Correction and Long-Term Performance Evaluation of Fine Particulate Mass Monitoring with Low-Cost Sensors

Carl Malings[2], Rebecca Tanzer[2], Aliaksei Hauryliuk[2], Provat K. Saha[2], Allen L. Robinson[2], Albert A. Presto[2], R. Subramanian[1]

[1]Center for Atmospheric Particle Studies, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213. Email: subu@cmu.edu (Corresponding Author)

[2]Center for Atmospheric Particle Studies, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213.

Abstract

Low-cost particulate sensors can allow for establishing a dense monitoring network to increase the spatial resolution of air quality information, which is particularly of interest in urban areas. However, these sensors are often affected by environmental factors such as temperature and humidity, the effects of which must be accounted for so that the accuracy of these sensors in field conditions can be quantified. In this paper, we conduct long-term tests of two types of low-cost particulate sensors: Met-One NPM and PurpleAir units. We assess the self-consistency of larger groups (12 to 25) of sensors, develop empirical equations for correcting the measurements of these sensors to better match those of regulatory-grade instruments, and assess the long-term performance of these sensors during deployments lasting over a year. These assessments are used to assess sensor performance in two different use cases: improving community awareness of air quality with a focus on short-term qualitative indications and providing accurate long-term quantitative information for health impact studies. We find that, for the short-term case, using either quadratic or piecewise-linear correction equations, either sensor can be used to provide reasonably accurate concentration information for $PM_{2.5}$ (mean absolute error on the order of 4 $\mu g/m^3$) in near-real time. For the long-term case, by applying in-field noise-adjustment, bias can be reduced below 1 $\mu g/m^3$. These results indicate the suitability of these sensors for supplementing regulatory-grade instruments in sparsely monitored regions, as well as for conducting hotspot mapping to better understand the variability of air quality in urban areas.

1. Introduction

The negative health impacts of exposure to particulate matter smaller than 2.5 microns ($PM_{2.5}$) are well documented (e.g. Schwartz et al. 1996; Pope et al. 2002; Brook et al. 2010). Even relatively small changes in particulate concentrations can have significant impacts on human health and mortality (Lepeule et al. 2012), while reductions in these levels, even in low concentration environments, can have substantial benefits (Apte et al. 2015). Accurate monitoring of $PM_{2.5}$ is thus important for a variety of applications, including long-term health studies, assessing the impacts of technology and/or regulatory changes on emissions, and supporting decision-making

36 for future regulatory efforts. Monitoring is especially of interest in urban areas where the high
37 density of exposed populations is coupled with higher variability in particulate concentrations due
38 to the large number and variety of sources (Jerrett et al. 2005; Karner et al. 2010; Eeftens et al.
39 2012); thus, a sparse monitoring network for $PM_{2.5}$ can lead to an incomplete understanding of its
40 spatial variability.

41 Recent advances in low-cost air quality sensing technologies have made it feasible for dense
42 networks of monitors to be deployed in urban areas, providing a neighborhood-resolution
43 understanding of air pollution (Snyder et al. 2013). Several pilot programs for monitoring air
44 quality at such high spatial resolution using these technologies have already begun (Jiao et al.
45 2016; English et al. 2017; Williams et al. 2018; Zimmerman et al. 2018). Efforts to characterize
46 the uncertainties associated with these low-cost instruments, particularly for long-duration field
47 deployment, are ongoing. Laboratory testing of several low-cost light-scattering particulate
48 sensors showed a linearity of results into the mg/m^3 range, but indicated that the sensors were
49 relatively less precise at concentrations below about 200 $\mu g/m^3$ (Wang et al. 2015). Evaluations of
50 low-cost Plantower PMS 1003 and 3003 units show good correlation (r above 0.88) with research-
51 grade instruments in laboratory and field conditions, although only a limited range of field
52 conditions were assessed (Kelly et al. 2017). Additional field testing of these sensors has shown a
53 significant effect of ambient humidity on their measurements (Jayaratne et al. 2018), better
54 performance at higher $PM_{2.5}$ concentrations, and varying correlation with different types of
55 reference instruments, e.g. r of 0.8 with a scattered light spectrometer versus 0.5 with a combined
56 light scattering nephelometer and beta attenuation monitoring instrument (Zheng et al. 2018).
57 Evaluation of the low-cost "Speck" monitors, using DSM501A optical dust sensors, indicated a
58 correlation below 0.7 with reference instruments, corresponding to a root-mean-square error of 10
59 $\mu g/m^3$ for outdoor measurements (Zikova et al. 2017a, 2017b). Investigations of other optical
60 particle counters (Alphasense OPC-N2) reinforce the need to correct their readings for relative
61 humidity, but indicate that inter-unit consistency is typically suitable to detect spatial trends if the
62 same type of sensors are used (Crilley et al. 2018). Assessments of these low-cost sensors must
63 also account for different use-cases; we consider two in this work. First, sensors may be used, e.g.
64 by community monitoring groups, to provide information on local air quality in real-time to
65 support individual health decisions. In this case, exact quantitative results are less important than
66 providing accurate indicators, e.g. that particulate concentrations are currently higher in one part
67 of a city than in another. Second, sensors may be used to determine long-term trends, e.g. for
68 quantifying the exposure of a population or the impacts of a new pollution-mitigation policy. In
69 this case, quantitatively accurate long-term performance is important, while short-term
70 performance is less so. Knowledge of the capabilities and limitations of these low-cost sensors
71 with respect to these use-cases is especially relevant considering that products such as the
72 PurpleAir sensor are already used by citizen scientists worldwide (www.purpleair.com).

73 In this paper, we provide evaluations of the long-term performance of two types of relatively low-
74 cost ($2000 or less) $PM_{2.5}$ sensors in field conditions in the city of Pittsburgh, Pennsylvania and

75   its surroundings. The ambient hourly $PM_{2.5}$ concentrations for this study are low (typically below
76   20 µg/m³) with respect to some previous field evaluations of these sensors (e.g. Kelly et al. 2017;
77   Jayaratne et al. 2018). We also suggest appropriate formulae for correcting for the influence of
78   humidity and temperature on instrument readings. We have focused our attention on field studies
79   due to the importance of assessing sensors in a similar environment to that in which they are to be
80   used (White et al. 2012; Piedrahita et al. 2014). By collocating several sensors with reference
81   instruments at two locations for two different time periods, we have developed a robust dataset
82   with which to calibrate correction models for the low-cost sensors, reflecting a wide range of
83   temperature (-20 to 43°C) and relative humidity (17 to 97%). Furthermore, by maintaining a small
84   number of sensors at these locations across multiple seasons (January 2017 to May 2018), we
85   evaluate their long-term performance and how this might be affected by different ambient
86   conditions.

2. Methods

2.1. RAMP Sensor Package and Attached Particulate Sensors

89   The Real-time Affordable Multi-Pollutant (RAMP) monitor (Figure 1) is a low-cost sensing
90   system collaboratively developed by SenSevere and the Center for Atmospheric Particle Studies
91   at Carnegie Mellon University. It incorporates five gas sensors, control circuits, batteries, and
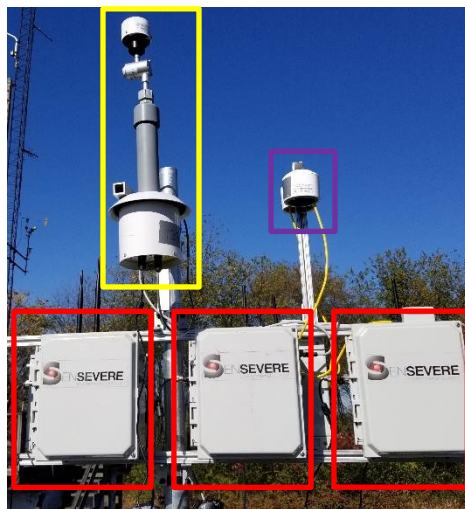92   wireless communication hardware.

94   Figure 1: Several RAMP monitors (red boxes) with connected Met-One NPM (yellow box) and
95   PurpleAir (purple box) $PM_{2.5}$ sensors.

96   In addition to its internal sensors, the RAMP can be connected to additional external instruments
97   for measuring $PM_{2.5}$. One such instrument is the Met-One Neighborhood Particulate Monitor
98   (NPM) sensor, which uses a forward light scattering laser. Previous research has assessed the
99   performance of two of these instruments over a two-month period in southern California, and
100   found only moderate correlations (R^2 between 0.5 and 0.7) between the instrument readings and

101 regulatory-grade instruments (AQ-SPEC 2015). The NPM is available for about $2000 or about
102 one tenth the price of regulatory-grade instruments measuring $PM_{2.5}$.

103 The PurpleAir $PM_{2.5}$ monitor (PPA) is also employed along with the RAMPs. This sensor
104 incorporates a pair of Plantower PMS 5003 laser particulate sensors, which provide measures of
105 $PM_{2.5}$ as well as of $PM_{1.0}$ and $PM_{10.0}$. Previous testing of three of these units over a two-month
106 period in southern California showed good correlation (R^2 above 0.9) with regulatory-grade
107 instruments (AQ-SPEC 2017). This sensor is available for about $250, or about one hundredth of
108 the price of a regulatory-grade instrument.

109 **2.2. Correction Methods**

110 Various methods were considered for correcting the raw readings of the low-cost $PM_{2.5}$ sensors
111 described above to better match regulatory-grade instruments. These methods applied various
112 combinations of functional forms, inputs, and thresholds. Two functional forms were considered:
113 linear and quadratic regression models. Possible inputs to these functions included the raw sensor
114 reading, hygroscopic growth factor fRH (as described below), the temperature T and/or relative
115 humidity RH recorded by the RAMP monitor to which the PM sensor was attached, and the
116 dewpoint DP (computed from T and RH). The hygroscopic growth factor corrects for particle
117 growth due to humidity, and is a nonlinear function of temperature and humidity. Furthermore,
118 humidity is known to affect particulate sensor performance (e.g. Jayaratne et al. 2018), and
119 temperature can affect the volatility of particulate constituents (e.g. Allen et al. 1997). These
120 factors, as well as the requirement that particulate mass be reported under specific temperature and
121 humidity conditions (US EPA 2016), prompted the inclusion of these inputs to proposed correction
122 models. Dewpoint was also considered as a non-linear function of temperature and humidity
123 related to condensation, and thus might serve as a proxy for the hygroscopic growth factor.

124 This hygroscopic growth factor is computed as:

$$\text{fRH}(T, RH) = 1 + \kappa_{\text{bulk}} \frac{a_w(T, RH)}{1 - a_w(T, RH)} \tag{1}$$

126 where:

$$a_w(T, RH) = RH \exp\left(\frac{4\sigma_w M_w}{\rho_w R T D_{\text{wet}}}\right)^{-1} \tag{2}$$

128 Parameters for this model are adapted from Petters and Kreidenweis (2007), as listed in Table 1.
129 Bulk particle composition factor κ_{bulk} was estimated from previous studies on the composition of
130 particulates in Pittsburgh (Cerully et al. 2015).

131 Table 1: Parameters used in hygroscopic growth factor calculation

Parameter	Value	Unit
κ_{bulk}	0.335	-
σ_w	0.072	N/m
M_w	0.018	kg/mol
ρ_w	1000	kg/m^3
R	8.314	J/mol K
D_{wet}	200	nm

132

133 Finally, thresholds were considered to define different subsets of the domain over which different
134 functional parameters could be applied, allowing for piecewise-linear or piecewise-quadratic
135 functions. Models without thresholds were considered, as well as models with single or multiple
136 threshold values chosen from among 5, 10, 15, 20, 30, 40, and 50 µg/m^3 (as determined from the
137 raw sensor reading). For reference, ambient concentrations in Pittsburgh typically range from 3 to
138 20 µg/m^3.

2.3. In-field Noise-adjustment Methods

140 We propose three methods to adjust for low-frequency noise (10^{-6} Hz or lower) in low-cost sensors
141 over the course of their field deployment. The first method, known as the "Deployment Records"
142 (DR) method, involves using a log of sensor deployment history to account for biases against a
143 reference instrument. In this case, the relative bias of a deployed sensor versus a "benchmark"
144 sensor is determined by computing the relative difference in readings from these sensors for the
145 last period during which they were collocated. The relative bias between this benchmark sensor
146 and a regulatory-grade instrument is also assessed based on its last collocation. Then, bias of the
147 deployed sensor to the regulatory-grade instrument is adjusted for, using the benchmark sensor as
148 an intermediate step. The second method, known as the "Site Percentiles" (SP) method, involves
149 computing the monthly 5$^{\text{th}}$ percentile of readings at a given deployment site, and then comparing
150 to the 5$^{\text{th}}$ percentile recorded at the nearest regulator monitoring station. Readings from the
151 deployed sensor are then adjusted so that these percentile values match. This is done with the
152 assumption that the 5$^{\text{th}}$ percentile represents a "background" level to which all sites in the region
153 are subject. A variation on this method, known as the "Average of Low readings" (AL) method,
154 uses the average of all readings in a month below 5 µg/m^3 as the target value to be matched. All
155 three methods rely on the availability of relatively frequent (e.g. hourly) data from regulatory-
156 grade instruments, and the first method relies on historical collocation data with these instruments.
157 The latter two methods of rectifying low-frequency noise by matching distribution parameters over
158 time are similar to those proposed by Moltchanov et al. (2015).

2.4. Data Collection

160 Sensor performance was assessed using data collected at two field sites, both coincident with
161 monitoring stations operated by the Allegheny County Health Department (ACHD), at which beta-

162   attenuation method (BAM) instruments provided hourly concentration measurements for
163   comparison (Hacker 2017). Although these instruments are not used for regulatory reporting, they
164   are recognized federal equivalent methods and provide hourly data necessary for short-term
165   comparisons.

166   One site, denoted as the "Lincoln" site, is located at 40.308ºN by 79.869ºW, is within 1 km of a
167   facility producing coke for steel manufacturing, and is nearby the only location in Allegheny
168   County which exceeded the annual EPA PM$_{2.5}$ standard in 2015-2017 (ACHD 2017). Average
169   PM$_{2.5}$ concentration at this site was 14.5 µg/m^3 in 2017, with a 1-hour maximum of 162 µg/m^3.
170   Here, one NPM sensor was operated for a total of 294 days from its deployment on April 24, 2017
171   until the end of data collection for this study on June 1, 2018. Additionally, for a period between
172   October 26, 2017 and February 12, 2018, a total of 12 NPM and 2 PPA sensors were collocated at
173   the site (although not all instruments were active for the entire period); during this time temperature
174   varied between -20 and 31°C and relative humidity varied from 22 to 97%. The second deployment
175   site, denoted as the "Lawrenceville" site, is located at 40.465ºN by 79.961ºW and is a community-
176   oriented monitoring site, part of the EPA's core monitoring network (Hacker 2017). Average PM$_{2.5}$
177   concentration at this site was 9.7 µg/m^3 in 2017, with a maximum 1-hour concentration of 67
178   µg/m^3. At this site, one NPM sensor was operated for a total of 380 days between January 13, 2017
179   and May 6, 2018. In addition, a total of 25 NPM and 9 PPA sensors were collocated at the site
180   between March 30, 2018 and June 4, 2018 (although again, not all instruments were present for
181   the entire period); temperature varied from -3 to 43°C and humidity varied between 17 and 97%.

182   Instruments at both sites were connected to RAMP monitors to allow for cellular data transmission.
183   For NPM sensors, data associated with instrument error codes, as well as likely erroneously high
184   readings (exceeding 10000 µg/m^3) were filtered from the data. For PPA sensors, readings from
185   both internal Plantower sensors were averaged to determine the PPA reading. Measurements from
186   these sensors were down-averaged to an hourly rate to allow for comparison with the reference
187   instruments.

188   **2.5. Assessment metrics**

189   To evaluate the performance of a sensor as compared to a reference (typically a regulatory-grade
190   instrument), the bias, mean absolute error, and correlation coefficient statistics are used. For n
191   measurements of concentration by the sensor (c) and reference ($\hat{c}$), bias is computed as:

192
$$\text{bias} = \frac{1}{n}\sum_{i=1}^{n}(c_i - \hat{c}_i) \tag{3}$$

193   mean absolute error (MAE) is evaluated as:

194
$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|c_i - \hat{c}_i| \tag{4}$$

195   and the Pearson correlation coefficient (r) is evaluated as:

$$196 \qquad r = \frac{\sum_{i=1}^{n}\left(c_i - \frac{1}{n}\sum_{j=1}^{n} c_j\right)\left(\hat{c}_i - \frac{1}{n}\sum_{j=1}^{n} \hat{c}_j\right)}{\sqrt{\sum_{i=1}^{n}\left(c_i - \frac{1}{n}\sum_{j=1}^{n} c_j\right)^2}\sqrt{\sum_{i=1}^{n}\left(\hat{c}_i - \frac{1}{n}\sum_{j=1}^{n} \hat{c}_j\right)^2}}, \qquad (5)$$

197 These statistics assess, respectively, the systematic differences between the sensor and reference
198 measurements over time, the average absolute difference in measurements taken at the same time,
199 and the degree to which the trends in measurements match (e.g. do measures from both sources
200 tend to be relatively "high" and "low" at the same times). Lower absolute values of bias and MAE
201 denote better agreement, while a value of r close to 1 denotes stronger correlation.

202 Performance of the instruments was also assessed from a classification perspective, using the
203 NAAQS 24-hour standard of 35 µg/m³ (www.epa.gov/criteria-air-pollutants/naaqs-table) as a
204 representative threshold, by assessing how often the sensor agreed with a reference instrument as
205 to whether this concentration was surpassed (note that this determination is made on an hourly
206 basis for this assessment, while the regulation cited above applies to daily averages). This is
207 quantified by the classification precision and recall, where classification precision indicates the
208 fraction of values of concentration c above threshold τ detected by the sensor which were also
209 detected by the reference:

$$210 \qquad \text{classification precision} = \frac{\sum_{i=1}^{n} \mathbb{I}(c_i > \tau)\mathbb{I}(\hat{c}_i > \tau)}{\sum_{i=1}^{n} \mathbb{I}(c_i > \tau)} \qquad (6)$$

211 and recall is the fraction of instances detected by the reference instrument which were also detected
212 by the sensor:

$$213 \qquad \text{classification recall} = \frac{\sum_{i=1}^{n} \mathbb{I}(c_i > \tau)\mathbb{I}(\hat{c}_i > \tau)}{\sum_{i=1}^{n} \mathbb{I}(\hat{c}_i > \tau)} \qquad (7)$$

214 where $\mathbb{I}$ is the indicator function, taking on value 1 when its argument is true and 0 otherwise.
215 Therefore, classification precision describes how often an event detected by the sensor actually
216 occurred (assuming the reference instrument reading is the "true" concentration) while recall
217 describes the fraction of actual events which were detected by the sensor. Values of classification
218 precision and recall close to 100% indicate better performance.

3. Results

3.1. Consistency between Sensors

221 To determine the consistency between sensors, pairwise comparisons of 1-hour-averaged data
222 were made among NPM and PPA sensors (i.e. NPM with NPM and PPA with PPA) collocated at
223 either site during the same period. Figure 2 presents the results of these inter-comparisons; only
224 results for sensors collocated for at least 3 days are presented. Overall, mutual correlations are
225 strong (typically $r > 0.9$) and are likely higher at the Lincoln site due to the wider range of
226 concentrations. Absolute differences in readings were typically below 2.5 µg/m³, while systematic
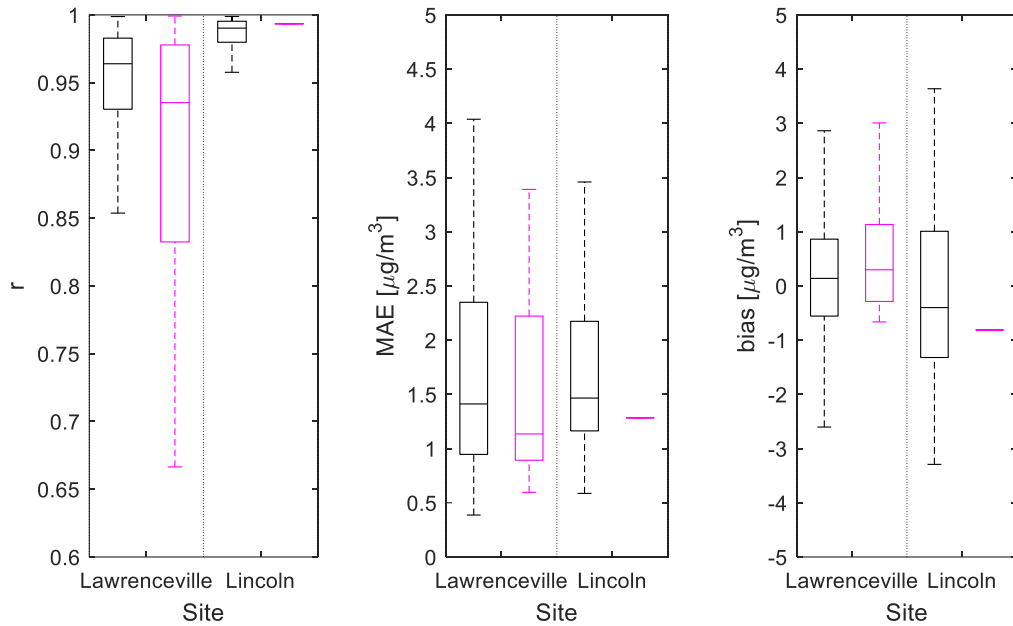227 biases between sensors were generally on the order of ±1 µg/m³.

7

Figure 2: Inter-comparison between sensors during collocation periods at both sites. Black boxplots indicate metric ranges for pairs of NPM sensors, and purple boxplots indicate ranges for pairs of PPA sensors. This represents 114 NPM pairs at Lawrenceville, 66 NPM pairs at Lincoln, 16 PPA pairs at Lawrenceville and 1 PPA pair at Lincoln.

Figure 3 compares hourly averages from collocated NPM and PPA sensors at Lawrenceville to each other as a function of humidity (the median readings of all sensors active at the site at the same time are shown). It is interesting to note that at low humidity, PPA readings are about twice that of the NPM, while at high humidity the ratio of readings approaches one; comparisons made between raw readings of both sensor types would therefore be heavily humidity-dependent.

238
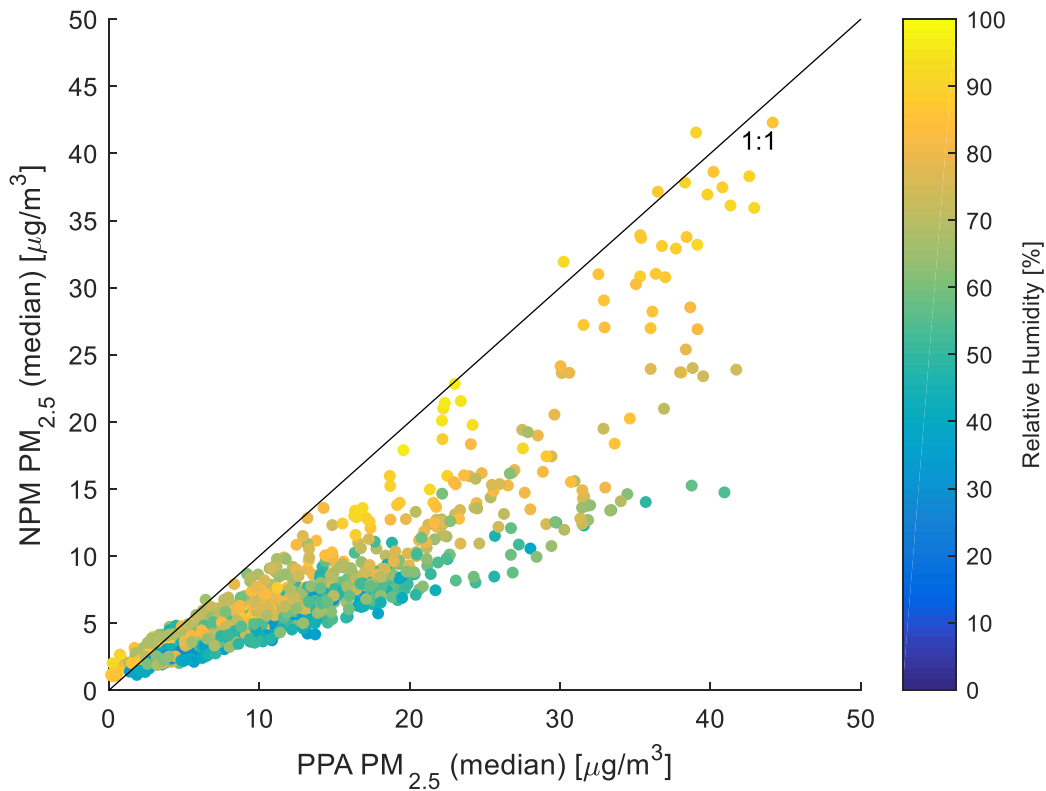
Figure 3: Comparison between medians of NPM and PPA sensors during collocation at the Lawrenceville site. Colors indicate relative humidity at the time of the measurements.

3.2. Accuracy of Low-Cost Sensors and Correction to BAM-Equivalence

Various functional forms as described in Section 2.2 were used to correct the hourly average readings of NPM and PPA sensors to match collocated data from the BAM regulatory-grade instruments. Models were calibrated using a combination of data collected at both the Lawrenceville and Lincoln sites from half of the sensors deployed to each site (the "training" set); model performance was evaluated on the other half of sensors at these sites (the "testing" set). Performance metrics for a subset of correction models on the testing set are presented in Table 2 (NPM) and Table 3 (PPA); the full set of performance metrics are included as supplementary information.

239

240
241

242

243
244
245
246
247
248
249
250

9

251   Table 2: Subset of performance metrics for NPM sensor correction models using one-hour
252 averages, as applied to a subset of n sensors set aside for testing. Results for the raw output of the
253   sensors are presented in the first line.

Met-One NPM Sensors

			Lawrenceville (n=13)			Lincoln (n=7)		
			r	MAE	bias	r	MAE	bias
Function	Inputs	Threshold		$[\mu g/m^3]$	$[\mu g/m^3]$		$[\mu g/m^3]$	$[\mu g/m^3]$
As-reported	NPM	None	0.60	4.47	-3.42	0.87	5.36	2.22
Linear	NPM	None	0.60	3.07	-1.76	0.87	4.62	0.89
Linear	NPM/fRH	None	0.73	2.96	-2.32	0.92	3.96	1.21
Linear	NPM/fRH, T, RH, DP	None	0.70	2.97	-0.95	0.92	4.05	1.37
Linear	NPM, T, RH	$30\,\mu g/m^3$	0.62	2.65	-0.82	0.91	3.99	0.91
Linear	NPM, T, RH, DP	$15\,\mu g/m^3$	0.68	2.69	-1.04	0.92	3.79	0.82
Linear	NPM/fRH, T, RH, DP	5, 10, 15, 20, 30, 40, $50\,\mu g/m^3$	0.70	2.86	-1.01	0.92	4.22	1.54
Quadratic	NPM	None	0.60	3.19	-1.99	0.88	4.60	0.97
Quadratic	NPM/fRH	None	0.73	2.75	-1.95	0.92	3.99	1.33
Quadratic	NPM/fRH, T, RH, DP	None	0.70	2.86	-1.16	0.92	4.02	1.33
Quadratic	NPM, T, RH	None	0.70	2.92	-1.28	0.92	3.88	0.99
Quadratic	NPM, T, RH, DP	$10\,\mu g/m^3$	0.69	2.73	-1.20	0.93	3.80	0.99

254

Table 3: Subset of performance metrics for PurpleAir sensor correction models using one-hour averages, as applied to a subset of n sensors set aside for testing. Results for the raw output of the sensors are presented in the first line.

PurpleAir PPA Sensors

			Lawrenceville (n=5)			Lincoln (n=1)		
			r	MAE	bias	r	MAE	bias
Function	Inputs	Threshold		$[\mu g/m^3]$	$[\mu g/m^3]$		$[\mu g/m^3]$	$[\mu g/m^3]$
As-reported	PPA	None	0.72	3.42	0.44	0.92	7.45	6.45
Linear	PPA	None	0.72	3.00	-2.19	0.92	4.05	-0.32
Linear	PPA/fRH	None	0.74	3.12	-1.34	0.91	4.34	-3.15
Linear	PPA/fRH, T, RH, DP	None	0.69	3.44	-1.27	0.91	4.02	-1.29
Linear	PPA, T, RH	20 $\mu g/m^3$	0.70	2.48	-0.28	0.95	3.62	-0.54
Linear	PPA, T, RH, DP	20 $\mu g/m^3$	0.72	2.41	-0.38	0.95	3.48	-0.51
Linear	PPA, T, RH, DP	5, 10, 15, 20, 30, 40, 50 $\mu g/m^3$	0.71	2.48	-0.44	0.94	3.56	-0.70
Quadratic	PPA	None	0.72	2.90	-2.03	0.92	4.10	-0.32
Quadratic	PPA/fRH	None	0.74	2.45	-0.84	0.89	4.45	-2.56
Quadratic	PPA/fRH, T, RH, DP	None	0.63	2.91	-0.62	0.93	3.83	-0.88
Quadratic	PPA, T, RH	30 $\mu g/m^3$	0.69	2.58	-0.45	0.95	3.29	-0.33
Quadratic	PPA, T, RH, DP	30 $\mu g/m^3$	0.69	2.61	-0.47	0.95	3.30	-0.38

The performance of each correction model as outlined above was scored using a heuristic combining various performance metrics across both collocation sites and penalizing the complexity of the model; see the supplementary materials for the resulting metrics. For selecting a final correction method for each type of sensor, performance across a range of concentrations experienced at both collocation sites was traded off against the complexity of the model (and therefore its propensity to overfit to training data).

Two equations were selected for the NPM sensors; first, a linear function of the raw signal corrected using a hygroscopic growth factor was identified as the model with the smallest number of free parameters giving the best overall performance:

$$[\text{corrected PM}_{2.5}] = \theta_1 \left(\frac{[\text{NPM PM}_{2.5}]}{\text{fRH}(T,RH)} \right) + \theta_0 \tag{8}$$

The hygroscopic growth factor is based on Pittsburgh-specific aerosol chemical composition, which may not be available at all locations. However, since factors such as temperature and relative humidity are readily available, a quadratic function of these which performed similarly well was considered as a more generalizable alternative:

11

273 $[\text{corrected PM}_{2.5}] = \alpha_0 + \alpha_1[\text{NPM PM}_{2.5}] + \alpha_2 T + \alpha_3 RH + \alpha_4[\text{NPM PM}_{2.5}]^2 +$

274 $\alpha_5[\text{NPM PM}_{2.5}]T + \alpha_6[\text{NPM PM}_{2.5}]RH + \alpha_7 T^2 + \alpha_8 TRH + \alpha_9 RH^2$ (9)

275 The final form selected for PPA sensors was a two-piece linear function of the sensor reading,

276 temperature, humidity, and dewpoint, with a threshold at 20 µg/m³:

277 $[\text{corrected PM}_{2.5}] = \begin{cases} \beta_0 + \beta_1[\text{PPA PM}_{2.5}] + \beta_2 T + \beta_3 RH + \beta_4 DP(T, RH) & \text{if } [\text{PPA PM}_{2.5}] > 20\,^{\mu g}/_{m^3} \\ \gamma_0 + \gamma_1[\text{PPA PM}_{2.5}] + \gamma_2 T + \gamma_3 RH + \gamma_4 DP(T, RH) & \text{if } [\text{PPA PM}_{2.5}] \leq 20\,^{\mu g}/_{m^3} \end{cases}$ (10)

278 Coefficients calibrated for these equations are listed in Table 4, along with their uncertainties.

279 Table 4: Coefficients for correction equations

Coefficient	Value Estimate	Standard Deviation	Unit
θ_0	1.52	0.16	$^{\mu g}/_{m^3}$
θ_1	1.94	0.020	N/A
α_0	0	2.9	$^{\mu g}/_{m^3}$
α_1	2.93	0.08	N/A
α_2	-0.11	0.08	$^{\mu g}/_{°Cm^3}$
α_3	0	0.08	$^{\mu g}/_{\%m^3}$
α_4	5.3×10^{-4}	1.5×10^{-4}	$^{m^3}/_{\mu g}$
α_5	-8.9×10^{-3}	1.2×10^{-3}	$°C^{-1}$
α_6	-2.7×10^{-2}	0.11×10^{-2}	$\%^{-1}$
α_7	2.9×10^{-3}	0.8×10^{-3}	$^{\mu g}/_{°C^2m^3}$
α_8	5.0×10^{-3}	1.0×10^{-3}	$^{\mu g}/_{°C\%m^3}$
α_9	0	6.0×10^{-4}	$^{\mu g}/_{\%^2m^3}$
β_0	75	11	$^{\mu g}/_{m^3}$
β_1	0.60	0.0090	N/A
β_2	-2.5	0.51	$^{\mu g}/_{°Cm^3}$
β_3	-0.82	0.11	$^{\mu g}/_{\%m^3}$
β_4	2.9	0.53	$^{\mu g}/_{°Cm^3}$
γ_0	21	2.1	$^{\mu g}/_{m^3}$
γ_1	0.43	0.013	N/A
γ_2	-0.58	0.090	$^{\mu g}/_{°Cm^3}$
γ_3	-0.22	0.023	$^{\mu g}/_{\%m^3}$
γ_4	0.73	0.098	$^{\mu g}/_{°Cm^3}$

280

281 Figure 4 plots hourly average readings from NPM and PPA sensors against regulatory-grade
282 instrument readings at the Lawrenceville site both before (above) and after (below) application of
283 correction equations. Before correction there is a clear effect of humidity on the readings; for the
284 NPM sensors particularly, many of the readings over 30 µg/m^3 correspond with periods of fog at
285 the site, indicating it may strongly affect the sensors. The corrections largely nullify this effect,
286 and reduce MAE by about 30% for both NPM and PPA sensors with respect to the raw signals.
287 However, there is still noticeable measurement noise about the identity line.

288

289
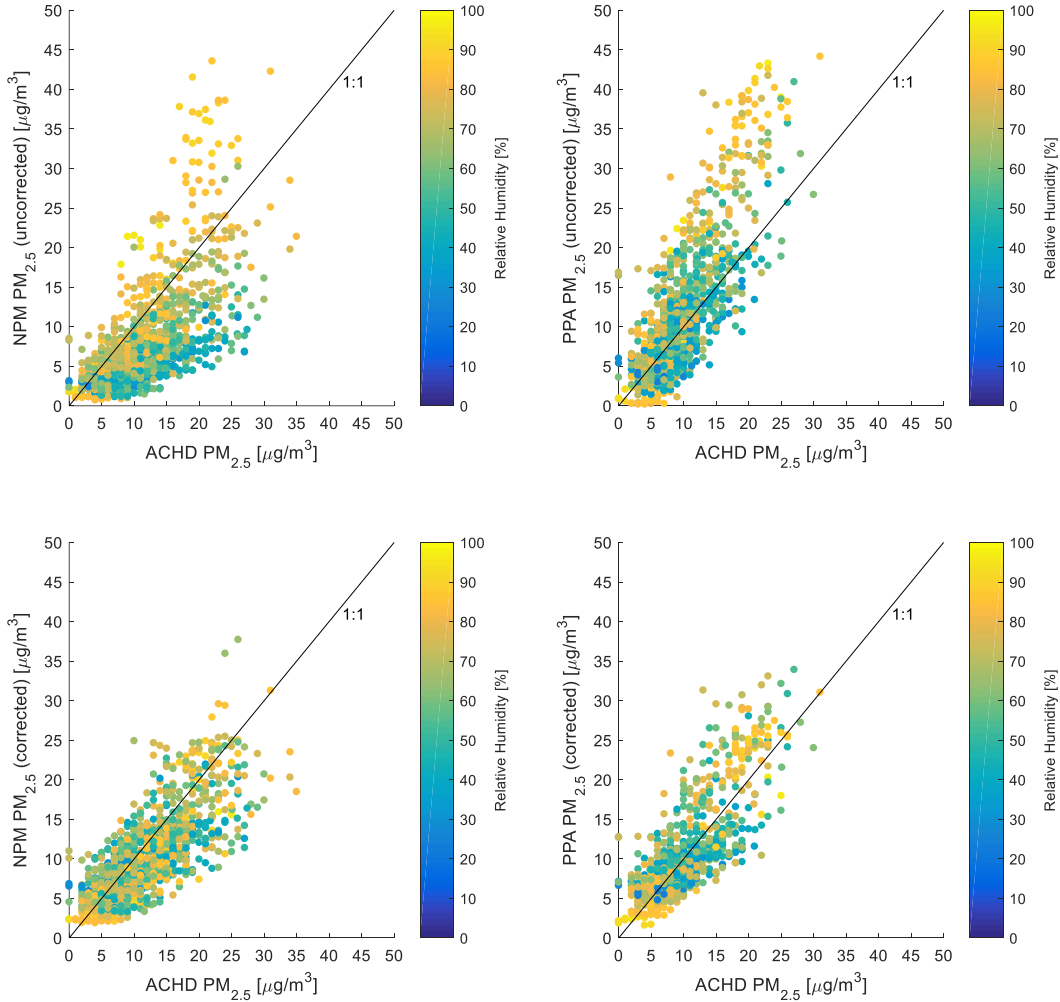
290 Figure 4: Comparison of median NPM (left) and PPA (right) sensor readings to the BAM
291 instrument during collocation at the Lawrenceville site, both before (above) and after (below)
292 correction (using Eq. 8 for NPM and Eq. 10 for PPA). Colors indicate relative humidity at the
293 time of the measurements.

294 Figure 5 summarizes the medians and ranges in performance of the corrected NPM and PPA hourly
295 averaged data across both collocation sites, using all sensors deployed to both sites (as opposed to
296 only the testing set), as well as specifying performance by different concentration ranges (0 to 10,
297 10 to 20, and higher than 20 µg/m^3). Correlation is typically better for NPM sensors (using either

13

correction equation), with r between 0.7 and 0.9, while for PPA sensors it ranges down to 0.5. Correlations also improve at higher concentrations. In terms of MAE, both sensors are between 3 and 5 $\mu g/m^3$. MAE also tends to increase as concentrations increase, but the PPA sensors appear to be less affected than NPM at concentrations above 20 $\mu g/m^3$; however, considering there were only two PPA sensors at the Lincoln site (where these higher concentrations were more common) this may be a sample size artefact. Although unbiased over the full range, the corrected sensor readings tend to be positively biased at low concentrations and negatively biased at high concentrations. This is opposite to the trend seen before correction and may be due to overcorrections at the extremes.
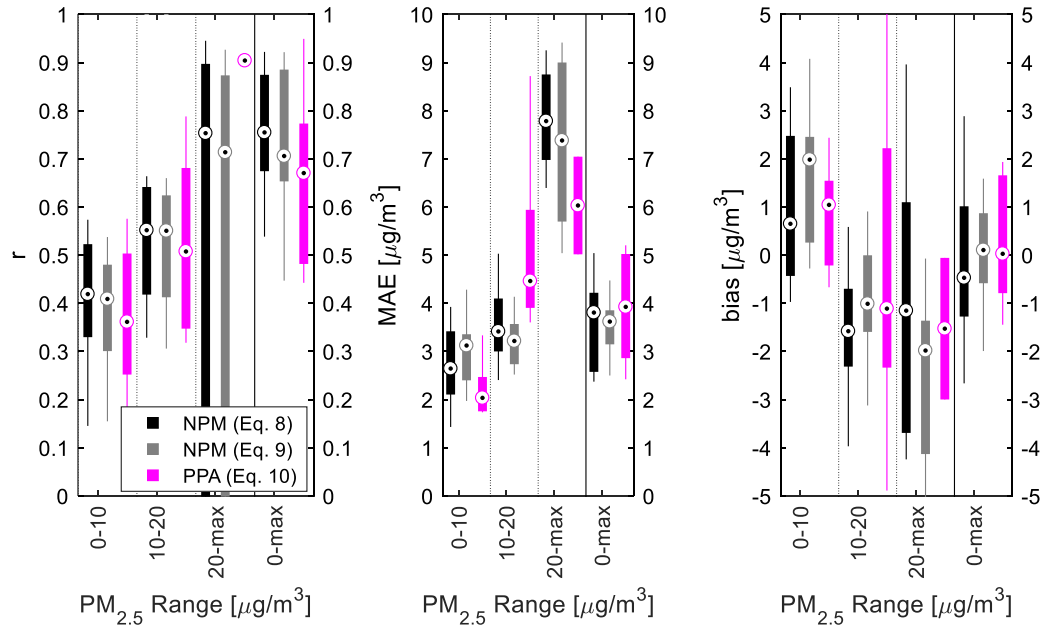
Figure 5: Comparison of sensor performance compared to BAM instruments during collocation at both the Lawrenceville and Lincoln sites. Performance metrics are plotted overall (0-max range) and by different PM$_{2.5}$ ranges (0-10, 10-20, 20-max).

Figure 6 assesses the ability of the sensors to correctly identify times when a threshold is passed; the timeline charts the number of hours per week with average concentrations above 35 $\mu g/m^3$ identified by the NPM sensor (corrected using Eq. 8) and/or the regulatory-grade instrument at the Lincoln site (results are not reported for the Lawrenceville site since hourly concentrations there surpassed the threshold less than 1% of the time). True positives occurred when both instruments detected an event; false positives are when only the NPM measured the event, and false negatives when the NPM failed to detect an event seen by the regulatory-grade instrument. A one hour "grace period" was used, i.e., if an event detection by one instrument leads or trails the other by up to an hour, this was still counted as a true positive. The classification precision of the sensor was 85% and its recall was 97%; for comparison, these values are 73% and 97% respectively when the un-

14

corrected signal of the NPM is used. Of the misclassifications, 41% were within 5 µg/m^3 of the threshold; the rest represented larger discrepancies between the instruments.
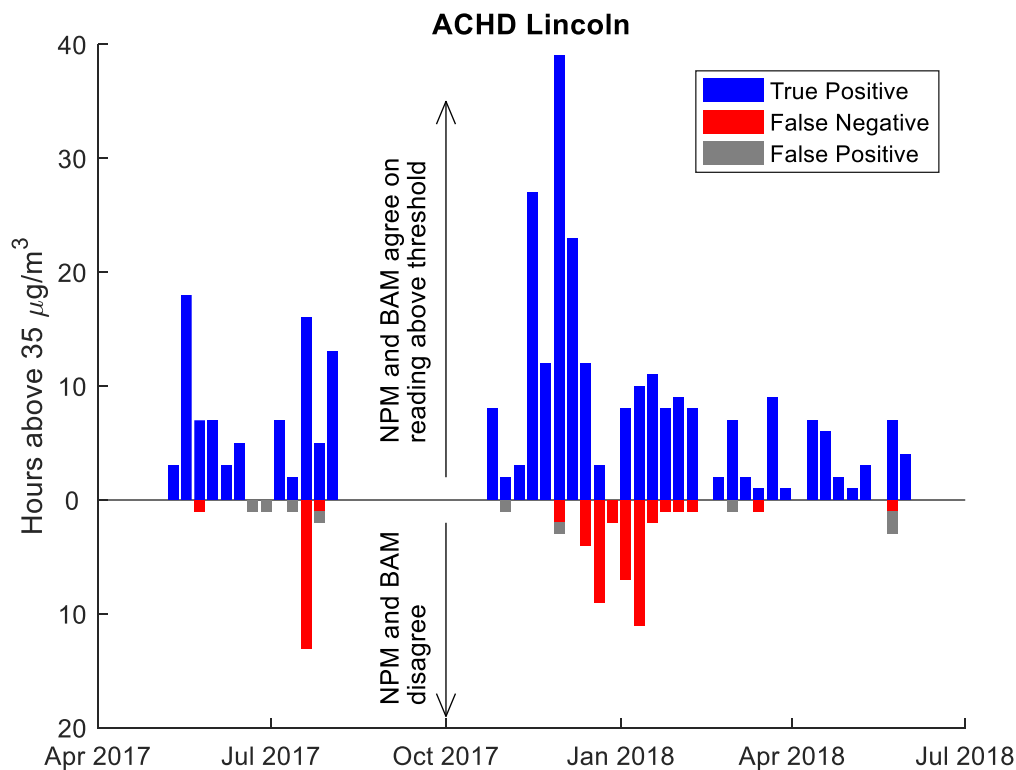

Figure 6: Detection of hourly high PM$_{2.5}$ events by NPM sensor at Lincoln. True positives (correct detections) are counted for each hour on a weekly basis, along with false positives (false event indications) and false negatives (missed events).

3.3. Long-Term Performance

Long-term assessment is necessary to categorize bias and assess data quality after extensive field use of sensors; previous studies of optical particle counters operating for up to four months have seen no evidence of significant drift (Crilley et al. 2018). The long-term performance of NPM sensors was assessed using data collected by the two sensors deployed at the Lawrenceville and Lincoln sites for extended periods (e.g. more than a year of data at the former site collected over a 16-month span). First, these data were used to assess the in-field noise-adjustment methods proposed in Section 2.3. Figure 7 shows the spread in weekly biases (difference between the weekly average readings of the corrected sensors and the regulatory-grade instruments) for both sites, both without noise-adjustment and with the three proposed noise-adjustment methods. Based on these results, the "average of low readings" method is best, reducing the median bias at the Lawrenceville site by half. However, there is still a significant spread in the weekly bias, indicating

15

339　that sensors may be experiencing changes in their low-frequency noise on relatively short
340　timescales.

341
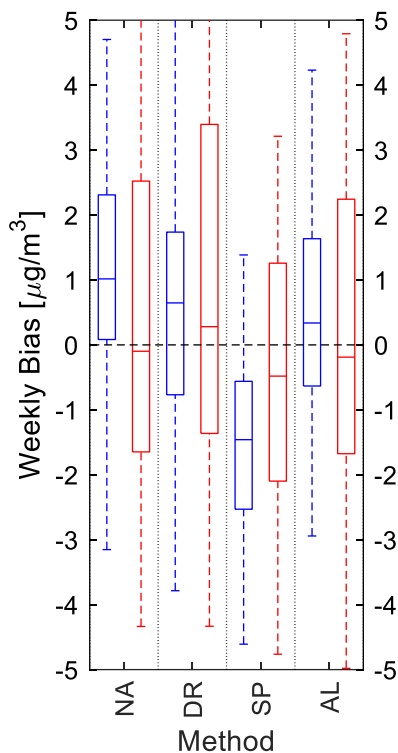
Figure 7: Performance of various noise-adjustment methods (NA – no noise-adjustment applied;
DR – noise-adjusted using deployment records; SP – noise-adjusted using percentiles of nearest
reference site; AL – noise-adjusted using averages of low readings at nearest reference site) in
reducing weekly biases. Performance is determined separately for the Lawrenceville (blue) and
Lincoln (red) sites. Corrections are performed using Eq. 8.

347　Figure 8 plots the distribution of absolute errors between the corrected and noise-adjusted sensor
348　data and the associated regulatory-grade instrument as a function of the period over which readings
349　are averaged. Solid lines indicate the mean absolute errors for these averaging periods, while the
350　shaded area indicates the interquartile range of absolute errors for different periods. While for
351　hourly averages, errors are on the order of 4 μg/m^3, for weekly averages this is reduced to about 2
352　μg/m^3, and for annual averages errors are below 1 μg/m^3.
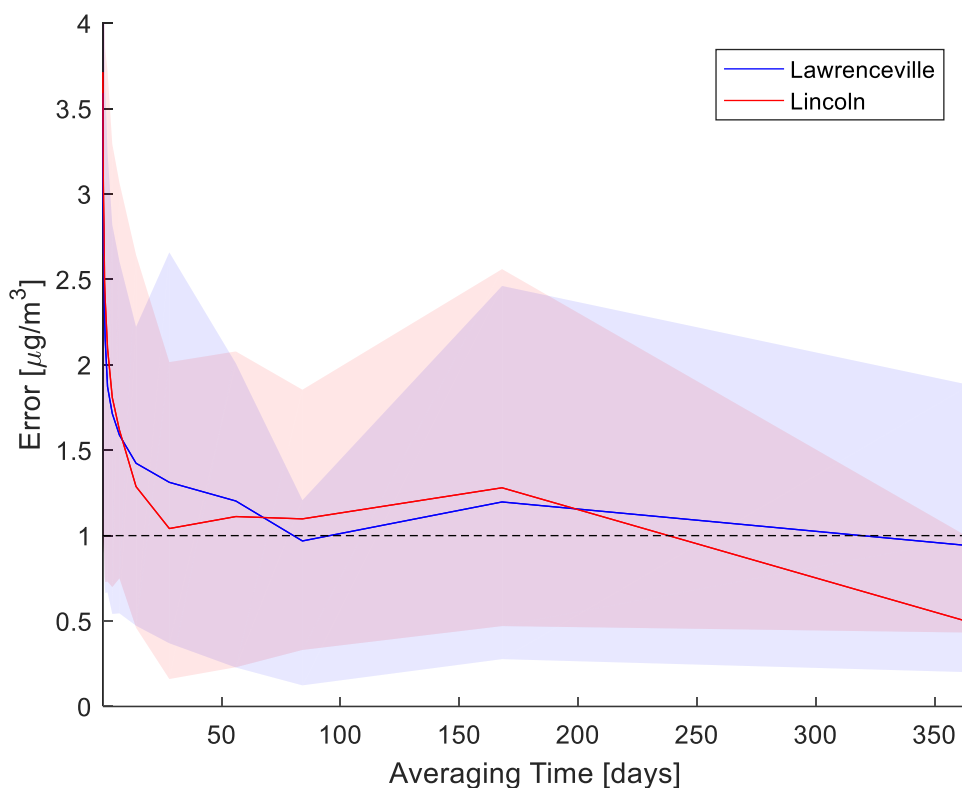
Figure 8: Mean absolute error in PM$_{2.5}$ measurements for two NPM sensors during long-term deployments as a function of averaging period. Measurements are corrected using Eq. 9 and noise-adjusted using the "average of low readings" method. Shaded regions indicate 25%-75% ranges in errors over periods.

4. Discussion

Testing of a relatively large number of NPM (up to 25 sensors at the Lawrenceville site) and PPA (up to 12 sensors at the Lawrenceville site) low-cost PM$_{2.5}$ sensors showed high mutual consistency between the sensors, with MAE typically below 2.5 µg/m^3 and correlation typically higher than 0.9. Systematic biases between instruments appeared to account for the largest fraction of the absolute differences; such biases may be assessed before and after field deployment using collocations, but these methods will likely fail because the biases vary over time, as evidenced by the relatively poor performance of the "deployment records" noise-adjustment method for field-collected data, which used this strategy. Furthermore, use of uncorrected sensor measurements is not advised, due to the major effect of humidity on the readings of both sensors (see Figure 3, Figure 4).

Correction equations were selected to balance accuracy and simplicity, with the selected equations capable of being implemented for real-time monitoring applications. For the NPM sensors, reasonably good performance was achieved with either a linear function incorporating a

17

hygroscopic growth correction term or a quadratic function involving temperature and humidity, while for PPA sensors a piecewise linear function of temperature, humidity, and dewpoint was selected. However, for both types of sensors, even following correction, relatively large differences in hourly averages (MAE of 4 $\mu g/m^3$, and double that for high concentrations) were observed with respect to the BAM regulatory-grade instruments. This lack of consistency with BAM instruments has also been observed in other works (e.g. Zheng et al. 2018) and may not be reconcilable with low-cost optical sensors. However, as data are averaged over longer periods, accuracy can be improved, especially if in-field noise-adjustment methods are also applied, such that long-term (1 year or more) averages are likely to be accurate within 1 $\mu g/m^3$. Furthermore, during tests for detecting hourly concentrations higher than 35 $\mu g/m^3$, the NPM sensor was able to correctly identify these events to within one hour more than 80% of the time; this indicates the potential for this sensor to be used to identify pollution hotspots.

In terms of use cases, the high level of mutual consistency and ability (with suitable noise-adjustment) to provide accurate long-term averages makes these low-cost sensors useful for large-scale mapping campaigns to determine long-term spatial patterns and temporal trends in PM$_{2.5}$. For real-time monitoring, although these sensors have the ability to detect hourly "spikes" with reasonable accuracy, concentration values should only be considered to be within about ± 5 $\mu g/m^3$ in typical ambient concentrations (with a wider margin for higher readings). Nevertheless, this is sufficient to provide qualitative indications of relative short-term air quality. The small size and ease of deployment of these units make them well suited to urban monitoring. PPA sensors also incorporate a pair of particulate sensors, allowing for internal self-consistency checks to flag possible erroneous data, while NPM sensors include PM$_{2.5}$ cyclones and inlet heaters which can protect the units from excessive dust and humidity (to which PPA sensors, which lack these features, may be more susceptible during longer deployments). Finally, we note that while these results are determined for the specific environment of Pittsburgh, Pennsylvania, we believe they will generalize to other areas of North America, Europe, and other cities which are characterized by hourly PM$_{2.5}$ mass concentrations typically less than 20 $\mu g/m^3$ over a wide temperature and humidity range.

Acknowledgements

References

ACHD. 2017. Air Quality Annual Data Summary for 2017: Criteria Pollutants and Selected Other Pollutants.

408 Allen G, Sioutas C, Koutrakis P, Reiss R, Lurmann FW, Roberts PT. 1997. Evaluation of the
409   TEOM® Method for Measurement of Ambient Particulate Mass in Urban Areas. Journal
410   of the Air & Waste Management Association 47:682–689;
411   doi:10.1080/10473289.1997.10463923.

412 Apte JS, Marshall JD, Cohen AJ, Brauer M. 2015. Addressing Global Mortality from Ambient
413   PM2.5. Environmental Science & Technology 49:8057–8066;
414   doi:10.1021/acs.est.5b01236.

415 AQ-SPEC. 2015. Met One Neighborhood Monitor Evaluation Report.

416 AQ-SPEC. 2017. PurpleAir PA-II Sensor Evaluation Report.

417 Brook RD, Rajagopalan S, Pope CA, Brook JR, Bhatnagar A, Diez-Roux AV, et al. 2010.
418   Particulate Matter Air Pollution and Cardiovascular Disease: An Update to the Scientific
419   Statement From the American Heart Association. Circulation 121:2331–2378;
420   doi:10.1161/CIR.0b013e3181dbece1.

421 Cerully KM, Bougiatioti A, Hite JR, Guo H, Xu L, Ng NL, et al. 2015. On the link between
422   hygroscopicity, volatility, and oxidation state of ambient and water-soluble aerosols in the
423   southeastern United States. Atmospheric Chemistry and Physics 15:8679–8694;
424   doi:10.5194/acp-15-8679-2015.

425 Crilley LR, Shaw M, Pound R, Kramer LJ, Price R, Young S, et al. 2018. Evaluation of a low-cost
426   optical particle counter (Alphasense OPC-N2) for ambient air monitoring. Atmospheric
427   Measurement Techniques 11:709–720; doi:10.5194/amt-11-709-2018.

428 Eeftens M, Tsai M-Y, Ampe C, Anwander B, Beelen R, Bellander T, et al. 2012. Spatial variation
429   of PM2.5, PM10, PM2.5 absorbance and PMcoarse concentrations between and within 20
430   European study areas and the relationship with NO2 – Results of the ESCAPE project.
431   Atmospheric Environment 62:303–317; doi:10.1016/j.atmosenv.2012.08.038.

432 English PB, Olmedo L, Bejarano E, Lugo H, Murillo E, Seto E, et al. 2017. The Imperial County
433   Community Air Monitoring Network: A Model for Community-based Environmental
434   Monitoring for Public Health Action. Environmental Health Perspectives 125;
435   doi:10.1289/EHP1772.

436 Hacker K. 2017. Air Monitoring Network Plan for 2018.

437 Jayaratne R, Liu X, Thai P, Dunbabin M, Morawska L. 2018. The Influence of Humidity on the
438   Performance of Low-Cost Air Particle Mass Sensors and the Effect of Atmospheric Fog.
439   Atmospheric Measurement Techniques Discussions 1–15; doi:10.5194/amt-2018-100.

440 Jerrett M, Burnett RT, Ma R, Pope CA, Krewski D, Newbold KB, et al. 2005. Spatial Analysis of
441   Air Pollution and Mortality in Los Angeles. Epidemiology 16:727–736;
442   doi:10.1097/01.ede.0000181630.15826.7d.

443 Jiao W, Hagler G, Williams R, Sharpe R, Brown R, Garver D, et al. 2016. Community Air Sensor
444   Network (CAIRSENSE) project: evaluation of low-cost sensor performance in a suburban
445   environment in the southeastern United States. Atmospheric Measurement Techniques
446   9:5281–5292; doi:10.5194/amt-9-5281-2016.

447 Karner AA, Eisinger DS, Niemeier DA. 2010. Near-Roadway Air Quality: Synthesizing the
448   Findings from Real-World Data. Environmental Science & Technology 44:5334–5344;
449   doi:10.1021/es100008x.

450 Kelly KE, Whitaker J, Petty A, Widmer C, Dybwad A, Sleeth D, et al. 2017. Ambient and
451   laboratory evaluation of a low-cost particulate matter sensor. Environmental Pollution
452   221:491–500; doi:10.1016/j.envpol.2016.12.039.

453 Lepeule J, Laden F, Dockery D, Schwartz J. 2012. Chronic Exposure to Fine Particles and
454   Mortality: An Extended Follow-up of the Harvard Six Cities Study from 1974 to 2009.
455   Environmental Health Perspectives 120:965–970; doi:10.1289/ehp.1104660.

456 Moltchanov S, Levy I, Etzion Y, Lerner U, Broday DM, Fishbain B. 2015. On the feasibility of
457   measuring urban air pollution by wireless distributed sensor networks. Science of The Total
458   Environment 502:537–547; doi:10.1016/j.scitotenv.2014.09.059.

459 Petters MD, Kreidenweis SM. 2007. A single parameter representation of hygroscopic growth and
460   cloud condensation nucleus activity. Atmospheric Chemistry and Physics 7:1961–1971;
461   doi:10.5194/acp-7-1961-2007.

462 Piedrahita R, Xiang Y, Masson N, Ortega J, Collier A, Jiang Y, et al. 2014. The next generation
463   of low-cost personal air quality sensors for quantitative exposure monitoring. Atmospheric
464   Measurement Techniques 7:3325–3336; doi:10.5194/amt-7-3325-2014.

465 Pope CA, Burnett RT, Thun MJ, Calle EE, Krewski D, Ito K, et al. 2002. Lung cancer,
466   cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. JAMA
467   287: 1132–1141.

468 Schwartz J, Dockery DW, Neas LM. 1996. Is daily mortality associated specifically with fine
469   particles? J Air Waste Manag Assoc 46: 927–939.

470 Snyder EG, Watkins TH, Solomon PA, Thoma ED, Williams RW, Hagler GSW, et al. 2013. The
471   Changing Paradigm of Air Pollution Monitoring. Environmental Science & Technology
472   47:11369–11377; doi:10.1021/es4022602.

473 US EPA. 2016. Quality Assurance Guidance Document 2.12: Monitoring PM2.5 in Ambient Air
474   Using Designated Reference or Class I Equivalent Methods.

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

478   White RM, Paprotny I, Doering F, Cascio WE, Solomon PA, Gundel LA. 2012. Sensors and
479   "apps" for community-based: Atmospheric monitoring. EM: Air and Waste Management
480   Association's Magazine for Environmental Managers 2012: 36–40.

481   Williams R, Vallano D, Polidori A, Garvey S. 2018. Spatial and Temporal Trends of Air Pollutants
482   in the South Coast Basin Using Low Cost Sensors.

483   Zheng T, Bergin MH, Johnson KK, Tripathi SN, Shirodkar S, Landis MS, et al. 2018. Field
484   evaluation of low-cost particulate matter sensors in high and low concentration
485   environments. Atmospheric Measurement Techniques Discussions 1–40;
486   doi:10.5194/amt-2018-111.

487   Zikova N, Hopke PK, Ferro AR. 2017a. Evaluation of new low-cost particle monitors for PM2.5
488   concentrations measurements. Journal of Aerosol Science 105:24–34;
489   doi:10.1016/j.jaerosci.2016.11.010.

490   Zikova N, Masiol M, Chalupa D, Rich D, Ferro A, Hopke P. 2017b. Estimating Hourly
491   Concentrations of PM2.5 across a Metropolitan Area Using Low-Cost Particle Monitors.
492   Sensors 17:1922; doi:10.3390/s17081922.

493   Zimmerman N, Presto AA, Kumar SPN, Gu J, Hauryliuk A, Robinson ES, et al. 2018. A machine
494   learning calibration model using random forests to improve sensor performance for lower-
495   cost air quality monitoring. Atmospheric Measurement Techniques 11:291–313;
496   doi:10.5194/amt-11-291-2018.