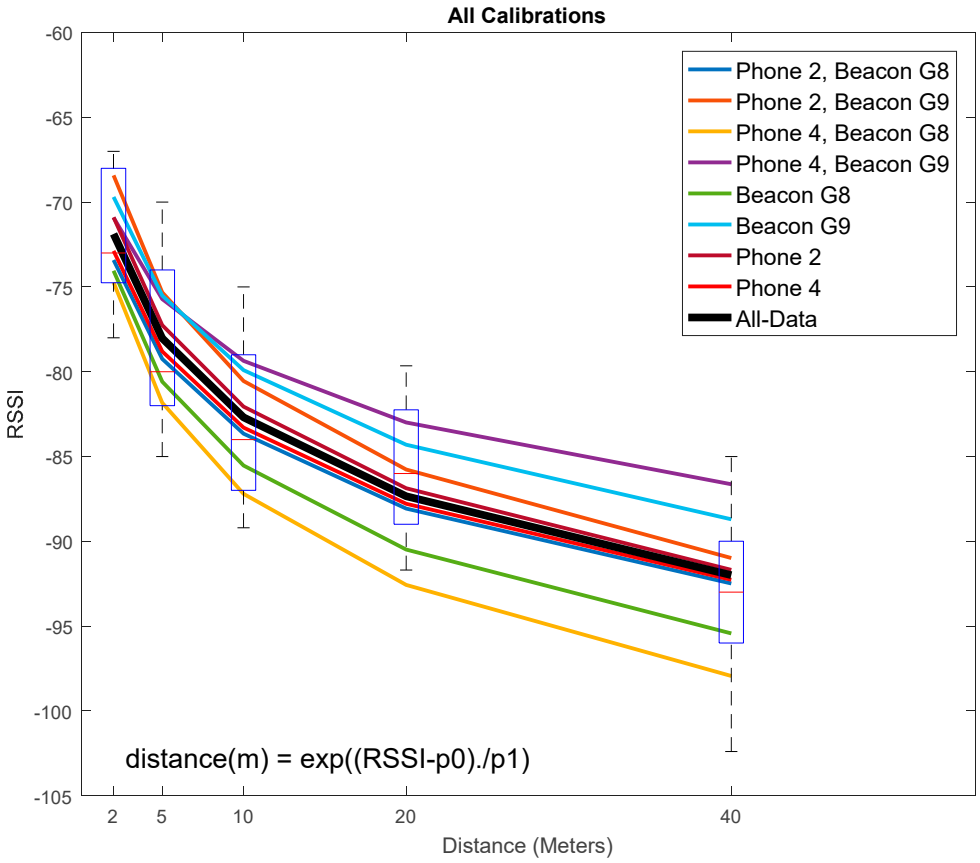

Figure S1. RSSI-to-distance calibrations for various calibration models. The bold black line shows a fit using aggregate data from both phones, and both beacons, while the thin lines are phone/beacon specific. Box and whisker plots show the distributions of the all the raw data, with whiskers representing 5th and 95th percentiles. Note that the outlying curves on the top and bottom of the plot are from phone 4, suggesting a performance issue with that phone.

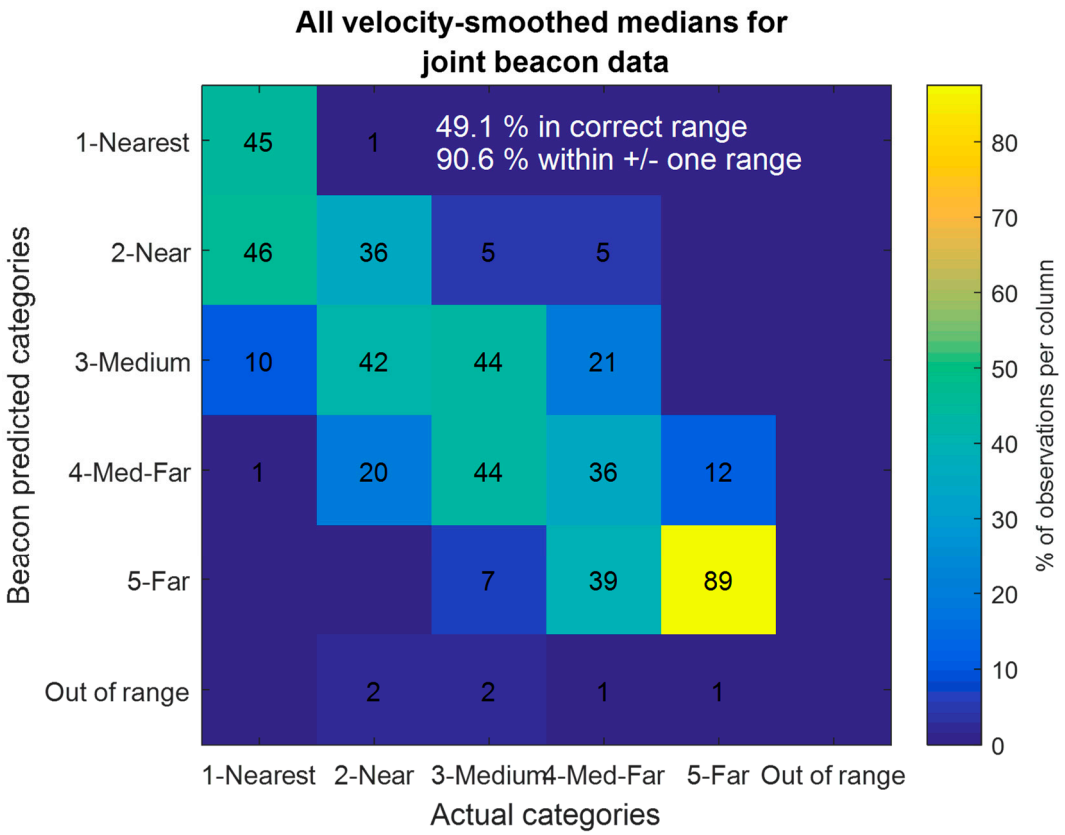

Figure S2. Modeled categories vs. known categories for all merged beacon signal data. Percentages add up to 100 by column, as the x-axis represent the known category values.

4. Participant protocol compliance

Compliance was calculated using the rolling standard deviation of one-hour segments of the minute beacon data and flagging hours in which the standard deviation of any available beacon signal in units of RSSI was greater than 2, excluding nighttime between 21:00 and 7:00. Standard deviation of 2 was selected based on noise variability during calibration periods and is in units of RSSI because the residuals are normally and evenly distributed throughout the distance categories, but residuals are not evenly distributed after passing through the exponential calibration curve in conversion to distance (m). Using this approach, average compliance was measured at 81.9%.

5. Beacon system validation

To understand signal measurement uncertainty, we can first look at the results of a simple test we conducted outdoors with one phone and one Beacon. When the body is directly between a phone and Beacon, the signal attenuation is equivalent to predicting a change in distance from ~1 meter to ~10 meters. When performed indoors, the results are usually less pronounced, due to signal reflectance aiding the Beacon signal to reach the phone. A validation test was performed as part of an outdoors cookstove test. The same distance ranges were prepared, and a user walked throughout each range for 20 minutes. Three phones were placed at the epicenter of the region arcs, along with the stove. There were more obstructions in this test than the open field, making it a more realistic scenario. There are still limitations with this approach, and it is not meant to represent all indoor use, which could be highly variable due to placement of equipment, home layouts, building materials, and user behaviors. Dedicated indoor testing in a variety of environments would provide a better understanding of the expected performance.

Validation testing for both deployments and all combinations of phones and beacons showed correct classification of distance categories on 30.9% of observations on average, and 67.4% of observations were within one distance zone of the correct zone. System performance was not

significantly different between the initial validation test and the outdoor cooking test. However, classification errors were not evenly distributed among distance categories, with lower matching success rates for the more distant ranges (Figure 3). This was expected, since the relative signal drop-off due to bodily interference is higher for closer ranges. Calibration showed inter-phone variability of 4.4m (RMS error) (Figure 1), suggesting that each of our phones would have benefitted from individual calibrations, though such variability is model specific for the phone and may not be the case with other phone models.

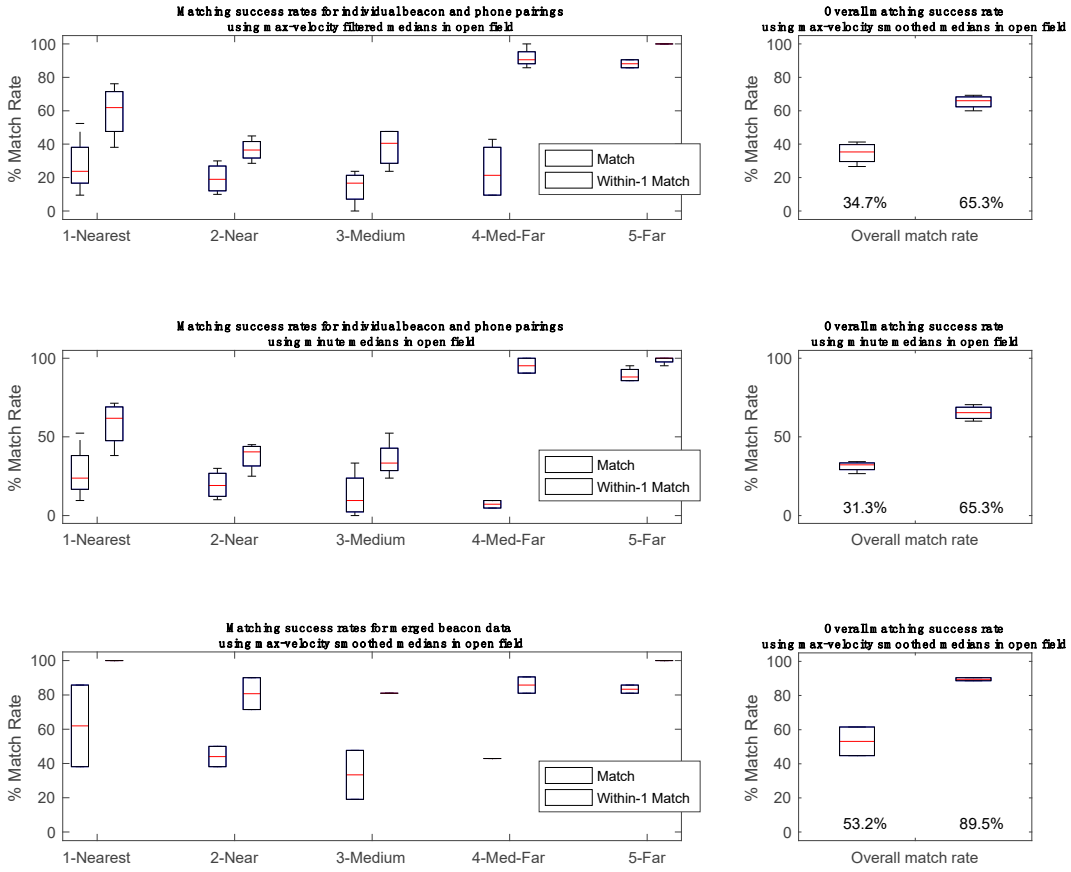

Figure S3. Performance from the validation deployment in an open field. Light colored boxes show the match rate, and dark boxes show the rate at which the algorithm predicted within one zone of the correct zone. Left frames show performance by distance zone, while right frames show overall performance. Top frames show match rates using the MV algorithm, the middle frames show rates using minute medians, and the bottom frames show match rates using the merged beacon data along with the MV algorithm.

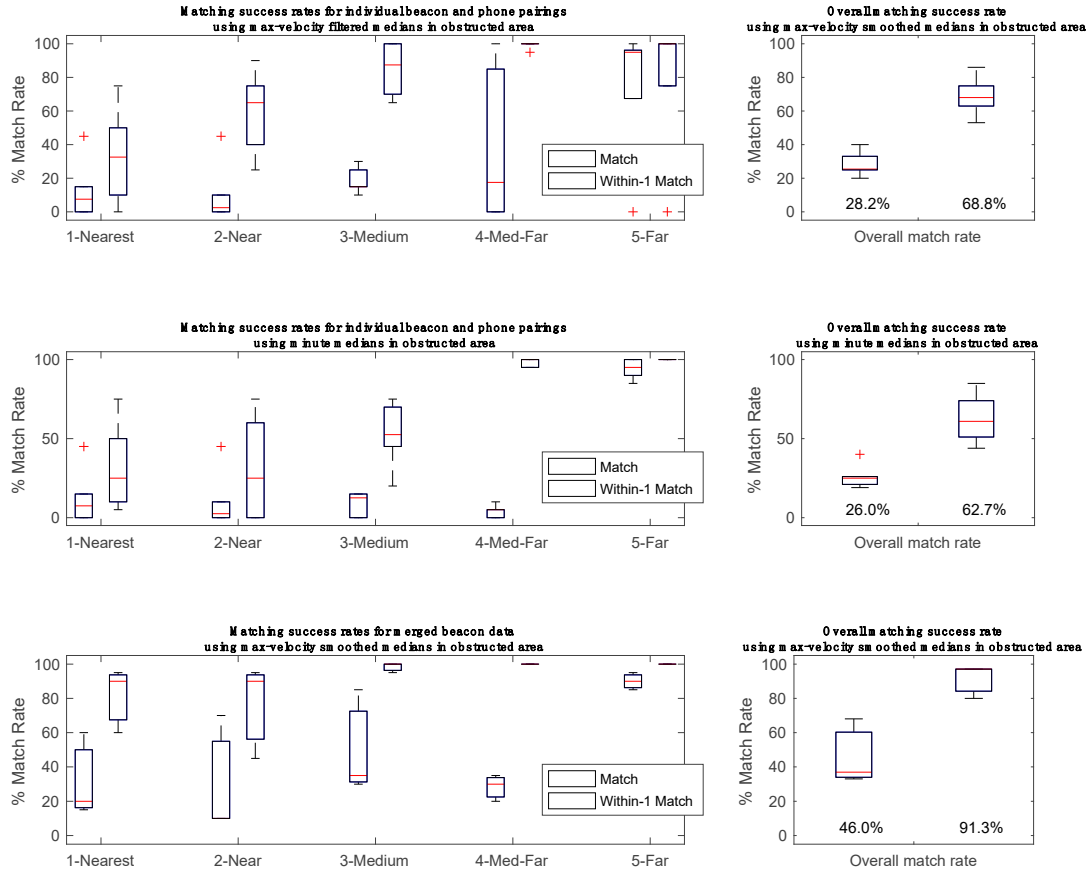

Figure S4. Performance from the test deployment with additional obstructions. Light colored boxes show the match rate, and dark boxes show the rate at which the algorithm predicted within one zone of the correct zone. Left frames show performance by distance zone, while right frames show overall performance. Top frames show match rates using the MV algorithm, the middle frames show rates using minute medians, and the bottom frames show match rates using the merged beacon data along with the MV algorithm.

The MV filter provided a 2.2% and 3.4% improvement over the simple medians for the direct match rate of the 'open field' and 'obstructed' data sets, and 0% and 6.1% improvement for the within-one match rate of the 'open field' and 'obstructed' data sets (Figure 2,3). The physical reasoning behind this approach suggests that it would improve performance in more variable and dynamic environments, with only minor potential drawbacks in outlying use cases.

Merging the data from both beacons worn by the user resulted in substantially better performance, since the attenuation effects were much less pronounced due to the improved direct line-of-sight to the phones at nearly all times. 53.1% of observations were correctly classified on average, while 89.5% of observations were classified within one zone for the 'open field' test, while in the obstructed data set the values were 46.0% and 91.3%. The merged data had errors that were more evenly distributed among categories.

6. CO and beacon modeling results

Table S1. Summary of time-activity states and defining criteria used in Equations 1 and 2.

State	Time-activity category	Proximity criteria	Cooking area CO concentration criteria
1	Away From Home	more than 90m away from both cooking areas	NA
2	Home Not Cooking	within 90m of any cooking area	CO < 10 ppm
3	Home Cooking (control)	within 90m of any cooking area	CO > 10 ppm

4	Home Cooking (Gyapa/Gyapa)	within 90m of any cooking area	CO > 10 ppm
5	Home Cooking (Philips/Philips)	within 90m of any cooking area	CO > 10 ppm
6	Home Cooking (Gyapa/Philips)	within 90m of any cooking area	CO > 10 ppm

Table S2. Summary of results from modeling personal CO exposure by cooking area CO.

	Personal vs. cooking area CO by zones (Equation 3)				**Daily average personal vs. cooking area CO (Equation 4)**			
	Expected value ppm (95% CI)	**Coefficient (95% CI)**	**% change (95% CI)**	**P-value**	**Expected value ppm (95% CI)**	**Coefficient (95% CI)**	**% change (95% CI)**	**P-value**
Intercept	0.1 (0.07, 0.16)	-2.27 (-2.69, -1.85)	NA	0.00	0.14 (0.07, 0.25)	-2.0 (-2.62, -1.38)	NA	0.00
log(Weighted cooking area CO)	2.74 (2.25, 3.33)	1.01 (0.81, 1.2)	173.54 (124.75, 232.92)	0.00	2.24 (1.5, 3.36)	0.81 (0.4, 1.21)	124.25 (49.87, 235.56)	0.00
Random effect by individual		0.35 (0.13, 0.94)				0.00		
Random error covariance		1.0 (0.76, 1.31)				1.28 (0.82, 2.01)		
Adjusted R-squared		0.63				0.28		
N		123				38		

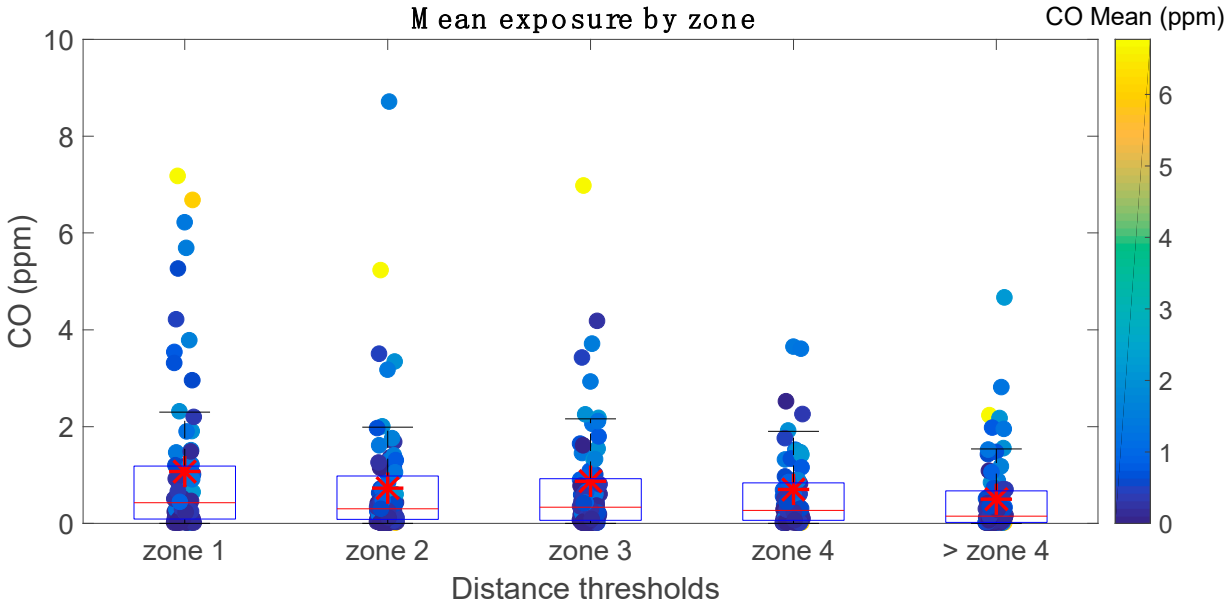

Figure S5. Mean exposure distributions categorized by zones. Marker colors indicate the participant's average exposure from the entire day, and red stars represent means by zone. Slope of

decreasing average exposure by zone was not found to be statistically significant by univariate linear regression.

7. Daily average modeling by stove group using only the data available with Beacons

Model information:

Number of observations	71
Fixed effects coefficients	4
Random effects coefficients	31
Covariance parameters	2

Formula:

LogPersonalCOMeans ~ 1 + StoveGroup + (1 | UserID)

Model fit statistics:

AIC	BIC	LogLikelihood	Deviance
274.86	288.44	-131.43	262.86

Fixed effects coefficients (95% CIs):

Name	Estimate	SE	tStat	DF	pValue	Lower	Upper
'(Intercept)'	-0.38888	0.689	-0.56442	67	0.57435	-1.7641	0.98636
'StoveGroup_C'	-1.1923	0.84385	-1.4129	67	0.16232	-2.8766	0.49207
'StoveGroup_B'	-1.0066	0.73792	-1.3642	67	0.17708	-2.4795	0.46626
'StoveGroup_A'	-1.2481	0.76329	-1.6351	67	0.10671	-2.7716	0.27544

Random effects covariance parameters (95% CIs):

Group: UserID (31 Levels)

Name1	Name2	Type	Estimate	Lower	Upper
'(Intercept)'	'(Intercept)'	'std'	6.4937e-07	NaN	NaN

Group: Error

Name	Estimate	Lower	Upper
'Res Std'	1.5406	1.307	1.8161

Figure S1: RSSI-to-distance calibrations for various calibration models. The bold black line shows a fit using aggregate data from both phones, and both beacons, while the thin lines are phone/beacon specific. Box and whisker plots show the distributions of the all the raw data, with whiskers representing 5th and 95th percentiles. Note that the outlying curves on the top and bottom of the plot are from phone 4, suggesting a performance issue with that phone. Figure S2: Modeled categories vs. known categories for all merged beacon signal data. Percentages add up to 100 by column, as the x-axis represent the known category values. Figure S3: Performance from the validation deployment in an open field. Light colored boxes show the match rate, and dark boxes show the rate at which the algorithm predicted within one zone of the correct zone. Left frames show performance by distance zone, while right frames show overall performance. Top frames show match rates using the MV algorithm, the middle frames show rates using minute medians, and the bottom frames show match rates using the merged beacon data along with the MV algorithm. Figure S4: Performance from the test deployment with additional obstructions. Light colored boxes show the match rate, and dark boxes show the rate at which the

algorithm predicted within one zone of the correct zone. Left frames show performance by distance zone, while right frames show overall performance. Top frames show match rates using the MV algorithm, the middle frames show rates using minute medians, and the bottom frames show match rates using the merged beacon data along with the MV algorithm. Figure S5: Mean exposure distributions categorized by zones. Marker colors indicate the participant's average exposure from the entire day, and red stars represent means by zone. Slope of decreasing average exposure by zone was not found to be statistically significant by univariate linear regression. Table S1: Summary of time-activity states and defining criteria used in Equations 1 and 2. Table S2: Summary of results from modeling personal CO exposure by cooking area CO.

References

1. Piedrahita, R.; Kanyomse, E.; Coffey, E.; Xie, M.; Hagar, Y.; Alirigia, R.; Agyei, F.; Wiedinmyer, C.; Dickinson, K.L.; Oduro, A.; et al. Exposures to and origins of carbonaceous PM2.5 in a cookstove intervention in Northern Ghana. Science of The Total Environment 2017, 576, 178–192.
2. Masson, N., Piedrahita, R., Hannigan, M., 2015. Quantification Method for Electrolytic Sensors in Long-Term Monitoring of Ambient Air Quality. Sensors 15, 27283–27302. doi:10.3390/s151027283
3. Piedrahita, R.; Xiang, Y.; Masson, N.; Ortega, J.; Collier, A.; Jiang, Y.; Li, K.; Dick, R.P.; Lv, Q.; Hannigan, M.; et al. The next generation of low-cost personal air quality sensors for quantitative exposure monitoring. Atmospheric Measurement Techniques 2014, 7, 3325–3336.
4. Anagnostopoulos, G.G., Deriaz, M. Accuracy Enhancements in Indoor Localization with the Weighted Average Technique. In proceeding of The Eighth International Conference on Sensor Technologies and Applications, SENSORCOMM. Lisbon, Portugal, 16–20 November 2014.