RESPONSE TO REVIEW 1

Reviewer Comment 1: The authors have compared mass concentrations with number concentration for some of the devices, I suggest removing the information on slope and intercept from the Table 3 as this is not informative but can mislead the reader with regards to the performance of these units.

Author Response 1: We feel that the slopes and intercepts demonstrate potential variability in sensor response. It is informative to know whether slopes and intercepts from different sensors of the same type are consistent.

Author Change 1: No Change

Reviewer Comment 2: It will help if the authors give more information on how the PM2.5 are calculated by the various manufacturers for the devices reporting this unit of measurement, including the size range each measured.

Author Response 2: We don't know the algorithms behind many of the sensors' reported concentrations or particle counts, as they tend to be proprietary information and the intellectual property of manufacturers.

Author Change 2: No Change

Reviewer Comment3: I am not sure the section describing the comparison of the high-time resolution of the device with respect to the reference unit is well described. Will having a time series plot of the 1-minute data from all devices (PM/ref PM and O3/ref O3) albeit for 24-hour period complement the conclusion drawn by this analysis?

Author Response 3: We don't believe showing a sample time period will help show this conclusion, as this analysis is based on overall measurement to measurement variation at one-minute time scales.

Author Change 3: We have clarified the language for Figure 6 to highlight that this analysis was done to evaluate if sensors tended to have smaller or larger measurement-to-measurement changes than the reference, and what these differences may indicate

Reviewer Comment 4: With regards to the difference in trend patterns (time/wind), have the authors considered the impact of the RH diurnal cycle on the PM sensors. Typically, high RH are observed at night-times, this may be masked in the wind trend analysis (high RH randomly spread across the wind directions). It is worth checking the time trend analysis using periods of low RH (say < 50%).

Author Response 4: We have explored examining RH (and temperature) as a cause of this difference; however, we were not able to explain the differences with these parameters.

Author Change 4: We have added text to the discussion explaining that we explored this avenue.

Reviewer Comment 5: Technical corrections P.2, line 19, add "was" after the phrase "the sensors . . .. " P. 5, line 118: what do the authors mean by ". . .challenge concentrations. . ." P. 6, line 149-150 rephrase ". . . the clause removing wind-blown snow . . .."

Author Response 5: We have made these corrections in the text.

Author Change 5: We have made these corrections in the text.

RESPONSE TO REVIEW 2

Reviewer Comment 1: As there was three of each sensor, I would have liked to see some discussion on the precision of the each sensor.

Author Response 1: We agree and have updated our results section to included precision data.

Author Change 1: We have added Root Mean Square Coefficient of Variation to Table 3 to represent precision and a general description of overall results to section 3

Reviewer Comment 2: Furthermore, throughout Section 3 I would have liked to have seen more discussion on the results and how they compare to previous studies in the literature.

Author Response 2: We have discussed some high-level comparisons with some other studies and programs in Section 4. Our study does not compare directly with other studies performed, as the sensors are likely to have different responses in different environments and exposure to different aerosol compositions and size distributions.

Author Change 2: We have included additional text in Section 4 about expected differences between studies

Reviewer Comment 3: Section 3: Did you see any evidence for baseline drift in any of the sensors over the 7 month period? For example, did the correlation/slope with respect to the reference instrument change in the first month compared to the last? It would be good to include some discussion on the how the different sensors performed in this regard, as in the literature

Author Response 3: We did not see significant baseline drift over this period of time. If there was any significant change in comparison, it resulted from sensor failure.

Author Change 3: We have added a statement to Section 3 regarding baseline shift.

Reviewer Comment 4: Page 11, line 246: To me, the TSI Air Assure was the best performing sensor in terms of accuracy relative to reference, based on table 3. Therefore, I would be interested to know if there was any humidity effects observed in this instrument like was observed for the OPC-N2 and Airbeam (fig 3). Was it just these two sensors that appeared to be affected by humidity?

Author Response 4: The TSI AirAssure did not have RH effects to the same extent as the OPC-N2 and Airbeam. Similar plots are provided in the supplemental information

Author Change 4: No Change

Reviewer Comment 5: Page 11, line 267: My take on Fig 4a is that was the sensors that report particle counts that best captured the diel pattern rather than those that report particle mass concentrations, despite the reference instrument also reporting particle mass concentrations. Perhaps the authors could comment on this.

Author Response 5: These particle count sensors do appear to exhibit the most similar diel patterns., particularly the TZOA and Airbeam. We will highlight this; however, we don't have any explanation as to why this may be the case based on the data we've seen in the study.

Author Change 5: We have added text mentioning that TZOA and Airbeam are particle counters but don't have evidence that this is why they performed better

Reviewer Comment 6: Page 12, line 284: I would be good if the authors could briefly indicate what was tried to explain why the PM sensors better captured the win direction trends compared to the diel as knowing what was not the cause will help avoid duplication of effort in future studies.

Author Response 6: We attempted to explain the differences by examining potential daily humidity and temperature affects, which are likely stronger associated with time of day than wind direction. However, these were not able to explain the differences between the ability to reproduce the different trends.

Author Change 6: We have added a sentence discussing using RH and T to try to explain differences in trend replication.

Reviewer Comment 7: Figure 5: it appears that PM sensors had a wide response range at a north wind direction unlike other direction, that wasn't observed for the ozone sensor. Was there a local soucrce in this direction that may affected the sensor response? This may help understand how aerosol compostion affects the sensor reading.

Author Response 7: The highest concentrations did come from the north where there is a railyard and likely other sources. However, we have no evidence from the data alone to suggest that particles from these sources are measured or reported differently, and any conclusion based on those sources would be speculation. It is also important to note that this may result to either muted or enhanced sensor response when exposed to higher concentration, when compared to reference measurements.

Author Change 7: We have added text to the discussion of Figure 5 about potential sources to the north

Reviewer Comment 8: Page 14, line 301: Why were the OPC-N2 and Airbeam the only sensors to the right of reference in Fig 6? Is due to instrument response time or other artefacts?

Author Response 8: We expect this is due to measurement noise in these sensors, where sensors to the left of reference may have slower response times.

Author Change 8: We have clarified the language for Figure 6 to highlight that this analysis was done to evaluate if sensors tended to have smaller or larger measurement-to-measurement changes than the reference, and what these differences may indicate

Reviewer Comment 9: Table 2: Please include the data capture for the OPC-N2 in this table

Author Response 9: Added OPC-N2 and other missing sensor measurement capture to Table 2. The OPC-N2 data is footnoted OPC-N2 measures at a different frequency (10s) than the other sensors (1 min), but the summary results are still based on 1-minute averages.

Author Change 9: Added OPC-N2 and other missing sensor measurement capture to Table 2. The OPC-N2 data is footnoted OPC-N2 measures at a different frequency (10s) than the other sensors (1 min), but the summary results are still based on 1-minute averages.

Long-term evaluation of air sensor technology under ambient conditions in Denver, Colorado

Stephen Feinberg[1,2], Ron Williams[2], Gayle S.W. Hagler[2], Joshua Rickard[3], Ryan Brown[4], Daniel Garver[4], Greg Harshfield[5], Phillip Stauffer[5], Erick Mattson[5], Robert Judge[6], Sam Garvey[7]

1. Oak Ridge Institute for Science and Education, Oak Ridge, TN 37830
2. U.S Environmental Protection Agency (EPA), Office of Research and Development, Research Triangle Park, NC 27711,
3. U.S. EPA Region 8, Denver, CO 80202,
4. U.S. EPA Region 4, Atlanta, GA 30303,
5. State of Colorado Department of Public Health and Environment (CDPHE),
6. U.S. EPA Region 1, Boston, MA 02109,
7. Jacobs Technology, Inc, Research Triangle Park, NC 27709

ABSTRACT

Air pollution sensors are quickly proliferating for use in a wide variety of applications, with a low price point that supports use in high density networks, citizen science, and individual consumer use. This emerging technology motivates the assessment under real-world conditions, including varying pollution levels and environmental conditions. A seven-month, systematic field evaluation of low-cost air pollution sensors was performed in Denver, Colorado over 2015-2016; the location was chosen to evaluate the sensors in a high altitude, cool, and dry climate. A suite of particulate matter (PM), Ozone (O_3), and nitrogen dioxide (NO_2) sensors were deployed in triplicate, and were collocated with Federal Equivalent Method (FEM) monitors at an urban regulatory site. Sensors were evaluated for their data completeness, correlation with reference monitors, and ability to reproduce trends in pollution data, such as daily concentration values and wind-direction patterns. Most sensors showed high data completeness when data loggers were functioning properly. The sensors displayed a range of correlations with reference instruments, from poor to very high (e.g. hourly-average PM Pearson correlations with reference measurements varied from 0.01 to 0.86). Some sensors showed a change in response to laboratory audits/testing from before the sampling campaign to afterwards, such as the Aeroqual, where the O_3 response slope changed from about 1.2 to 0.6. Some PM sensors measured wind-direction and time of day trends similar to those measured by reference monitors, while others did not. This study showed different results for sensor performance than previous studies performed by the U.S. EPA and others, which could be due to different geographic location, meteorology, and aerosol properties. These results imply that continued field testing is necessary to understand emerging air sensing technology.

1. INTRODUCTION

Next generation air monitoring (NGAM) is a quickly evolving and expanding field. Low-cost air pollution sensors have improved the access for both citizens and researchers to obtain pollutant concentration data in more locations. Many new sensors are now sold and marketed to consumers, and come with messaging on implications for health. In addition to improving the accessibility of measurement data, air pollution sensors have been used to supplement ambient air monitoring by providing high spatial density and high time-resolution measurements (Mead et al., 2013; Snyder et al., 2013; Kaufman et al., 2017). Low-cost air pollution sensors have the potential to be important enablers of smart cities and the Internet of things (IoT), especially in terms of forecasting and health messaging in megacities with significant variability in microenvironments (Mead et al., 2013; Kumar et al., 2015; Ramaswami et al, 2016). Sensors also enable new techniques for mobile monitoring. (McKercher and Vanos 2017; Woodall et al., 2017). However, without a proper understanding of sensor data quality and calibration, low-cost sensors have the potential to mislead interested community and research groups (Rai et al., 2017). Evaluating how well these sensors perform in both laboratory and field environments is critical for understanding their possible uses in research, citizen science, and consumer use, for individual exposure assessment.

Low-cost air pollution sensors, with purchase prices ranging from the low hundreds to the low thousands of dollars per pollutant, have been developed for both particulate and gas phase pollutants, including ozone (O_3) and nitrogen dioxide (NO_2). Particulate matter (PM) sensors typically measure particle counts using light scattering principles. By using light scattering to measure an ensemble of particles, sensors can be produced that are miniaturized, lower cost, and provide real-time data. However, this detection approach can result in bias and inaccuracy from measurement artifacts (Gao et al., 2015; Holstius et al., 2014). Some sensors, such as the OPC-N2 (Alphasense) measure single particles and allocate them into size bins. This approach is subject to measurement artifacts due to humidity effects, potential particle coincidence, and assumes particles are spherical and of a homogenous density (Mukherjee et al., 2017). Gas phase sensors produce a signal through the reaction of the target gases with electrochemical or metal oxide sensors. However, the reactive agents used in these types of sensors may degrade over time, and measurement artifacts may also exist, such as cross-interferences and impacts of temperature (Rai et al., 2017). Therefore, it is necessary to evaluate sensor performance in long-term, real-world study conditions (Alastair Lewis and Peter Edwards, 2016; Williams et al., 2014).

The evaluation of low-cost air pollution sensors and their performance is continually evolving (McKercher et al., 2017b). Many sensors are evaluated in laboratory settings by exposure to known concentrations of gasses and PM, with PM often being evaluated by well-defined aerosol, such as polystyrene latex, in controlled conditions (Wang et al., 2015; Lewis et al., 2016; Manikonda et al., 2016). In outdoor, field settings, sensors are often evaluated to determine their performance in comparison with reference methods (Borrego et al. 2016; Jiao et al., 2017; Crilley et al., 2017; Mukherjee et al., 2017; Hagan et al; 2018). Correlations of low-cost sensors have been found to vary from study to study, spanning from negligible to high correlations. Recent studies have shown the correlation between sensors and reference measurements can be improved by the application of correction factors for environmental conditions such as relative humidity (Crilley et al., 2017) or multivariate models and machine learning (Cross et al., 2017; Zimmerman et al, 2018; Hagan et al., 2018).

There are relatively few efforts that exist to systematically examine air pollution sensor technology performance, that test a variety of replicate sensor types against reference monitors in a real-world environment. In the United States, the U.S. EPA and the South Coast Air Quality Management District (SCAQMD) have developed field- and laboratory-testing programs for both gas and particulate matter sensors. These efforts represent specific geographic locations and concentration ranges (U.S. EPA, 2017; SCAQMD, 2017). For example, EPA's Community Air Sensor Network (CAIRSENSE) project tested a variety of gas-phase and particulate matter sensors in Atlanta, GA, under conditions that were high temperature, high humidity, and fairly low ambient concentrations (e.g., hourly $PM_{2.5}$ ranging 0 to 40 $\mu g/m^3$) (Jiao et al., 2016). The SCAQMD AQ-SPEC program similarly conducts field testing of sensor technology in Diamond Bar, California, at a near-road location nominally two months. Evaluation of identical sensors by the EPA and SCAQMD has revealed that the sensor performance may vary by geographical region. For example, Jiao et al., (2016) found Airbeam sensor correlations to be moderate ($r^2 \approx 0.43$), SCAQMD (2017) reported much stronger correlations ($r^2 \approx 0.74$). This might be a result of from both different concentration ranges as well as the optical properties of the aerosol being measured.

The Community Air Sensor Network (CAIRSENSE) project was a multi-year, multi-location project that focused on evaluating performance characteristics and limitations of low-costs sensors. A prior CAIRSENSE study in Atlanta, Georgia was conducted in 2014 and early 2015 (Jiao, et al., 2016). Atlanta was chosen to test the sensors' performance in the face of higher temperatures and humidity. For the second part of the CAIRSENSE study, Denver Colorado was chosen to test the sensors' performance under conditions of high altitude, dry and lower temperature conditions. Beyond assessing sensor performance through correlation with a reference monitor, this study also investigates the degree to which data from sensors is able to produce similar temporal, wind-direction, and transient event trends in comparison to a high time-resolution reference monitors.

2. METHODS

Sensors for this study were selected based on cost, commercial availability, market prevalence, capability, and applicability to EPA research objectives. Table 1 lists the sensors chosen for this study, pollutants measured by each sensor, and the measurement principle used by each sensor. Cost information for these sensors are published on the EPA's Air Sensor Toolbox (U.S. EPA 2017). Two different Dylos units were used for this study. Unit 1 was a Dylos DC1100, while units 2 and 3 were Dylos DC1100 Pro models, where the Pro models are advertised to have increased sensitivity for smaller particles. The Shinyei, Dylos, Airbeam, Aeroqual, and Cairclip sensors were used in both the Denver and Atlanta studies (Jiao et al., 2016). Additionally, several of these sensors have been evaluated in laboratory or short term ambient settings (Air Sensor Toolbox reference; Sousan et al., 2016; SCAQMD 2017; etc.).

Air pollution sensors were acquired and deployed in triplicate. Before deployment, laboratory sensor response audits were performed for all of the available sensors. PM sensors were zero-checked in a clean room environment, all reporting <2 $\mu g/m^3$ values under those conditions, except for the Air Assure. The software for the Air Assure performs its own zeroing, therefore they were operated 'as-is'. A pre-deployment sensor response audit was not performed for the TZOA as it was received shortly before deployment. Sensor output was not adjusted based on the calibration audits in order to reflect

Table 1 Sensors used during the CAIRSENSE-Denver Study

Sensor	Pollutant(s) Measured	Principle of Operation
Aeroqual SM-50	O_3	Electrochemical Sensor
TSI Air Assure	PM	Light Scattering
AirCasting AirBeam	PM	Light Scattering
Cairpol Cairclip	$NO_2 + O_3$	Electrochemical Sensor
Dylos DC1100/DC1100 Pro	PM	Laser particle counter
Alphasense OPC-N2	PM	Laser particle counter
Shinyei PMS-SYS-1	PM	Light Scattering
AirViz Speck	PM	Light Scattering
TZOA PM Research Sensor	PM	Laser particle counter

their 'out of the box' performance. Sensor responses were also audited by either ~~challenge concentrations~~recording their responses to known concentrations (Aeroqual and CairClip sensors), or in a clean air environment (PM sensors) after the end of the measurement period, to evaluate possible sensor drift. Laboratory audit results are presented in the supplemental information.

Sensors were deployed at the downtown Denver Continuous Ambient Monitoring Program (CAMP) regulatory monitoring site (Latitude: 39.751184; Longitude: -104.987625) from September 2015 to March 2016. The CAMP site was operated by the state of Colorado for the duration of the study. Sensors were placed in a ventilated, multi-level shelter designed to allow ambient air circulation and prevent intrusion from precipitation, as shown in Figure. 1. A full description of the shelter has been previously reported (Jiao, 2016). The sensors were connected to data loggers stored in weatherproof enclosures attached to the bottom of the shelter. Most of the sensors were connected to Arduino (single-board) microprocessors with either Ethernet (IEEE 802.3 standard) or Recommended Standard 232 (RS-232) serial communication cables. The OPC-N2 and Speck sensor data were logged using laptops, and the TZOA data was stored internally on secure digital (SD) cards. To comply with EPA data security requirements, the cloud based storage capability of the Air Assure sensors was disabled, and these units reported data locally via the Arduino microprocessors with onboard memory. The Cairclip sensor measures the combined signal from NO_2 and O_3. Therefore, both NO_2 and O_3 measurements from the Cairclip were determined by subtracting the opposite (collocated) reference measurement. The Dylos units also measure multiple particle size fractions. In this study, the "small" particle size fraction, as described by the manufacturer, was used for $PM_{2.5}$ comparisons. TZOA sensors did not have a real-time

Figure 1 Sensor deployment shelter

clock, and only measured time as the elapsed number of milliseconds since the device was powered on. Therefore, field operators were required to accurately record start and end times as a means of establishing the sensor response time series.

A total of four Arduino microprocessors and three laptops were used simultaneously for data logging. Between the data loggers, laptops and onboard data storage, there were many different sensor data output formats. Separate data scripts were developed to process each different data format into similarly formatted files for each air pollution sensor type. Once data collections were initiated in September 2015, the sensors were operated with little or no intervention through the entirety of the study. Noted interventions included restarting data systems when they 'locked up', or removing ~~wind-blown~~ snow from the shelves housing the sensors during a major winter snowstorm.
Federal Equivalent Method (FEM) measurements at the Denver monitoring site were collected using a Teledyne 400E O_3 monitor, Teledyne 200EU NO_2 analyzer and a GRIMM EDM 180 Dust monitor, which measured $PM_{2.5}$ and PM_{10} mass at one-minute intervals using optical detection. All sensors and monitors collected pollutant data at one-minute intervals or less. One-minute values were used to generate concentrations at multiple time intervals, with primarily one-hour averages used for data analysis. All averaging and other data processing was performed using the following software: RStudio version 0.98.1103, R version 3.2.2, and the ggplot2, scales, plyr, lattice, corrplot, and 'data.table' (extension of 'data.frame') packages.

Sensor data were recovered from the connected laptops and SD cards connected to the data loggers. Most sensors reported data in one-minute intervals. The Alphasense OPC-N2 units recorded concentrations every ten seconds. These measurements were used to calculate one-minute averages. The TZOA sensors reported data based on time elapsed from turning on each unit. The start times for

each unit and total elapsed time for each measurement were combined to generate 5-second time stamps for the TZOA measurements. These values were then used to calculate one-minute averages.

In order to best replicate actual use by non-experts and avoid biasing the results towards a positive direction, minimal screening of data was performed. Quality assurance screening consisted primarily of removing data where there was a clear malfunction of the sensor, such as non-numeric data output, or when a sensor (e.g., Cairclip unit 1) became 'stuck', reporting a repeated, value (value = 255) for long time spans. These types of errors had previously been identified for the output of this sensor type. The Aeroqual units had significant numbers of measurements that, for some reason, were reported as zero. These were possibly due to the inability of the sensor to detect trace concentrations, and were therefore not screened out of the data.

Timestamps for all sensors except the TZOA were recorded in Mountain Standard Time. As previously mentioned, TZOA timestamps were generated by combining the initial recording time and the elapsed time reported by the sensors. One-minute measurements and averages were used to calculate 5-minute and hourly averages. Hourly averages were further used to calculate 12-hour and daily averages. FEM measurements from the State of Colorado instruments were also recorded at one-minute intervals and averaged in the same manner as the sensor data. Data from all sensors and reference instruments were stored in separate data files and combined based on timestamps for analyses using 'R' scripts.

Sensors were also investigated for how well they replicated different trends in the regulatory monitor measurement data. The trends analyzed included average sensor responses based on time of day and wind direction. In order to evaluate these trends, different normalized sensor responses were used. The normalized average sensor response for the diel (daily, 24-hour) patterns was calculated as the average concentration for a given hour divided by the average concentration for the hour beginning at 12:00 PM. The normalized average sensor response for wind direction data was defined as the mean concentration for each 10-degree wind 'bin', divided by the average concentration of the 170 to 180-degree bin. The sensor response times were also analyzed by calculating the average one-minute relative sensor response, as defined by the distribution of the one-minute concentration differences divided by the average sensor response.

3. RESULTS & DISCUSSION

Table 2 shows a summary of data completeness from the air pollution sensors, including the total percentage of minutes measured, percentage of measurements missed, by not logging data, and the percentage of completely missing data. The majority of missing data was due to events where the sensor and data loggers were inoperative. The most significant of these events was due to snow intrusion into the monitoring platform in December 2015, which caused units to shut down. Most sensors had a very high data capture rate throughout the study when the units were on (and operational). The Cairclip units had significant amounts of missing data, likely due to data transmission errors from the universal asynchronous receiver-transmitter (UART) serial communication system. In the previous Atlanta study as well as in a Newark-based citizen science study (Kaufman et al., 2017), Cairclip units with identical sensors but different universal serial bus (USB) data connections were used and did not have significant amounts of missing data.

Table 2 Sensor Data Completeness

Sensor	Measurement %	Sensor on and not Logging %	Completely Missing %	Comments
Aeroqual	82%	0%	18%	45% of logged values were 0
	73%	0%	27%	42% of logged values were 0
	81%	5%	13%	32% of logged values were 0
Air Assure	87%	0%	13%	
	87%	0%	13%	
	87%	0%	13%	
Airbeam	74%	0%	25%	
	62%	6%	32%	
	62%	6%	32%	
Cairclip	29%	53%	18%	56% of logged values were 255[a]
	63%	13%	24%	No data before 10/8/15
	63%	23%	13%	
Dylos	82%	0%	18%	
	82%	0%	18%	
	72%	1%	27%	
OPC-N2	77%	0%	23%	
	76%	0%	24%	
	71%	0%	29%	59% of logged values were 0
Shinyei	82%	0%	18%	
	73%	0%	27%	
	87%	0%	13%	
Speck	92%	0%	8%	
	93%	0%	7%	
	96%	0%	4%	
TZOA	61%	0%	39%	
	47%	0%	53%	
	47%	0%	53%	

[a]255 represented a communication or other unknown sensor failure

Measurements from air pollution sensors and regulatory monitors were time-averaged at multiple intervals for comparison. The time intervals included 5-minute, hourly, 12-hour, and daily averages. For each set of time averaging, regressions were calculated to evaluate sensor correlation and bias when compared to regulatory measurements. Additionally, intercomparisons were made between sensors of the same pollutant type (e.g., correlations between PM sensors). Table 3 displays a summary of regression statistics for sensors when compared to regulatory measurements as well as precision calculations for 1-hour time averages. The precision was calculated as the root mean square (RMS) of the hourly coefficients of variation. In general, correlations were greatest at the 1-hour time average. Correlations in general improved slightly with increasing length of the averaging period up to hourly averages. Reduced correlations for most sensors at the 12-hour and daily averages may be a result of a lower number of data points. In contrast to most other measurements, sensors that reported data for

Table 3 Regression and Precision Results for CAIRSENSE sensors (1-hour time averaged)

Formatted Table

Sensor	Pollutant	Reference Average Concentration[1]	Slope	Intercept	Pearson Correlation, R	RMS Precision (%)	Number of Hourly Measurements
Aeroqual SM-50	O_3, ppb	18.8 ppb	0.56	-0.004	0.93	73	3325
			0.58	-0.004	0.92		2963
			0.77	-0.004	0.96		3279
TSI Air Assure	PM, µg/m³	7.8 µg/m³	1.14	2.64	0.8	41	3486
			1.13	-0.04	0.78		3486
			1.19	-1.38	0.81		3486
AirCasting AirBeam	Particle Count, hundreds of particles per cubic foot (hppcf)	7.8 µg/m³	273	-323	0.82	6	3028
			278	-124	0.84		2539
			322	-352	0.82		2532
Cairpol Cairclip	O_3, ppb	18.8 ppb	NA[2]	NA[2]	NA[2]	NA[2]	738
			-0.04	-23.6	-0.06		2831
			1.03	-39.0	0.46		2852
Cairpol Cairclip	NO_2, ppb	26.8 ppb	NA[2]	NA[2]	NA[2]	NA[2]	738
			0.65	-10	0.87		2831
			0.67	-15	0.84		2852
Dylos DC1100/DC1100 Pro	"Small" Particle Count, hppcf	7.8 µg/m³	64	-152	0.86	15	3324
			428	-1182	0.78		3324
			431	-941	0.73		2937
Dylos DC1100/DC1100 Pro	"Large" Particle Count, hppcf	12.0 µg/m³	1.3	5.5	0.40	10	3324
			5.7	73	0.33		3324
			4.9	84	0.27		2937
Alphasense OPC-N2	$PM_{2.5}$, µg/m³	7.8 µg/m³	0.4	-0.30	0.45	108	2969
			0.49	-1.66	0.34		2939
			0.07	0.60	0.11		2735
Alphasense OPC-N2	PM_{10}, µg/m³	19.6 µg/m³	0.45	2.98	0.47	101	2969
			0.54	-1.06	0.68		2939
			0.12	2.86	0.20		2735
Shinyei PMS-SYS-1	$PM_{2.5}$, µg/m³	7.8 µg/m³	0.58	0.24	0.71	20	3325
			0.54	0.8	0.72		2963
			0.42	4.35	0.01[a]		3486
AirViz Speck	$PM_{2.5}$, µg/m³	7.8 µg/m³	0.76	13	0.24	37	3557
			0.74	15	0.40		3584
			0.62	10	0.35		3971
TZOA PM Research Sensor	Particle Count, hppcf	7.8 µg/m³	NA[2]	NA[2]	NA[2]	17[b]	2341
			6.68	1.37	0.66		1838
			6.75	2.16	0.72		1836

coarse PM (Dylos) or PM_{10} (OPC-N2) showed improved correlations with increasing averaging time for those measurements. The correlations for all the time averaging periods can be found in the

[1]Average Concentration calculated for hours with valid sampling data.
[2]Correlation results not shown due to large amount of missing or invalid data
[a] Shinyei Unit 3's correlation improved to 0.84 when only considering data from October 16 and later
[b]TZOA Unit 1 was excluded from RMS precision calculations

supplemental information. Sensors that measured particle count had better precision than those measuring particle mass concentrations. Figure 2 shows a Pearson correlation (R) plot for 1-hour average reference (SoC) and PM sensor measurements~~measurements of the PM sensors~~. The PM units show high correlation among sensors of the same model, except for when one sensor in a group had significant issues. Of the $PM_{2.5}$ sensors, the Air Assure, Airbeam, and Dylos (R = 0.73 to 0.86) units exhibited the highest correlation with reference measurements. Dylos unit 1 had the highest linearity, however it had the lowest particle count response, both of which are likely explained by not detecting the smallest particles as effectively as other units. Cairclip unit 1 rarely properly transmitted data throughout the study, leading to its low correlations. Cairclip units 2 and 3 had more sporadic data transmission issues. All Cairclip units recovered data properly once returned to the lab after the field campaign where their internal data storage was used. The response from Shinyei unit 3 changed in mid-

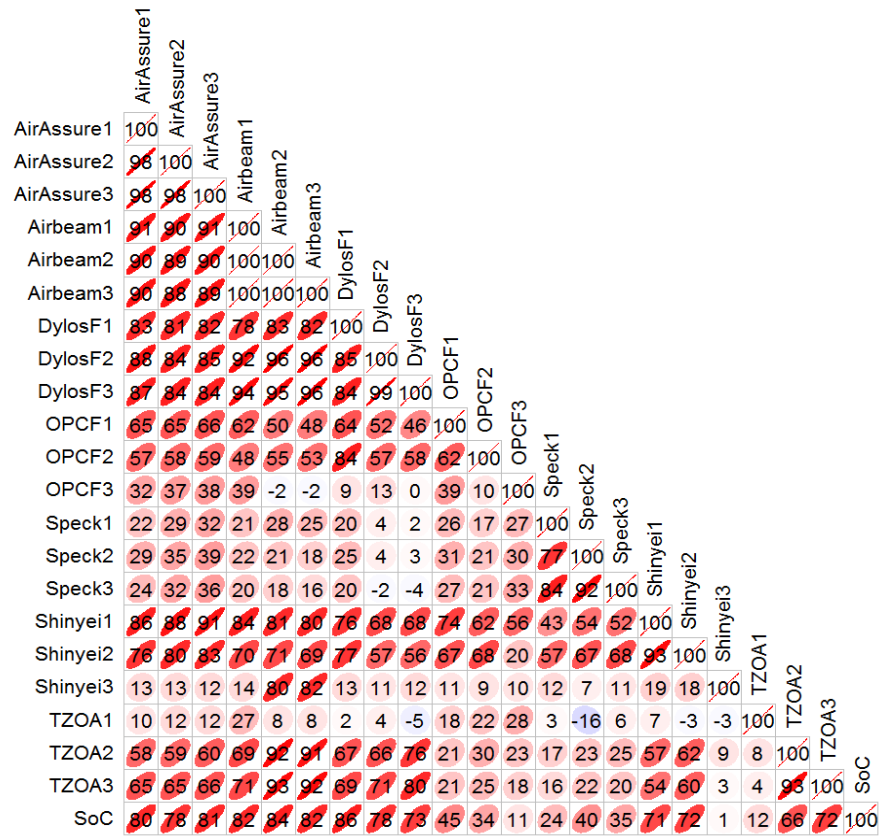


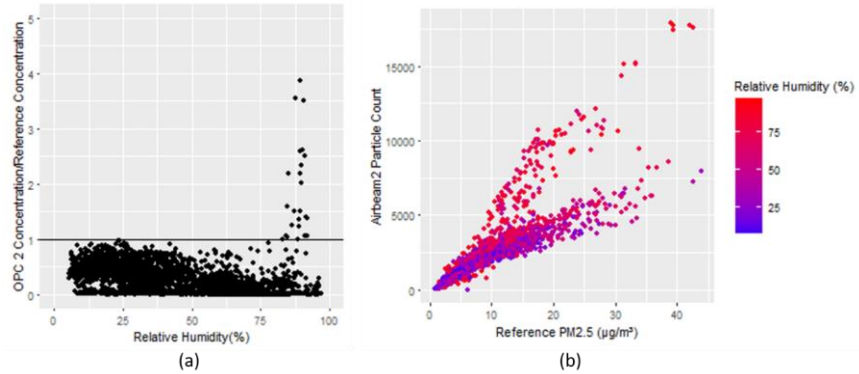
Figure 2 Correlation (*r*100*) plot for sensors measuring fine PM. Ellipses represent the overall scatter of the data (1-hour averaged measurements)

October. The correlation between the unit and the reference monitor was initially 0.01, then increased to 0.84 when comparing only the data starting October 16 and later.

Several sensor models were used in both the Atlanta and Denver CAIRSENSE evaluation campaigns. Both studies deployed the Airbeam, Dylos, and Shinyei PM sensors. In all cases except for Shinyei unit 3, these sensors showed greater linearity in Denver than in Atlanta, when comparing 12-hour averages. When only considering data after October 16, Shinyei unit 3 also had higher correlation in Denver than in Atlanta. This may be due to less noise caused by lower humidity in Denver than in Atlanta. Aeroqual and Cairclip air pollution sensors were also deployed in both Atlanta and Denver. O_3 measured by the Aeroqual units showed similar correlations in both locations (R^2 = 0.82 to 0.94 in Atlanta, R^2 = 0.85 to 0.92 in Denver). O_3 measured by Cairclip units 2 and 3 in Denver showed poorer correlations than the Cairclip units used in Atlanta (R^2 = 0.00 to 0.21 in Denver versus R^2 = 0.68 to 0.88 in Atlanta). However,

250 NO_2 measured by Cairclip units 2 and 3 in Denver was more highly correlated than in Atlanta (R^2 = 0.71
251 to 0.76 in Denver versus 0.57 in Atlanta).

252 While Denver is not necessarily known for high humidity, humidity artifacts were observed in some
253 sensors. Figure 3a shows the $PM_{2.5}$ concentrations measured by one of the OPC-N2 against relative
254 humidity. At RH around 90%, the PM concentration spikes significantly, suggesting that humidity is
255 interfering with the sensor response measurement. This behavior is similar to that observed by Sousan
256 et al., (2016). Some other instruments also had different responses based on humidity. Figure 3b shows
257 hourly particle counts measured by an Airbeam sensor against $PM_{2.5}$ concentration measured by the
258 reference instrument, stratified by relative humidity. There appear to be two separate relations
259 between reference measured concentrations and sensor measured particle counts, with a greater
260 particle count response occurring more at higher humidity. This relationship was observed in each of the
261 Airbeam sensors. An example of humidity relationships from each sensor type can be found in the
262 supplemental information.



263

264 **Figure 3 OPC 2 $PM_{2.5}$ and Relative Humidity (a) and Hourly Average FRM $PM_{2.5}$ concentration and**
265 **Airbeam Particle count stratified by Relative Humidity (b)**

266

267 In addition to understanding the precision of air pollution sensors and how well they correlate with
268 reference measurements, it is also important to understand how well a sensor can capture trends and
269 distributions of pollutant concentrations. There are many ways to examine these trends and
270 distributions. Figure 4 shows the diel patterns of $PM_{2.5}$ (a) and O_3 (b) reference and sensor
271 measurements respectively. The results, for each sensor, represent the measurements of the best
272 performing unit for each sensor type/model, as determined by R-squared values. The various PM air
273 pollution sensors have a wide range of comparisons to the reference monitor. Two sensors (TZOA and
274 Airbeam) show similar patterns throughout the day, while some other sensors do not reflect the
275 reference diel pattern at all (e.g., OPC, Speck, etc.). It is interesting to note that both the TZOA and
276 Airbeam measure particle count; however, there is no basis to say why these sensors performed better
277 than those measuring mass concentrations.The Aeroqual sensor diel pattern was similar to that of the

reference O_3 monitor. The nature of the calculation of O_3 and NO_2 by subtraction, and missing data from the Cairclip sensors, prevented this analysis from providing meaningful results.

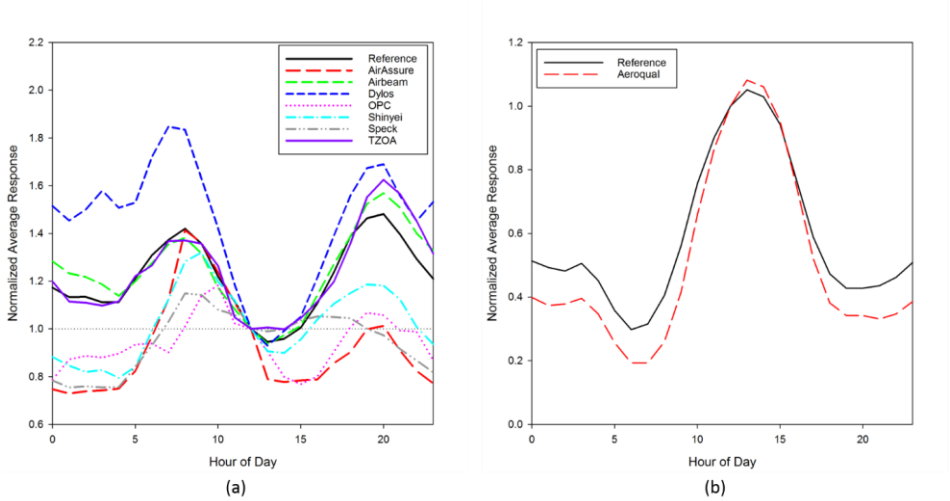


Figure 4 Diel patterns for $PM_{2.5}$ (a) O_3 (b) sensor and reference measurements.

Air quality measurements are also known to be dependent on wind direction, and it is important to know if these differences were reflected in the sensor measurements. Figure 5 shows the normalized average sensor response $PM_{2.5}$ (a) and O_3 (b) response of the sensors and the reference monitors respectively. The reference monitor response is represented by the black line. Both the highest concentrations and greatest variation from the reference monitor concentrations occurred when winds were from the north, where there are multiple large roadways and a railyard. However, there was no other evidence to suggest that these sources contributed to differences in the measurement trends. The sensors generally compared more favorably with the reference monitors when examining the wind direction dependence of concentration. This is most apparent in the OPC-N2 sensor, where the sensor trends track the trends measured by the reference monitor. This increases the confidence that sensors may be useful in studies that pair wind direction with concentration to determine potential bearings or

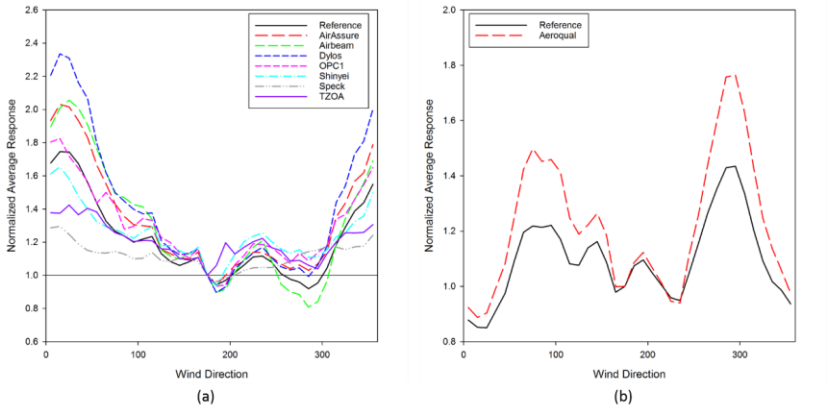
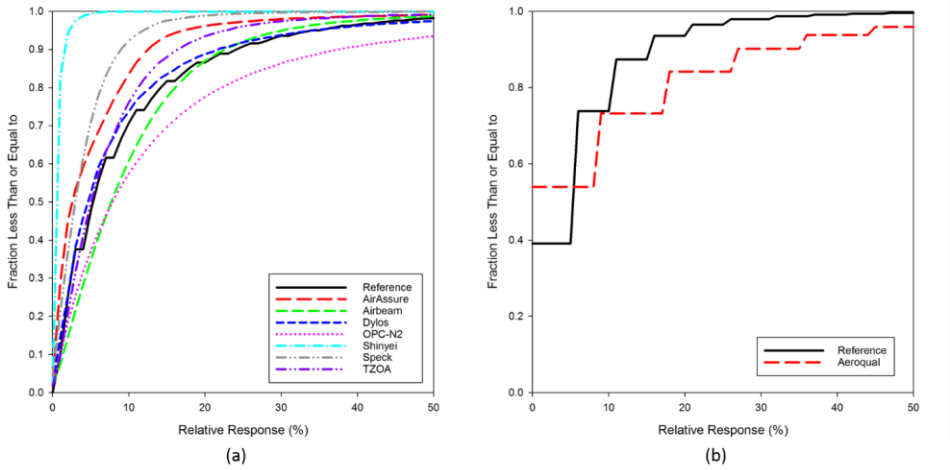


Figure 5 Wind direction patterns for $PM_{2.5}$ (a) O_3 (b) sensor and reference measurements.

locations of pollution sources to supplement source apportionment and receptor modeling. It also raises questions as to why an air pollution sensor would be able to reproduce wind direction trends but not necessarily reproduce daily concentration measurement patterns. We undertook exploration of this perplexing result, but were not able to determine a clearly identifiable cause. While RH and temperature do have time of day variation that is not reflected in wind direction, we were unable to use these parameters to explain the differences between time of day and wind direction trends.

The high-time resolution data collected for this study allowed for the examination of air pollution sensor response trends compared to that of regulatory air pollution monitors. Figure 6 shows a cumulative distribution function (CDF) for the relative change in sensor and regulatory monitor response between 1-minute measurements for $PM_{2.5}$ (a) and O_3 (b) sensor and reference measurements respectively. The relative response was calculated as the absolute value of the difference between consecutive one-minute measurements divided by the mean measurement over the entire study period for each sensor/monitor. If the reference monitor were considered a perfect measurement, sensor curves to the left and above the reference monitor line have smaller relative changes than the reference monitor, indicating a ~~would be~~ slower ~~to respond~~response to changes in concentration, while curves below and to the right of the monitor line would signify larger measurement-to-measurement changes that the reference monitor, indicating potential high levels of measurement noise. Most PM monitors exhibited a slower response to changes in concentration than the reference monitor. The OPC-N2 and Airbeam sensors were the only ones with curves to the right of the reference monitor, suggesting that they may have more noise in their measurements. The Aeroqual sensor showed more O_3 measurement noise when compared to the reference measurement.



316

317 **Figure 6 Cumulative distribution functions for 1-minute response differences for $PM_{2.5}$ (a) O_3 (b)**
318 **sensor and reference measurements.**

319

320 **4. Conclusions**

321 Nine different air pollution sensor devices were deployed in triplicate with collocated air pollution
322 reference monitors in Denver, CO over an extended operational timeline of longer than six months. The
323 sensors showed a wide range of correlations with reference measurements, but tended to have high
324 correlation with sensors of the same model. PM sensors deployed in both Denver and Atlanta had
325 higher correlations with reference monitors in Denver than in Atlanta. This is likely due to less humidity
326 related response in Denver. Aeroqual O_3 measurements in Denver showed similar linearity to those
327 measured in Atlanta. Cairclip O_3 correlations were lower in Denver than in Atlanta, but NO_2 correlations
328 were higher. Sensors that have also been evaluated by the South Coast Air Quality Management District
329 (SCAQMD) tended to show similar results in terms of correlation (SCAQMD, 2017). However, in all cases,
330 sensors' performance in this long-term field deployment was less than that of laboratory based
331 comparisons performed in this study and others (U.S. EPA, 2017). It is not surprising that the results of
332 this study for PM sensors varied from other studies, as the responses to optical measurement
333 techniques used by these sensors are likely influenced by aerosol composition. This study demonstrates
334 the need for long-term, real-world evaluation studies for current and future air pollution sensors, that
335 ~~will need to~~should be performed in locations with different air pollutant concentration ranges and
336 aerosol characteristics.

337 Several air pollution sensors were able to capture variations in important trends, such as diel patterns
338 and wind direction dependence on concentration. However, the OPC-N2 units showed similar results as
339 reference monitor measurement data when analyzing the wind direction trends, but not when analyzing
340 'time-of-day' trends. These promising results show that sensors have the possibility for supplementing

measurement research capabilities when interested in air pollution trends such as those dependent on wind direction. Analyses of wind direction based air pollutant trends could be useful for possible identification of source locations or regions, especially with the use of a sensor-based network.

Data Availability

The CAIRSENSE dataset will be available at the EPA environmental dataset gateway (https://edg.epa.gov) (EPA 2018) where the dataset can be retrieved by searching for "CAIRSENSE Denver." Project data can also be requested from the corresponding author.

Author Contribution

Stephen Feinberg performed the data analysis and prepared the manuscript with contribution from all co-authors. Ron Williams was the Principal Investigator and advised data analysis and manuscript preparation. Gayle Hagler was Co-Investigator and advised data analysis and manuscript preparation. Joshua Rickard, Ryan Brown, Daniel Garver, and Robert Judge provided discussion and assistance in the establishment of the study design and execution. Greg Harshfield, Phillip Stauffer, and Erick Mattson carried out day-to-day operations. Sam Garvey was in charge of the establishment of the initial site, data recovery, and principal operation of the sensors.

Disclaimer

This article was supported by Jacobs Technology, Inc. (contract EP-C-15-008) for the National Exposure Research Laboratory, U.S. Environmental Protection Agency (U.S. EPA), Research Triangle Park, NC. It has been reviewed by the U.S. EPA and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

5. References

Alastair Lewis and Peter Edwards, Validate personal air -pollution sensors. Nature 535, 29-31 (2016).

Borrego, C., Costa, A. M., Ginja, J., Amorim, M., Coutinho, M., Karatzas, K., Sioumis, T., Katsifarakis, N., Konstantinidis, K., De Vito, S., Esposito, E., Smith, P., André, N., Gérard, P., Francis, L. A., Castell, N., Schneider, P., Viana, M., Minguillón, M. C., Reimringer, W., Otjes, R. P., von Sicard, O., Pohle, R., Elen, B., Suriano, D., Pfister, V., Prato, M., Dipinto, S., and Penza, M.: Assessment of air quality microsensors versus reference methods: The EuNetAir joint exercise, Atmospheric Environment, 147, 246-263, https://doi.org/10.1016/j.atmosenv.2016.09.050, 2016.

Crilley, L. R., Shaw, M., Pound, R., Kramer, L. J., Price, R., Young, S., Lewis, A. C., and Pope, F. D.: Evaluation of a low-cost optical particle counter (Alphasense OPC-N2) for ambient air monitoring, Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2017-308, in review, 2017.

Cross, E. S., Williams, L. R., Lewis, D. K., Magoon, G. R., Onasch, T. B., Kaminsky, M. L., Worsnop, D. R., and Jayne, J. T.: Use of electrochemical sensors for measurement of air pollution: correcting interference response and validating measurements, Atmos. Meas. Tech., 10, 3575-3588, https://doi.org/10.5194/amt-10-3575-2017, 2017.

Gao, M., Cao, J., and Seto, E.: A distributed network of low-cost continuous reading sensors to measure spatiotemporal variations of PM2.5 in Xi'an, China, Environmental Pollution, 199, 56-65, https://doi.org/10.1016/j.envpol.2015.01.013, 2015.

Hagan, D. H., Isaacman-VanWertz, G., Franklin, J. P., Wallace, L. M. M., Kocar, B. D., Heald, C. L., and Kroll, J. H.: Calibration and assessment of electrochemical air quality sensors by co-location with regulatory-grade instruments, Atmos. Meas. Tech., 11, 315-328, https://doi.org/10.5194/amt-11-315-2018, 2018.

Holstius, D. M., Pillarisetti, A., Smith, K. R., and Seto, E.: Field calibrations of a low-cost aerosol sensor at a regulatory monitoring site in California, Atmos. Meas. Tech., 7, 1121-1131, https://doi.org/10.5194/amt-7-1121-2014, 2014.

Jiao, W., Hagler, G., Williams, R., Sharpe, R., Brown, R., Garver, D., Judge, R., Caudill, M., Rickard, J., Davis, M., Weinstock, L., Zimmer-Dauphinee, S., and Buckley, K.: Community Air Sensor Network (CAIRSENSE) project: evaluation of low-cost sensor performance in a suburban environment in the southeastern United States, Atmos. Meas. Tech., 9, 5281-5292, https://doi.org/10.5194/amt-9-5281-2016, 2016.

Kaufman, A., Williams, R., Barzyk, T., Greenberg, M., O'Shea, M., Sheridan, P., Hoang, A., Ash, C., Teitz, A., Mustafa, M., and Garvey, S.: A Citizen Science and Government Collaboration: Developing Tools to Facilitate Community Air Monitoring, Environmental Justice, 10, 51-61, 10.1089/env.2016.0044, 2017.

Kumar, P., Morawska, L., Martani, C., Biskos, G., Neophytou, M., Di Sabatino, S., Bell, M., Norford, L., and Britter, R.: The rise of low-cost sensing for managing air pollution in cities, Environ. Int., 75, 199–205, 2015.

Lewis, A. C., Lee, J. D., Edwards, P. M., Shaw, M. D., Evans, M. J., Moller, S. J., Smith, K. R., Buckley, J. W., Ellis, M., Gillot, S. R., and White, A.: Evaluating the performance of low cost chemical sensors for air pollution research, Faraday Discussions, 189, 85-103, 10.1039/C5FD00201J, 2016.

403 Manikonda, A., Zíková, N., Hopke, P. K., and Ferro, A. R.: Laboratory assessment of low-cost PM
404 monitors, Journal of Aerosol Science, 102, 29-40, https://doi.org/10.1016/j.jaerosci.2016.08.010, 2016.

405 McKercher, G. R., and Vanos, J. K.: Low-cost mobile air pollution monitoring in urban environments: a
406 pilot study in Lubbock, Texas, Environmental Technology, 1-10, 10.1080/09593330.2017.1332106,
407 2017a.

408 McKercher, G. R., Salmond, J. A., and Vanos, J. K.: Characteristics and applications of small, portable
409 gaseous air pollution monitors, Environmental Pollution, 223, 102-110,
410 https://doi.org/10.1016/j.envpol.2016.12.045, 2017b.

411 Mead, M., Popoola, O., Stewart, G., Landshoff, P., Calleja, M., Hayes, M., Baldovi, J., McLeod, M.,
412 Hodgson, T., Dicks, J., Lewis, A. C., Cohen, J., Baron, R., Saffell, J., and Jones, R.: The use of
413 electrochemical sensors for monitoring urban air quality in low-cost, high-density networks, Atmos.
414 Environ., 70, 186–203, 2013.

415 Mukherjee, A., Stanton, G. L., Graham, R. A., and Roberts, T. P.: Assessing the Utility of Low-Cost
416 Particulate Matter Sensors over a 12-Week Period in the Cuyama Valley of California, Sensors, 17,
417 10.3390/s17081805, 2017.

418 Rai, A. C., Kumar, P., Pilla, F., Skouloudis, A. N., Di Sabatino, S., Ratti, C., Yasar, A., and Rickerby, D.: End-
419 user perspective of low-cost sensors for outdoor air pollution monitoring, Science of The Total
420 Environment, 607-608, 691-705, https://doi.org/10.1016/j.scitotenv.2017.06.266, 2017.

421 Ramaswami, A., Russell, A. G., Culligan, P. J., Sharma, K. R., and Kumar, E.: Meta-principles for
422 developing smart, sustainable, and healthy cities, Science, 352, 940-943, 10.1126/science.aaf7160, 2016.

423 Snyder, E. G., Watkins, T. H., Solomon, P. A., Thoma, E. D., Williams, R. W., Hagler, G. S. W., Shelow, D.,
424 Hindin, D. A., Kilaru, V. J., and Preuss, P. W.: The Changing Paradigm of Air Pollution Monitoring, Environ.
425 Sci. Technol., 47, 11369–11377, doi:10.1021/es4022602, 2013.

426 Sousan, S., Koehler, K., Hallett, L., and Peters, T. M.: Evaluation of the Alphasense Optical Particle
427 Counter (OPC-N2) and the Grimm Portable Aerosol Spectrometer (PAS-1.108), Aerosol science and
428 technology: the journal of the American Association for Aerosol Research, 50, 1352-1365,
429 10.1080/02786826.2016.1232859, 2016.

430 South Coast Air Quality Management District Air Quality Sensor Performance Evaluation Center.
431 http://www.aqmd.gov/aq-spec, 2017, accessed December 5, 2017.

432 U.S. EPA. Air Sensor Toobox. https://www.epa.gov/air-sensor-toolbox, 2017, accessed December 5,
433 2017.

434 Wang, Y., Li, J., Jing, H., Zhang, Q., Jiang, J., and Biswas, P.: Laboratory Evaluation and Calibration of
435 Three Low-Cost Particle Sensors for Particulate Matter Measurement, Aerosol Science and Technology,
436 49, 1063-1077, 10.1080/02786826.2015.1100710, 2015.

437 Williams, R., Watkins, T., Long, R. Low cost sensor calibration options. Environmental Manager, p 10-15,
438 January 2014.

439 Woodall, M. G., Hoover, D. M., Williams, R., Benedict, K., Harper, M., Soo, J.-C., Jarabek, M. A., Stewart,
440 J. M., Brown, S. J., Hulla, E. J., Caudill, M., Clements, L. A., Kaufman, A., Parker, J. A., Keating, M.,

441 Balshaw, D., Garrahan, K., Burton, L., Batka, S., Limaye, S. V., Hakkinen, J. P., and Thompson, B.:
442 Interpreting Mobile and Handheld Air Sensor Readings in Relation to Air Quality Standards and Health
443 Effect Reference Values: Tackling the Challenges, Atmosphere, 8, 10.3390/atmos8100182, 2017.

444 Zimmerman, N., Presto, A. A., Kumar, S. P. N., Gu, J., Hauryliuk, A., Robinson, E. S., Robinson, A. L., and R.
445 Subramanian: A machine learning calibration model using random forests to improve sensor
446 performance for lower-cost air quality monitoring, Atmos. Meas. Tech., 11, 291-313,
447 https://doi.org/10.5194/amt-11-291-2018, 2018