

Fine Particle Mass Monitoring with Low-Cost Sensors: Corrections and Long-Term Performance Evaluation

Carl Malings^{1,2}, Rebecca Tanzer¹, Aliaksei Hauryliuk¹, Provat K. Saha¹, Allen L. Robinson¹,
Albert A. Presto¹, R Subramanian^{1,2}

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

²OSU-EFLUVE, CNRS, Université Paris-Est Créteil, 61 Avenue du Général de Gaulle, 94000
Créteil, France

Abstract

Low-cost fine particle mass (PM_{2.5}) sensors enable dense networks to increase the spatial resolution of air quality monitoring. However, these sensors are affected by environmental factors such as temperature and humidity and their effects on ambient aerosol, which must be accounted for to improve the in-field accuracy of these sensors. We conducted long-term tests of two low-cost PM_{2.5} sensors: Met-One NPM and PurpleAir PA-II units. We found a high level of self-consistency within each sensor type after testing 25 NPM and 9 PurpleAir units (and after rejecting several malfunctioning PurpleAir units). We developed two types of corrections for the low-cost sensor measurements to better match regulatory-grade data. The first correction accounts for aerosol hygroscopic growth using particle composition and corrects for particle mass below the optical sensor size cut-point by collocation with reference Beta Attenuation Monitors (BAM). A second, fully-empirical correction uses linear or quadratic functions of environmental variables based on the same collocation dataset. Either model yielded comparable improvements over raw measurements. Sensor performance was assessed for two use cases: improving community awareness of air quality with short-term qualitative comparisons of sites and providing long-term quantitative information for health impact studies. For the short-term case, either sensor can provide reasonably accurate concentration information (mean absolute error of ~4 µg/m³) in near-real time. For the long-term case, tested using year-long collocations at one urban background and one near-source site, error in the annual average was reduced below 1 µg/m³. These sensors are thus suitable for supplementing regulatory-grade instruments in sparsely monitored regions, neighborhood-scale monitoring, and for better understanding spatial patterns and temporal air quality trends across urban areas.

1. Introduction

The negative health impacts of exposure to particulate matter smaller than 2.5 micrometers (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). Reductions in PM_{2.5}, even in low concentration

36 environments, can have substantial benefits (Apte et al. 2015). Accurate monitoring of PM_{2.5} is
37 thus important for a variety of applications, including long-term health studies, assessing the
38 impacts of technology and/or regulatory changes on emissions, and supporting decision-making
39 for future regulatory efforts or to alter individual behavior in real-time. Monitoring is especially
40 of interest in urban areas where the high density of exposed populations is coupled with higher
41 variability in particulate concentrations due to the large number and variety of sources (Jerrett et
42 al. 2005; Karner et al. 2010; Eeftens et al. 2012). Thus, a sparse monitoring network can lead to
43 an incomplete understanding of PM_{2.5} spatial variability and its subsequent health impacts. Recent
44 advances in low-cost air quality sensing technologies have made it feasible for dense networks of
45 monitors to be deployed in urban areas, providing a neighborhood-scale understanding of air
46 pollution (Snyder et al. 2013). Several pilot programs for monitoring air quality at such high spatial
47 resolution using these technologies are underway (Jiao et al. 2016; English et al. 2017; Williams
48 et al. 2018; Zimmerman et al. 2018).

49 Most low-cost particulate mass sensors make use of optical measurement techniques (Wang et al.
50 2015; Kelly et al. 2017; Rai et al. 2017). It is well-known that these optical methods do not
51 generally agree with measurements obtained from instruments operating on different principles
52 (Watson et al. 1998; Wilson et al. 2002; Chow et al. 2008; Solomon and Sioutas 2008; Burkart et
53 al. 2010). For example, work with low-cost optical PM_{2.5} sensors (Plantower model PMS3003)
54 showed good correlation (r of 0.8) with a scattered light spectrometer versus low correlation (r of
55 0.5) with a beta attenuation monitoring (BAM) instrument (Zheng et al. 2018). There are several
56 reasons for these disagreements. First, for regulatory-grade instruments, particulate mass must be
57 reported under specific temperature (20-23°C) and humidity (30-40%) conditions (US EPA
58 2016b), while most low-cost sensors report data at ambient conditions, leading to discrepancies
59 with regulatory-grade instruments (including the BAM instruments used in this work, which are
60 recognized as federal equivalent methods for PM_{2.5} mass measurement). As ambient humidity
61 increases, hygroscopic growth of particles occurs, which increases their light scattering coefficient
62 (Cabada et al. 2004), and therefore the mass reported by optical sensors. Field testing of low-cost
63 optical PM_{2.5} sensors has shown the significant effect of ambient humidity on their measurements
64 (Jayaratne et al. 2018; Zikova et al. 2017a, 2017b). A challenge is that hygroscopic growth is
65 particle composition dependent (Petters and Kreidenweis 2007). Accounting for such growth is
66 needed to reduce these humidity effects when comparing the optical sensor to reference monitors.
67 Further, low-cost optical sensors are usually limited to measuring particles larger than 0.3
68 micrometers (Koehler and Peters 2015; Zhou and Zheng 2016), and so will underreport PM_{2.5}.
69 This is corrected for during factory calibration by adjusting the instrument output to match that of
70 a reference PM_{2.5} mass measurement of the same calibration “smoke” (Liu et al. 2017). Differences
71 between particle size distribution and composition used for the factory calibration and the ambient
72 aerosol during deployment can therefore cause further errors.

73 Assessments of these low-cost sensors must also account for different use cases (Rai et al. 2017),
74 for which different performance goals might be appropriate (e.g. Williams et al. 2014). We

75 consider two use cases in this work. First, sensors may be used, e.g. by community monitoring
76 groups, to provide information on local air quality in real-time to support individual decisions, for
77 example about where to go for a walk in a city to avoid highly polluted areas. In this case, exact
78 quantitative results are less important than providing accurate indicators, e.g. that PM
79 concentrations are currently higher in one part of a city than in another. Second, sensors may be
80 used to determine long-term trends, e.g. for quantifying the exposure of a population or the impacts
81 of a new pollution-mitigation policy. In this case, quantitatively accurate long-term performance
82 is important. Knowledge of the capabilities and limitations of these low-cost sensors with respect
83 to these use cases is especially relevant considering that products such as the PurpleAir sensor are
84 already used by citizen scientists worldwide (www.purpleair.com).

85 In this paper, we provide evaluations of the long-term performance of two types of relatively low-
86 cost (under \$2000 for the NPM and \$250 for the PurpleAir) PM_{2.5} sensors in field conditions in
87 the city of Pittsburgh, Pennsylvania and its surroundings. The ambient hourly PM_{2.5} concentrations
88 for this study are low (typically below 20 µg/m³) compared to previous field evaluations of these
89 sensors (e.g. Kelly et al. 2017; Jayaratne et al. 2018). We also propose and evaluate both physics-
90 based and fully-empirical methods to correct for the influence of humidity and temperature on
91 sensor readings, thereby making them more comparable to BAM instrument data. We have focused
92 our attention on field studies due to the importance of assessing sensors in a similar environment
93 to that in which they are to be used (White et al. 2012; Piedrahita et al. 2014). In Pittsburgh, like
94 in other urban areas, PM_{2.5} is composed of regionally transported (aged) aerosol and fresh
95 vehicular emissions (Tan et al. 2014). Additionally, a metallurgical coke producing facility is a
96 major local point source. Hence, we develop a calibration equation through collocation with a
97 reference monitor at an urban background site that represents aged background PM and a source-
98 oriented site near the major point source. We further evaluate these models across multiple seasons
99 (January 2017 to May 2018) at both locations, as well as at a roadside location where vehicular
100 contribution to PM_{2.5} below the sensor size cut-point should be highest, and a more rural location.

101 **2. Methods**

102 **2.1. RAMP Sensor Package and Attached PM_{2.5} Sensors**

103 The Real-time Affordable Multi-Pollutant (RAMP) monitor is a low-cost sensing system
104 collaboratively developed by SenSevere and the Center for Atmospheric Particle Studies at
105 Carnegie Mellon University (Zimmerman et al. 2018). It incorporates five gas sensors, electronics,
106 batteries, and wireless communication hardware. In addition to its internal sensors, the RAMP can
107 be connected to external instruments for measuring PM_{2.5}. One such instrument is the Met-One
108 Neighborhood Particulate Monitor (NPM) sensor, which uses a forward light scattering laser. The
109 unit is also equipped with an inlet heater and PM_{2.5} cyclone. Previous research has assessed the
110 performance of two of these instruments over a two-month period in southern California, and
111 found moderate correlations (r between 0.7 and 0.8) with regulatory-grade instruments (AQ-SPEC

112 2015). The NPM is available for about \$2000 or about one tenth the price of regulatory-grade
113 instruments measuring PM_{2.5}. A total of 50 NPM units have been deployed alongside RAMPs.

114 The PurpleAir PM_{2.5} monitor (PPA) is also deployed along with the RAMPs. This sensor
115 incorporates a pair of Plantower PMS 5003 laser sensors, which provide measures of PM_{2.5} as well
116 as of PM_{1.0} and PM₁₀. Previous testing of three of these units over a two-month period in southern
117 California showed good correlation (r above 0.9) with regulatory-grade instruments (AQ-SPEC
118 2017). This sensor is available for about \$250, or about one hundredth of the price of a regulatory-
119 grade instrument. Initial laboratory testing of a batch of 30 PurpleAir units found 7 to be defective;
120 these defects were identified due to low correlations (r < 0.7) between the data provided by each
121 units' pair of Plantower sensors. These defective sensors are not considered in this paper. A total
122 of 20 PurpleAir units have been deployed with RAMPs in the Pittsburgh area.

123 **2.2. Data Collection**

124 Sensor performance was assessed using data collected at four field sites - one corresponding to an
125 “urban background”, one impacted by industrial emissions, one by vehicle emissions, and one
126 more rural site - coincident with monitoring stations operated by the Allegheny County Health
127 Department (ACHD) or Pennsylvania Department of Environmental Protection (DEP), at which
128 BAM instruments provided hourly concentration measurements for comparison (Hacker 2017;
129 McDonnell 2017). Although these instruments are not used for regulatory reporting, they are
130 recognized federal equivalent methods and provide hourly data for Air Quality Index calculations.
131 This section describes the two sites used for correction method development and long-term testing.
132 Two additional regulatory sites which were used to test the correction methods are described in
133 Section 3.3.

134 The “Lincoln” site (AQS#42-003-7004, 40.308°N by 79.869°W) is a “source-dominated” site
135 within 1 km of a facility producing coke for steel manufacturing that is the largest primary PM_{2.5}
136 point source in Allegheny county. This part of Allegheny County exceeded the annual and 24-hour
137 Environmental Protection Agency (EPA) PM_{2.5} standards over 2015-2017 (ACHD 2017). This site
138 is illustrative of a “fence line” monitoring application, where monitors are placed in proximity to
139 a known emission source. Average PM_{2.5} concentration at this site (based on the BAM) was 14.5
140 µg/m³ in 2017, with a one-hour maximum of 162 µg/m³. Here, one NPM sensor was operated for
141 a total of 294 days from April 24, 2017 until the end of data collection for this study on June 1,
142 2018. Additionally, between October 26, 2017 and February 12, 2018 (109 days), a total of 12
143 NPM and 2 PurpleAir sensors were collocated at the site (although not all instruments were active
144 for the entire period).

145 The “Lawrenceville” deployment site (AQS#42-003-0008, 40.465°N by 79.961°W) is an urban
146 background site located in an urban residential and commercial neighborhood, and part of the
147 EPA’s NCore monitoring network (Hacker 2017). Average PM_{2.5} concentration at this site (based
148 on the BAM) was 9.7 µg/m³ in 2017, with a maximum one-hour concentration of 67 µg/m³. At

149 this site, one NPM sensor was operated for a total of 380 days between January 13, 2017 and May
150 6, 2018. In addition, a total of 25 NPM and 9 PurpleAir sensors were collocated at the site between
151 March 30, 2018 and June 4, 2018 (66 days, although again, not all instruments were present for
152 the entire period). Five NPM sensors were collocated at both Lincoln and Lawrenceville at
153 different times; none of the PurpleAir sensors were collocated at both sites.

154 Instruments at all sites were connected to RAMP monitors to allow for cellular data transmission.
155 For NPM sensors, data associated with instrument error codes, as well as likely erroneously high
156 readings (exceeding $10,000 \mu\text{g}/\text{m}^3$) were removed from the dataset. For PurpleAir sensors,
157 readings from both internal Plantower sensors were averaged to determine the PurpleAir reading.
158 Measurements from both types of sensors were down-averaged from their collection rate (roughly
159 one measurement every 12 seconds) to an hourly rate to allow for comparison with the reference
160 instruments.

161 2.3. Physics-based (Hygroscopic Growth and Size Distribution) Correction Methods

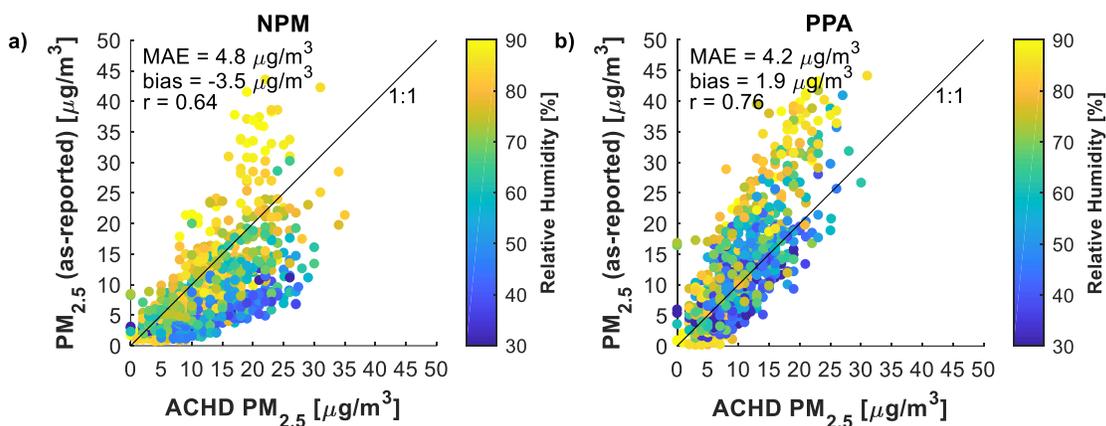
162 Figure 1 compares the as-reported data from the NPM and PurpleAir sensors to the BAM
163 instrument at the Lawrenceville site. There are sizeable discrepancies (up to $20 \mu\text{g}/\text{m}^3$ in some
164 cases) in the values, with humidity clearly having an effect. A method was sought to correct the
165 readings of the low-cost sensors to better match those of the federal equivalent BAM instruments.
166 As a starting point, the hygroscopic growth factor is the ratio of particulate mass at a given
167 humidity and temperature to that at 22°C and 35% relative humidity (the conditions at which
168 regulatory data are reported), and is calculated as follows:

$$169 \quad fRH(T, RH) = 1 + \kappa_{\text{bulk}} \frac{a_w(T, RH)}{1 - a_w(T, RH)} \quad (1)$$

170 The hygroscopicity of bulk aerosol (κ_{bulk}) is evaluated considering seasonal changes in particle
171 composition observed in Pittsburgh; these are accounted for by dividing the year into summer
172 (May to September inclusive), winter (November to March inclusive), and other periods (with the
173 “other” period using an average of the summer and winter compositions). Within each period, it is
174 assumed that the aerosol composition and size distribution are constant over time and throughout
175 the urban area. Seasonal aerosol compositions in the Pittsburgh area are obtained from Gu et al.
176 (2018), and literature κ -values for the major non-refractory aerosol components sulfate, nitrate,
177 ammonium, and organic matter are used (Cerully et al. 2015; Petters and Kreidenweis 2007); a
178 sensitivity analysis for this compositional information is provided in the results (Section 3.2) and
179 in the supplemental materials (Section S.2). Water activity is calculated as:

$$180 \quad a_w(T, RH) = RH \exp\left(\frac{4\sigma_w M_w}{\rho_w R T D_p}\right)^{-1} \quad (2)$$

181 where σ_w , M_w , and ρ_w represent the surface tension, molecular weight and density of water,
182 respectively; T is the absolute temperature, R is the ideal gas constant, RH is ambient relative
183 humidity; and D_p is the particle diameter (see Table S.1 for details).



184

185 Figure 1: Comparison of one-hour-average NPM (a) and PurpleAir (b) as-reported sensor
 186 readings to the BAM instrument during collocation at the Lawrenceville site. Each point
 187 indicates the median across all sensors of the given type present at the site. Colors indicate
 188 relative humidity at the time of the measurements. A breakdown of these results by relative
 189 humidity is provided in the supplemental materials (Table S.2).

190 Correction of low-cost sensor readings using the hygroscopic growth factor alone was found to be
 191 insufficient (see supplemental materials Figure S.9), likely due to differences between the factory
 192 calibration aerosols and ambient aerosol in Pittsburgh. Therefore, the hygroscopic growth
 193 correction was combined with an additional linear correction:

194

$$[\text{corrected PM}_{2.5}] = \theta_1 \left(\frac{[\text{PM}_{2.5} \text{ as reported}]}{f_{RH}(T, RH)} \right) + \theta_0 \quad (3)$$

195 The coefficients θ_0 and θ_1 were estimated using a combination of data collected at both the urban
 196 background Lawrenceville and source-dominated Lincoln sites from half of the sensors deployed
 197 to each site (the “training” set). Correction model performance was evaluated on the other half of
 198 sensors at these sites (the “testing” set), as well as at independent sites (see Section 3.3).
 199 Coefficients were set using typical linear regression techniques, minimizing the error between the
 200 corrected sensor measurements and the collocated BAM instrument at each site. These coefficients
 201 were estimated separately for the different time periods (summer, winter, other) for each of the
 202 low-cost sensor types (NPM, PurpleAir). This was necessary to account for the different responses
 203 of each type of sensor. For example, seasonal changes in particle size distributions lead to changes
 204 in the θ_0 term as more or less of the particulate matter mass falls below the 300nm detection size
 205 cut-point for optical sensors.

206 2.4. Empirical Correction Methods

207 The hygroscopic growth factor correction method described above is based on information about
 208 the specific aerosol chemical composition of the sensor deployment area, which may not be
 209 available at all locations. However, since factors such as temperature and relative humidity are

210 more readily available, other more generalizable, empirical correction equations were developed
 211 using these data. Dewpoint (DP) was considered as a factor related to condensation that might
 212 serve in place of the hygroscopic growth factor; temperature (T) and relative humidity (RH) were
 213 also considered. Various combinations of the as-reported sensor readings and the above
 214 environmental parameters were fit using linear and quadratic regression models to correct the data.
 215 The forms of the empirical corrections were selected by trading off performance (across a range
 216 of concentrations experienced at both collocation sites) against functional complexity (details are
 217 provided in the supplemental materials, Section S.3). For NPM sensors, a quadratic function of the
 218 sensor reading, temperature, and humidity was selected:

$$219 \quad [\text{corrected PM}_{2.5}]_{\text{NPM}} = \alpha_0 + \alpha_1[\text{PM}_{2.5}]_{\text{NPM}} + \alpha_2 T + \alpha_3 RH + \alpha_4 [\text{PM}_{2.5}]_{\text{NPM}}^2 + \\ 220 \quad \alpha_5 [\text{PM}_{2.5}]_{\text{NPM}} T + \alpha_6 [\text{PM}_{2.5}]_{\text{NPM}} RH + \alpha_7 T^2 + \alpha_8 TRH + \alpha_9 RH^2 \quad (4)$$

221 The form selected for PurpleAir sensors was a two-piece linear function of the sensor reading,
 222 temperature, humidity, and dewpoint, with a threshold at $20 \mu\text{g}/\text{m}^3$:

$$223 \quad [\text{corrected PM}_{2.5}]_{\text{PPA}} = \begin{cases} \beta_0 + \beta_1[\text{PM}_{2.5}]_{\text{PPA}} + \beta_2 T + \beta_3 RH + \beta_4 DP(T, RH) & \text{if } [\text{PM}_{2.5}]_{\text{PPA}} > 20 \mu\text{g}/\text{m}^3 \\ \gamma_0 + \gamma_1[\text{PM}_{2.5}]_{\text{PPA}} + \gamma_2 T + \gamma_3 RH + \gamma_4 DP(T, RH) & \text{if } [\text{PM}_{2.5}]_{\text{PPA}} \leq 20 \mu\text{g}/\text{m}^3 \end{cases} \quad (5)$$

224 Coefficients calibrated for these equations (using standard regression techniques) along with their
 225 uncertainties are provided in the supplemental materials (Table S.4).

226 **2.5. In-field Drift-adjustment**

227 A somewhat random, not-necessarily-monotonic fluctuation (e.g. a “random walk”) taking place
 228 over a period of weeks or months was observed in field-deployed NPM sensors when Eq. (4) is
 229 applied (see supplemental materials Figure S.5). The reason for this is likely due to seasonal
 230 changes in aerosol properties and/or sensor behaviors which are not captured by this equation. This
 231 was observed to affect monthly average $\text{PM}_{2.5}$ readings by up to $4 \mu\text{g}/\text{m}^3$ at the Lawrenceville and
 232 Lincoln sites. Insufficient data were available to assess whether the same phenomenon occurs for
 233 PurpleAir sensors. We propose three methods to adjust for this drift in sensor response over the
 234 course of their field deployment. Note that here we use “drift” to refer to any changes in the
 235 baseline or “zero” reading of the sensor.

236 The first adjustment method, known as the “Deployment Records” (DR) method, involves using a
 237 log of sensor deployment history to account for biases against a reference instrument. This method
 238 involves adjusting the measurements of all sensors to match that of one “benchmark” sensor during
 239 periods when sensors are collocated. The benchmark sensor is then collocated with a regulatory-
 240 grade instrument while other sensors are deployed in the field. The relative bias of a deployed
 241 sensor versus the regulatory-grade instrument can then be estimated using the benchmark as an
 242 intermediary (i.e. the biases of all sensors versus the benchmark are assessed during their
 243 collocations, and the bias of the benchmark versus the regulatory-grade instrument is assessed
 244 during its collocation; the bias of any deployed sensor versus the regulatory-grade instrument is

245 then estimated as the sum of the above biases). The second method, known as the “Fifth
 246 Percentiles” (5P) method, involves computing the monthly 5th percentile of readings at a given
 247 deployment site, and then comparing to the 5th percentile recorded at the nearest regulatory
 248 monitoring station. Readings from the deployed sensor are then adjusted so that these percentiles
 249 match. This is done with the assumption that the 5th percentile represents a “background” level to
 250 which all sites in the region are subject. The third method is a variation of the 5P method, known
 251 as the “Average of Low readings” (AL) method, which uses the average of all readings in a month
 252 below 5 µg/m³ as the target value to be matched. All three methods rely on the availability of
 253 relatively frequent (e.g. hourly) data from regulatory-grade instruments, and the first method relies
 254 on historical collocation data with these instruments. Diagrams depicting each of these proposed
 255 methods are provided in the supplemental materials (Figure S.6). The latter two methods of
 256 rectifying drift by matching distribution parameters over time are similar to those proposed by
 257 Tsujita et al. (2005) and used by Moltchanov et al. (2015).

258 2.6. Assessment metrics

259 To evaluate the performance of a sensor as compared to a reference (typically a regulatory-grade
 260 instrument), the bias, mean absolute error, and correlation coefficient (r) statistics were used
 261 (details are provided in the supplemental materials, Section S.5). Performance of the instruments
 262 was also assessed from a classification perspective, using the EPA’s National Ambient Air Quality
 263 Standards 24-hour standard of 35 µg/m³ (www.epa.gov/criteria-air-pollutants/naaqs-table) as a
 264 representative threshold, by assessing how often the sensor agreed with a reference instrument as
 265 to whether this concentration was surpassed. This determination was made on an hourly basis for
 266 this assessment, while the regulation cited above applies to daily averages. This comparison was
 267 therefore conservative, and we would expect better performance for daily averages based on the
 268 results of Section 3.6. Classification precision indicates the fraction of values of concentration c
 269 above threshold τ detected by the sensor which were also detected by the reference:

$$270 \quad \text{classification precision} = \frac{\sum_{i=1}^n \mathbb{I}(c_i > \tau) \mathbb{I}(\hat{c}_i > \tau)}{\sum_{i=1}^n \mathbb{I}(c_i > \tau)} \cdot 100\% \quad (6)$$

271 where c_i is the reading of the sensor and \hat{c}_i the reading of the reference instrument at time i of n ,
 272 and \mathbb{I} is the indicator function, taking on value 1 when its argument is true and 0 otherwise.
 273 Classification recall is the fraction of instances detected by the reference instrument which were
 274 also detected by the sensor:

$$275 \quad \text{classification recall} = \frac{\sum_{i=1}^n \mathbb{I}(c_i > \tau) \mathbb{I}(\hat{c}_i > \tau)}{\sum_{i=1}^n \mathbb{I}(\hat{c}_i > \tau)} \cdot 100\% \quad (7)$$

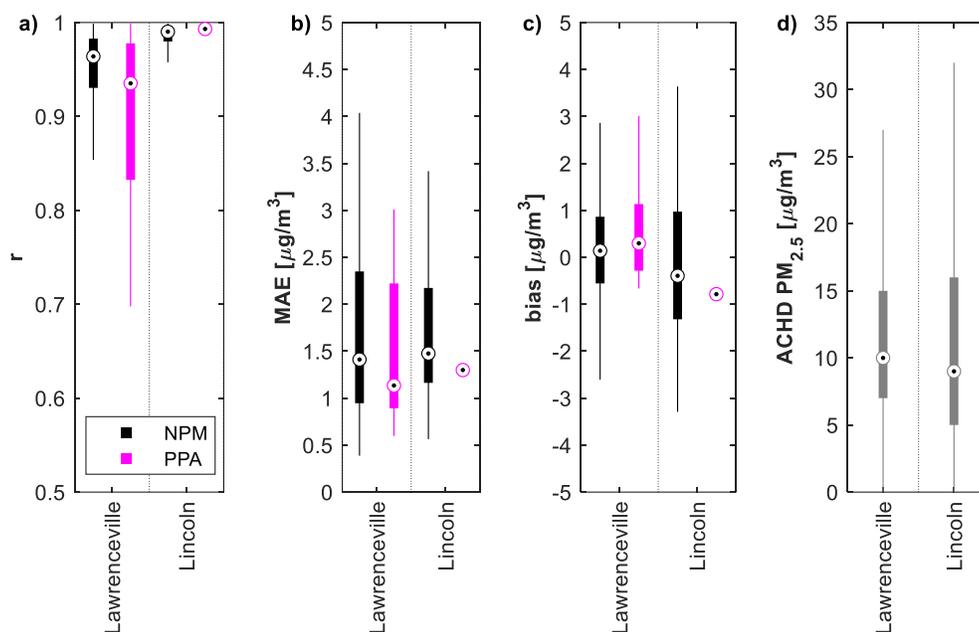
276 Therefore, classification precision describes how often an event detected by the sensor actually
 277 occurred (assuming the reference instrument reading was the “true” concentration) while recall
 278 describes the fraction of actual events which were detected by the sensor. Values of these metrics
 279 close to 100% indicate better performance.

280 **3. Results**

281 In this section, first, the mutual consistency of the as-reported data from the low-cost PM sensors
282 is quantified, to address how comparisons might be made without applying corrections. Second,
283 the quantitative performance of the proposed correction methods is assessed for the short-term use
284 case envisioned for these sensors. Finally, the long-term performance of these sensors is analyzed,
285 including contributions of the proposed drift-adjustment methods.

286 **3.1. Consistency between Sensors**

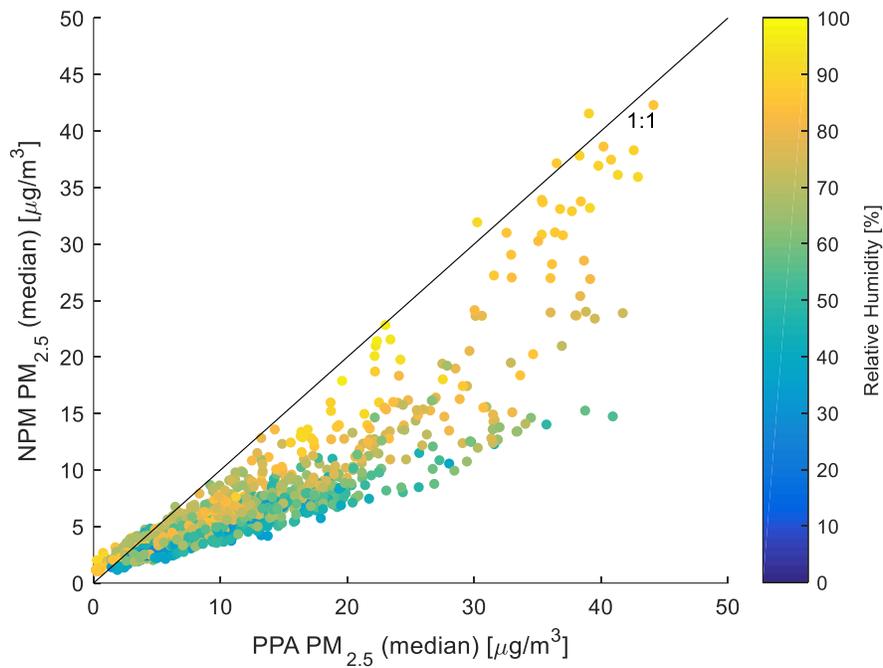
287 To determine the consistency between sensors, pairwise comparisons of 1-hour-averaged data
288 were made among NPM and PurpleAir sensors (i.e. NPM with NPM and PurpleAir with
289 PurpleAir) collocated at either the Lawrenceville or Lincoln site during the same period. At
290 Lawrenceville, during the RAMP collocations, temperature varied between -20 and +31°C and
291 relative humidity varied from 22% to 97%; at Lincoln, temperature varied from -3 to +43°C and
292 humidity varied between 17% and 97% (as measured by the RAMPs' onboard sensors). Figure 2
293 presents the results of these inter-comparisons; only results for sensors collocated for at least 3
294 days (36 1-hour averages) are presented. Overall, mutual correlations were strong (typically $r >$
295 0.9) and were higher at the Lincoln site likely due to the wider range of concentrations. Absolute
296 differences in as-reported readings were typically about $2 \mu\text{g}/\text{m}^3$ or less, which includes systematic
297 biases between sensors generally on the order of $\pm 1 \mu\text{g}/\text{m}^3$. This is similar to prior results for
298 Alphasense OPC-N2 optical $\text{PM}_{2.5}$ sensors, which are more than twice the price of PurpleAir units
299 (Crilley et al. 2018).



300

301 Figure 2: Inter-comparison of as-reported one-hour-average data between sensors during
 302 collocation periods at both sites. In the boxplots, circles with dots denote the median, thick bars
 303 denote the interquartile range, and thin bars denote the 95% confidence range. Black boxplots
 304 indicate metric ranges for pairs of NPM sensors, and purple boxplots indicate ranges for pairs of
 305 PurpleAir sensors. This represents 114 NPM pairs at Lawrenceville, 66 NPM pairs at Lincoln,
 306 16 PurpleAir pairs at Lawrenceville and 1 PurpleAir pair at Lincoln. For reference, the ranges of
 307 concentrations measured by BAM instruments at the sites during the same time are depicted in
 308 panel d.

309 Figure 3 compares hourly averages of as-reported data from NPM sensors at Lawrenceville to
 310 those collected by PurpleAir sensors at Lawrenceville as a function of humidity (the median
 311 readings of all sensors active at the site at the same time are shown). At low humidity, PurpleAir
 312 readings were about twice that of the NPM, while at high humidity the ratio of readings approached
 313 one; comparisons made between raw readings of the two sensor types would therefore be heavily
 314 humidity-dependent. There are several likely causes for these differences. First, the NPM
 315 possesses an inlet heater with a 4-second residence time which activates when relative humidity
 316 exceeds 40%. However, this residence time may not be sufficient to totally remove humidity
 317 effects on the NPM. Second, these instruments are calibrated differently. The NPM is calibrated
 318 with $0.6\mu\text{m}$ polystyrene latex spheres (Met One 2018), while PurpleAir Plantowers are calibrated
 319 with ambient aerosol across several cities in China (Wang 2019). They therefore respond
 320 differently when exposed to a common aerosol which differs from their calibration aerosols.

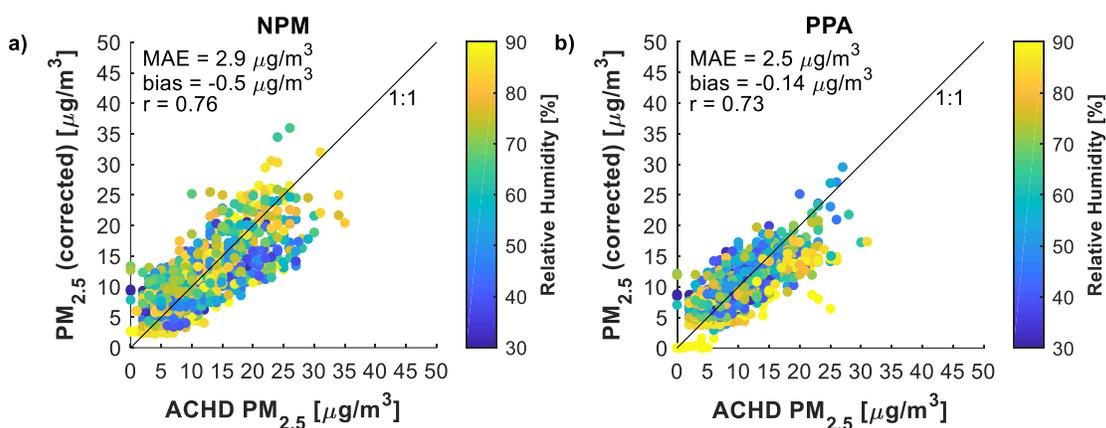


321

322 Figure 3: Comparison between medians of as-reported one-hour-average data of 25 NPM and 9
 323 PurpleAir sensors during collocation at the Lawrenceville site. Colors indicate relative humidity
 324 at the time of the measurements.

325 **3.2. Correction of Low-Cost Sensors towards a Federal Equivalent Method**

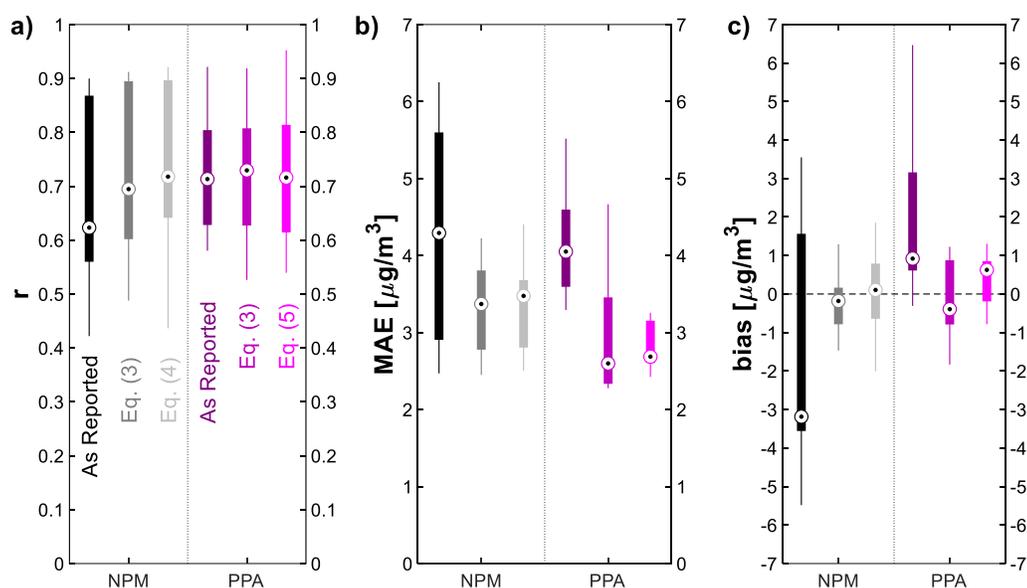
326 Figure 4 plots median hourly-average readings from NPM and PurpleAir sensors collocated at the
 327 Lawrenceville site corrected using Eq. (3) against the ACHD regulatory-grade (BAM) instrument
 328 readings. This correction decreased MAE by about 40% for both NPM and PurpleAir sensors with
 329 respect to their as-reported values and reduced bias significantly, but there was still noticeable
 330 measurement noise ($r \sim 0.75$) about the identity line.



331

332 Figure 4: Comparison of one-hour-average NPM (a) and PurpleAir (b) sensor readings to the
 333 BAM instrument during collocation at the Lawrenceville site after correction using Eq. (3), with
 334 appropriate coefficients for NPM and PurpleAir. Each point indicates the median across all
 335 sensors of the given type present at the site (including both “training” and “testing” sensors).
 336 Colors indicate relative humidity at the time of the measurements. A breakdown of these results
 337 by relative humidity is provided in the supplemental materials (Table S.3).

338 Figure 5 assesses the performance of the designated “testing” set of low-cost sensors deployed to
 339 the Lawrenceville and Lincoln sites during the March to June (at Lawrenceville) and October to
 340 February (at Lincoln) collocation periods. The figure compares as-reported data to data corrected
 341 using the hygroscopic-growth-based approach of Eq. (3) (with appropriate coefficients for NPM
 342 or PurpleAir sensors) and data corrected using the fully-empirical approaches of Eq. (4) for NPM
 343 or Eq. (5) for PurpleAir. In all cases hourly-averaged data were used. In terms of correlation
 344 (Figure 5a), no improvement was made for PurpleAir sensors, while only a modest improvement
 345 resulted from correction of the NPM sensors. In terms of MAE (Figure 5b) and bias (Figure 5c),
 346 however, both correction approaches resulted in noticeable improvements. For NPM sensors, both
 347 the physics-based Eq. (3) and fully-empirical Eq. (4) gave comparable performance. For PurpleAir
 348 sensors, the fully-empirical approach of Eq. (5) provided a smaller spread of MAE and bias results
 349 as compared to Eq. (3), while the median MAE of both approaches were almost the same, and the
 350 median bias of Eq. (5) was slightly worse. Overall both correction approaches improved upon the
 351 as-reported data and there was no significant difference between their performance.



352

353 Figure 5: Performance metrics of one-hour-average as-reported and corrected sensor data
 354 compared to BAM instruments during collocation at both the Lawrenceville and Lincoln sites.

355 Results shown relate to a total of 17 NPM and 5 PurpleAir sensors of the “testing” set.
 356 Corrections are performed using either the approach of Eq. (3), with appropriate coefficients for
 357 NPM or PurpleAir, or the approaches of Eq. (4) for NPM and Eq. (5) for PurpleAir.

358 Table 1 presents the calibrated coefficients for the approach of Eq. (3) for both NPM and PurpleAir
 359 sensors during the summer, winter, and for other periods (calibrated coefficients for Eqs. (4) and
 360 (5) are provided in the supplemental materials, Table S.4). Note that for both NPM and PurpleAir
 361 sensors, the value of θ_0 (the linear intercept term) was larger in summer than in winter. This could
 362 be explained by the fact that during summertime in Pittsburgh, as in most urban areas (Asmi et al.
 363 2011), particles smaller than 300 nm optical diameter are a larger fraction of $\text{PM}_{2.5}$ (see the
 364 supplemental materials Figure S.8), necessitating a larger correction. For θ_1 (the linear slope term),
 365 while the values for summer and winter were the same for NPM sensors, for PurpleAir sensors the
 366 value was higher in the winter. However, the hygroscopic growth factor (for the same temperature
 367 and relative humidity) was also higher in winter, as winter-time aerosol has a larger contribution
 368 from more hygroscopic inorganic aerosol. Thus, the net result was a lower impact of seasonal
 369 changes in the hygroscopic growth factor on the PurpleAir readings, indicating that the PurpleAir
 370 sensor may be less susceptible to humidity-driven changes. The internal structure of the PurpleAir
 371 unit may contribute to this; the plastic shell enclosing the Plantower sensors and associated
 372 electronic circuits can trap heat inside the unit, leading to lower relative humidity within the device.
 373 During tests at the Lawrenceville site, RH inside the PurpleAir was found to be 9.7 percentage
 374 points lower on average than outside, while T was 2.7°C higher.

375 Table 1: Calibrated coefficients for Eq. (3). Values following “±” represent the standard
 376 deviations in the coefficient estimates.

		Met-One NPM		PurpleAir PPA	
θ_0	Summer	5.28 ± 0.09	$\mu\text{g}/\text{m}^3$	5.4 ± 0.4	$\mu\text{g}/\text{m}^3$
	Winter	2.03 ± 0.08	$\mu\text{g}/\text{m}^3$	-0.3 ± 0.2	$\mu\text{g}/\text{m}^3$
	Other	1.68 ± 0.13	$\mu\text{g}/\text{m}^3$	3.7 ± 0.1	$\mu\text{g}/\text{m}^3$
θ_1	Summer	1.50 ± 0.01		0.62 ± 0.03	
	Winter	1.50 ± 0.01		1.25 ± 0.01	
	Other	1.76 ± 0.02		0.83 ± 0.01	

377

378 3.3. Performance Assessment at Other Regulatory Sites

379 To further assess the performance of these sensor corrections at locations independent of where
 380 they were developed, several sensors were tested at two additional sites. The “Parkway East” site
 381 (AQS#42-003-1376, 40.437°N by 79.864°W) represents a roadside location (Hacker 2017), and
 382 thus may have a different, vehicular traffic-influenced particle composition and size distribution
 383 than either the urban background or coke oven-impacted sites at which the corrections were
 384 developed. Between September 6 and 27, 2018 (21 days), two PurpleAir sensors were collocated
 385 at this site. Data from these sensors were corrected using Eq. (3). These provided comparable
 386 results to testing at the Lincoln and Lawrenceville sites (median r of 0.71, median MAE of 2.7
 387 $\mu\text{g}/\text{m}^3$, median bias of 0.36 $\mu\text{g}/\text{m}^3$). For reference, the average concentration at this site during the
 388 same time was 10.6 $\mu\text{g}/\text{m}^3$.

389 The “DEP Johnstown” site (AQS#42-021-0011, 40.310°N by 78.915°W) is in Cambria county,
 390 about 90 kilometers east of Pittsburgh (McDonnell 2017). While possessing a similar overall
 391 climate to Pittsburgh, it represents a more rural site. From April 3 to 6, 2017 (3 days), a single
 392 NPM sensor was deployed at this site. Data from this sensor was corrected using Eq. (3), and gave
 393 performance within the ranges observed at the other sites (r of 0.62, MAE of 1.9 $\mu\text{g}/\text{m}^3$, bias of -
 394 0.99 $\mu\text{g}/\text{m}^3$). For reference, the average concentration at this site during the same time was 6.2
 395 $\mu\text{g}/\text{m}^3$.

396 3.4. Sensitivity Analysis to Aerosol Composition

397 While the hygroscopic growth correction method discussed earlier used aerosol composition data
 398 from Aerosol Mass Spectrometer (AMS) measurements, not all locations have such data.
 399 However, aerosol composition data is also collected on a regular basis (one 24-hour sample every
 400 three to six days) by regulatory agencies such as the US EPA, and the data are publicly available
 401 (https://aqs.epa.gov/aqsweb/airdata/download_files.html). For example, aerosol composition data
 402 from Washington county, a site 35 kilometers from Pittsburgh, was used as a proxy for Pittsburgh
 403 aerosol composition; this resulted in a difference of less than 1% in the corrected $\text{PM}_{2.5}$
 404 concentration values. A sensitivity analysis for the hygroscopic-growth-based correction approach

405 with respect to aerosol composition was also performed using a range of plausible compositions
406 from the EPA Chemical Speciation Network (EPA Air Data). Full details are provided in the
407 supplemental materials (Section S.2). Briefly, the organic component fraction varied from 0.3 to
408 1, the sulfate component varied from 0 to 0.8, nitrate varied from 0 to 0.8, and ammonium varied
409 from 0 to 0.3. Overall, using this range of alternate chemical composition information in Eq. (3)
410 changed the resulting corrected $PM_{2.5}$ concentrations by up to 10% for typical cases, and up to
411 25% in extreme cases (see Figure S.4). Thus, for US sites where no local composition information
412 are available, publicly-available information from the nearest site in the EPA network can be used.
413 A similar approach may be possible for other countries.

414 **3.5. Short-Term Performance**

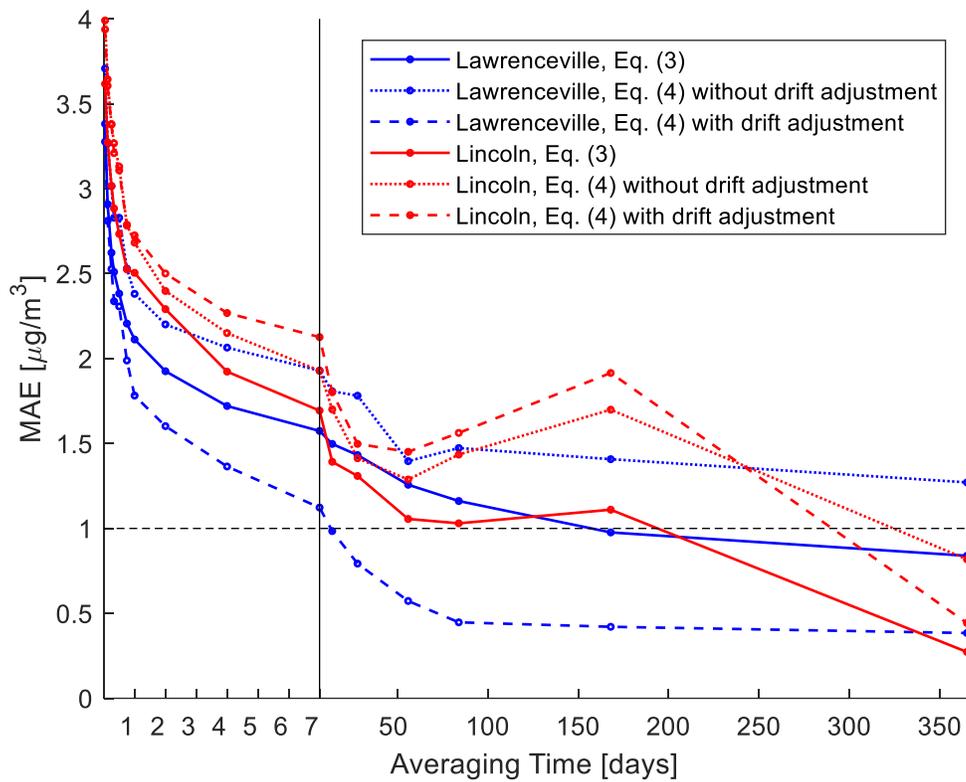
415 The US EPA has a short-term standard for $PM_{2.5}$ based on 24-hour average concentrations, set at
416 $35 \mu\text{g}/\text{m}^3$ (US EPA 2016a). We use this concentration level to test the performance of these low-
417 cost sensors under a short-term use case, where they might be used to alert citizens to potentially
418 unhealthy outdoor conditions. Although the EPA standard applies to a 24-hour average
419 concentration, we test the performance of the low-cost sensors using one-hour averages in order
420 to better mimic a near real-time alert scenario. This test is performed for the Lincoln site only since
421 hourly concentrations at Lawrenceville surpassed the threshold less than 1% of the time. True
422 positives occurred when both the NPM sensor (corrected using Eq. (3)) and BAM detected an
423 event (i.e. an hour when the average $PM_{2.5}$ concentration was higher than $35 \mu\text{g}/\text{m}^3$); false positives
424 were when only the NPM measured the event, and false negatives when the NPM failed to detect
425 an event seen by the BAM. The classification precision (Eq. (6)) of the sensor was 85% and its
426 classification recall (Eq. (7)) was 71% at the Lincoln site; for comparison, these values were 61%
427 and 78% respectively for the un-corrected, as-reported NPM data. Of the misclassifications, 15%
428 occurred when the BAM measured average concentrations between 30 and $40 \mu\text{g}/\text{m}^3$; the rest
429 represented larger discrepancies between the instruments. A one hour “grace period” was also
430 considered, i.e., if an event detection by one instrument leads or trails the other by up to an hour,
431 this was still counted as a true positive. With this grace period, the classification precision was
432 90% and classification recall was 97%, versus 73% and 97% respectively for the uncorrected data.
433 A graphical presentation of the results is provided in the supplemental materials (Figure S.15).

434 **3.6. Long-Term Performance**

435 Long-term assessment is necessary to categorize bias and assess data quality after extensive field
436 use of sensors. Additionally, long-term deployments can be used to generate data for
437 epidemiological studies, to evaluate different air quality models, and for verification of satellite
438 retrievals, which previously relied on sparse networks of expensive reference monitors. Previous
439 studies of lower-cost optical particle counters operating for up to four months report no evidence
440 of significant drift (Crilley et al. 2018). The long-term performance of NPM sensors was assessed
441 using data collected by the two sensors deployed at the Lawrenceville and Lincoln sites for a much
442 more extended period (e.g. more than a year of data at Lawrenceville collected over a 16-month

443 span). First, data corrected using Eq. (4) were used to assess the in-field drift-adjustment methods
 444 proposed in Section 2.5 to eliminate the “random walk” behavior observed when this correction
 445 approach was used over long periods. Based on these results, the “average of low readings” method
 446 worked best, reducing both the median bias and spread in biases at the Lawrenceville site.
 447 However, there were no clear improvements for these metrics at the Lincoln site (see supplemental
 448 materials Figure S.7 for details).

449 Figure 6 plots the MAE of the corrected sensor data with and without drift-adjustment (using the
 450 AL method) compared to the associated regulatory-grade instrument, as a function of averaging
 451 period. For weekly averages error was below about $2 \mu\text{g}/\text{m}^3$. For annual averages, errors were
 452 about or below $1 \mu\text{g}/\text{m}^3$, which is about 10% of the annual average concentrations for Pittsburgh.
 453 Drift-adjustment of measurements corrected with the fully-empirical Eq. (4) improved the
 454 performance at the Lawrenceville site (where concentrations are typically lower) to exceed that of
 455 Eq. (3); here, the errors fall below $1 \mu\text{g}/\text{m}^3$ for quarterly or seasonal averages. At the Lincoln site
 456 the drift-adjustment method tended to do nothing, or to slightly increase errors; this indicates that
 457 drift adjustment may not be required (or even suitable) for all locations or all sensors.



458
 459 Figure 6: Mean absolute error in $\text{PM}_{2.5}$ measurements for two NPM sensors during long-term
 460 deployments as a function of averaging period (note the differing horizontal axis scale on either
 461 side of the vertical black line). Solid lines represent measurements corrected using Eq. (3);
 462 dotted lines indicate measures corrected using Eq. (4) but not drift-adjusted; dashed lines

463 indicate measures corrected using Eq. (4) and drift-adjusted using the AL method. Points along
464 the lines indicate which specific averaging times were evaluated.

465 **4. Discussion**

466 Testing of a relatively large number of NPM (25 sensors at the Lawrenceville site) and PurpleAir
467 (9 sensors at the Lawrenceville site) low-cost PM_{2.5} sensors showed high mutual consistency
468 between the sensors, with mean inter-unit disagreement typically below 2.5 $\mu\text{g}/\text{m}^3$ and correlation
469 typically higher than 0.9. Systematic biases between instruments appear to account for the largest
470 fraction of the absolute differences; such biases may be assessed before and after field deployment
471 using collocations, but this may not fully account for in-field differences due to changes in aerosol
472 composition and size distributions over time (see supplemental materials Section S.2).

473 The first proposed correction equation was designed to account for two of the main factors
474 contributing to differences between optical measurements and the BAM instrument readings. First,
475 a hygroscopic growth factor was used to account for the increase in measured particle mass due to
476 ambient humidity. Second, a linear correction was applied to account for mismatches between the
477 size distribution and chemical composition of the factory calibration aerosol and the (ambient)
478 aerosol to be measured. We also evaluated alternative empirical correction equations which did
479 not rely on the assumptions necessary for estimating hygroscopic growth. For both NPM and
480 PurpleAir sensors, both correction approaches achieve similar performance, although even
481 following correction, relatively large differences in hourly averages (MAE of 3 to 4 $\mu\text{g}/\text{m}^3$) were
482 observed with respect to the BAM regulatory-grade instruments. This lack of consistency with
483 BAM instruments has also been observed previously (e.g. Zheng et al. 2018) and may not be
484 reconcilable with low-cost optical sensors. However, as data were averaged over longer periods,
485 accuracy was improved, such that longer-term (1 year or even seasonal) averages were likely to
486 have errors below 1 $\mu\text{g}/\text{m}^3$ (or about 10% of long-term average concentrations).

487 The proposed correction approaches handle aerosol chemical composition in different ways. The
488 empirical correction approach is calibrated based on an (implicitly assumed) long-term average
489 composition. Thus, when the true composition is close to this average, the empirical method is
490 likely to perform well. When the composition is closer to the seasonal averages than to the long-
491 term average, the hygroscopic-growth-based correction approach (which explicitly uses seasonal
492 average composition information) is likely to perform better. When the composition differs from
493 both the seasonal and long-term average compositions, neither approach will perform as well. For
494 the hygroscopic-growth-based correction approach, sensitivity of the results to compositional
495 variations was analyzed and found to only have a small effect in most cases (see supplemental
496 materials Section S.2). For the empirical approach, differing performance due to varying chemical
497 composition is not evaluated explicitly; however, its contribution is part of the overall error
498 associated with this method, assessed both in the short-term (contributing to the MAE and bias
499 noted in Figure 5) and in the long-term (contributing to the monthly bias noted for the “no drift
500 adjustment” method).

501 The efficacy of several proposed in-field drift-adjustment methods were also evaluated, although
502 these methods were only seen to have noticeable benefits for one of two sensors on which they
503 were tested, and so their utility is still unclear. The “average of low readings” (AL) method
504 adjusted for drift in the NPM sensor deployed at the Lawrenceville site, improving its performance
505 (see Figure 6). The same method made little impact for the Lincoln site. This could indicate either
506 that the Lincoln sensor did not experience significant drift, and therefore was not in need of
507 adjustments, or that such drift was not adjusted for by the method. Overall, while in-field sensor
508 drift should be corrected for, more research and in-field verification of drift-adjustment methods
509 are needed.

510 The NPM (with or without correction) detected 97% of occurrences when the BAM recorded $PM_{2.5}$
511 higher than $35 \mu\text{g}/\text{m}^3$. However, in the uncorrected low-cost sensor data, 27% of values above 35
512 $\mu\text{g}/\text{m}^3$ were not observed by the BAM; the corrections presented here (Eq. (3)), which account for
513 aerosol hygroscopic growth, reduced this error to 10%. Additionally, short-term performance of
514 the sensors after corrections met EPA recommendations for educational or informational
515 monitoring activities (Williams et al. 2014) (see supplemental materials, Figure S.13). Together,
516 these results indicate the potential for these sensors (after accounting for humidity effects) to be
517 used for assessing relative pollution levels in different neighborhoods.

518 The high level of mutual consistency and ability (with suitable corrections) to provide accurate
519 long-term averages makes these low-cost sensors useful for large-scale mapping campaigns to
520 determine long-term spatial patterns and temporal trends in $PM_{2.5}$. For real-time monitoring,
521 although these sensors can detect hourly “spikes” reasonably well, concentration values are only
522 accurate within about $\pm 4 \mu\text{g}/\text{m}^3$. Nevertheless, this is sufficient for qualitative indications of
523 relative short-term air quality, as indicated by the high concentration detection performance
524 (Section 3.5). The small size and ease of deployment of these units make them well suited to urban
525 monitoring. The low-cost (sub-\$250 each) PurpleAir sensors also incorporate a pair of optical
526 sensors, allowing for internal self-consistency checks to flag possible erroneous data. The cyclone
527 and inlet heater of the (sub-\$2,000 each) NPM sensors can protect the units from excessive dust
528 and humidity (to which PurpleAir sensors, which lack these features, may be more susceptible
529 during longer deployments). Finally, we note that these results are determined for the specific
530 environment of Pittsburgh, Pennsylvania; however, they can be generalized to other areas in
531 developed or OECD countries which are characterized by annual $PM_{2.5}$ mass concentrations less
532 than $20 \mu\text{g}/\text{m}^3$ and across both urban background (e.g. Lawrenceville) and source-impacted (e.g.
533 Lincoln and Parkway East) sites. Especially for the Plantower (PurpleAir) sensor, for which a two-
534 part correction equation was found optimal even within the range of $PM_{2.5}$ concentrations observed
535 in Pittsburgh, the response may be different at higher concentrations found in developing countries
536 like India or China. Similar to low-cost electrochemical gas sensors, which are designed to operate
537 at higher concentrations and therefore require specialized calibrations for lower ambient
538 concentrations (e.g. Malings et al. 2019), these low-cost $PM_{2.5}$ sensor may operate well using
539 factory calibrations in high-concentration environments, but require additional corrections such as

540 those presented here at lower ambient concentrations. Field tests of these instruments and
541 calibration techniques at sites throughout the USA and around the world is the subject of ongoing
542 work (e.g. Subramanian et al. 2018) and beyond the scope of this manuscript.

543 Considering future low-cost PM_{2.5} sensor deployments, the use of correction Eq. (3) is
544 recommended where information on particle composition is available (whether for the area in
545 question of for a nearby area with similar characteristics); otherwise, Eqs. (4) or (5) can be used.
546 The coefficients presented here for those corrections can be used as a starting point. Where
547 possible, however, new coefficients should be determined via collocation with regulatory-grade
548 instruments to account for local conditions, and depending on local conditions, further drift
549 adjustments using the techniques presented here (or others) may be necessary.

550 **Acknowledgements**

551 This work is part of the Center for Air, Climate and Energy Solution (CACES). Funding was
552 provided by the Environmental Protection Agency (Assistance Agreement Nos. RD83587301 and
553 83628601). It has not been formally reviewed by EPA. The views expressed in this document are
554 solely those of authors and do not necessarily reflect those of the Agency. EPA does not endorse
555 any products or commercial services mentioned in this publication. This work is also supported by
556 the Heinz Endowment Fund (Grants E2375 and E3145). The authors would like to thank Eric
557 Lipsky, Naomi Zimmerman, and S. Rose Eilenberg for assistance with instrument setup and
558 operation, as well as Ellis Robinson for providing chemical composition data from the AMS and
559 Rishabh Shah for providing data on particle size distributions. We thank Ruben Mamani-Paco and
560 Kelsea Palmer for the RAMP deployment at DEP Johnstown as part of an environmental
561 engineering laboratory class at St Francis University. The authors would also like to thank the staff
562 of the ACHD and Pennsylvania DEP for allowing us to deploy RAMP and PM sensors at their
563 sites. Finally, the authors would like to thank Spyros Pandis for helpful discussions.

564 **References**

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

567 Apte JS, Marshall JD, Cohen AJ, Brauer M. 2015. Addressing Global Mortality from Ambient
568 PM_{2.5}. *Environmental Science & Technology* 49:8057–8066;
569 doi:10.1021/acs.est.5b01236.

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

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

572 Asmi A, Wiedensohler A, Laj P, Fjaeraa A-M, Sellegri K, Birmili W, et al. 2011. Number size
573 distributions and seasonality of submicron particles in Europe 2008–2009. *Atmospheric*
574 *Chemistry and Physics* 11:5505–5538; doi:10.5194/acp-11-5505-2011.

- 575 Brook RD, Rajagopalan S, Pope CA, Brook JR, Bhatnagar A, Diez-Roux AV, et al. 2010.
576 Particulate Matter Air Pollution and Cardiovascular Disease: An Update to the Scientific
577 Statement From the American Heart Association. *Circulation* 121:2331–2378;
578 doi:10.1161/CIR.0b013e3181d8e1.
- 579 Burkart J, Steiner G, Reischl G, Moshhammer H, Neuberger M, Hitzemberger R. 2010.
580 Characterizing the performance of two optical particle counters (Grimm OPC1.108 and
581 OPC1.109) under urban aerosol conditions. *Journal of Aerosol Science* 41:953–962;
582 doi:10.1016/j.jaerosci.2010.07.007.
- 583 Cabada JC, Khlystov A, Wittig AE, Pilinis C, Pandis SN. 2004. Light scattering by fine particles
584 during the Pittsburgh Air Quality Study: Measurements and modeling. *Journal of*
585 *Geophysical Research* 109; doi:10.1029/2003JD004155.
- 586 Cerully KM, Bougiatioti A, Hite JR, Guo H, Xu L, Ng NL, et al. 2015. On the link between
587 hygroscopicity, volatility, and oxidation state of ambient and water-soluble aerosols in the
588 southeastern United States. *Atmospheric Chemistry and Physics* 15:8679–8694;
589 doi:10.5194/acp-15-8679-2015.
- 590 Chow JC, Doraiswamy P, Watson JG, Chen L-WA, Ho SSH, Sodeman DA. 2008. Advances in
591 Integrated and Continuous Measurements for Particle Mass and Chemical Composition.
592 *Journal of the Air & Waste Management Association* 58:141–163; doi:10.3155/1047-
593 3289.58.2.141.
- 594 Crilley LR, Shaw M, Pound R, Kramer LJ, Price R, Young S, et al. 2018. Evaluation of a low-cost
595 optical particle counter (Alphasense OPC-N2) for ambient air monitoring. *Atmospheric*
596 *Measurement Techniques* 11:709–720; doi:10.5194/amt-11-709-2018.
- 597 Eeftens M, Tsai M-Y, Ampe C, Anwander B, Beelen R, Bellander T, et al. 2012. Spatial variation
598 of PM_{2.5}, PM₁₀, PM_{2.5} absorbance and PM_{coarse} concentrations between and within 20
599 European study areas and the relationship with NO₂ – Results of the ESCAPE project.
600 *Atmospheric Environment* 62:303–317; doi:10.1016/j.atmosenv.2012.08.038.
- 601 English PB, Olmedo L, Bejarano E, Lugo H, Murillo E, Seto E, et al. 2017. The Imperial County
602 Community Air Monitoring Network: A Model for Community-based Environmental
603 Monitoring for Public Health Action. *Environmental Health Perspectives* 125;
604 doi:10.1289/EHP1772.
- 605 EPA Air Data. https://aqs.epa.gov/aqsweb/airdata/download_files.html.
- 606 Gu P, Li HZ, Ye Q, Robinson ES, Apte JS, Robinson AL, et al. 2018. Intra-city variability of PM
607 exposure is driven by carbonaceous sources and correlated with land use variables.
608 *Environmental Science & Technology*; doi:10.1021/acs.est.8b03833.
- 609 Hacker K. 2017. Air Monitoring Network Plan for 2018.

610 Jayaratne R, Liu X, Thai P, Dunbabin M, Morawska L. 2018. The influence of humidity on the
611 performance of a low-cost air particle mass sensor and the effect of atmospheric fog.
612 *Atmospheric Measurement Techniques* 11:4883–4890; doi:10.5194/amt-11-4883-2018.

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

616 Jiao W, Hagler G, Williams R, Sharpe R, Brown R, Garver D, et al. 2016. Community Air Sensor
617 Network (CAIRSENSE) project: evaluation of low-cost sensor performance in a suburban
618 environment in the southeastern United States. *Atmospheric Measurement Techniques*
619 9:5281–5292; doi:10.5194/amt-9-5281-2016.

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

623 Kelly KE, Whitaker J, Petty A, Widmer C, Dybwad A, Sleeth D, et al. 2017. Ambient and
624 laboratory evaluation of a low-cost particulate matter sensor. *Environmental Pollution*
625 221:491–500; doi:10.1016/j.envpol.2016.12.039.

626 Koehler KA, Peters TM. 2015. New Methods for Personal Exposure Monitoring for Airborne
627 Particles. *Current Environmental Health Reports* 2:399–411; doi:10.1007/s40572-015-
628 0070-z.

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

632 Liu D, Zhang Q, Jiang J, Chen D-R. 2017. Performance calibration of low-cost and portable
633 particular matter (PM) sensors. *Journal of Aerosol Science* 112:1–10;
634 doi:10.1016/j.jaerosci.2017.05.011.

635 Malings C, Tanzer R, Hauryliuk A, Kumar SPN, Zimmerman N, Kara LB, et al. 2019.
636 Development of a general calibration model and long-term performance evaluation of low-
637 cost sensors for air pollutant gas monitoring. *Atmospheric Measurement Techniques*
638 12:903–920; doi:10.5194/amt-12-903-2019.

639 McDonnell P. 2017. Commonwealth of Pennsylvania Department of Environmental Protection
640 2016 Annual Ambient Air Monitoring Network Plan.

641 Met One. 2018. NPM 2 Operation Manual, Rev. B.

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

- 645 Petters MD, Kreidenweis SM. 2007. A single parameter representation of hygroscopic growth and
646 cloud condensation nucleus activity. *Atmospheric Chemistry and Physics* 7:1961–1971;
647 doi:10.5194/acp-7-1961-2007.
- 648 Piedrahita R, Xiang Y, Masson N, Ortega J, Collier A, Jiang Y, et al. 2014. The next generation
649 of low-cost personal air quality sensors for quantitative exposure monitoring. *Atmospheric*
650 *Measurement Techniques* 7:3325–3336; doi:10.5194/amt-7-3325-2014.
- 651 Pope CA, Burnett RT, Thun MJ, Calle EE, Krewski D, Ito K, et al. 2002. Lung cancer,
652 cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *JAMA*
653 287: 1132–1141.
- 654 Rai AC, Kumar P, Pilla F, Skouloudis AN, Di Sabatino S, Ratti C, et al. 2017. End-user perspective
655 of low-cost sensors for outdoor air pollution monitoring. *Science of The Total Environment*
656 607–608:691–705; doi:10.1016/j.scitotenv.2017.06.266.
- 657 Schwartz J, Dockery DW, Neas LM. 1996. Is daily mortality associated specifically with fine
658 particles? *J Air Waste Manag Assoc* 46: 927–939.
- 659 Snyder EG, Watkins TH, Solomon PA, Thoma ED, Williams RW, Hagler GSW, et al. 2013. The
660 Changing Paradigm of Air Pollution Monitoring. *Environmental Science & Technology*
661 47:11369–11377; doi:10.1021/es4022602.
- 662 Solomon PA, Sioutas C. 2008. Continuous and Semicontinuous Monitoring Techniques for
663 Particulate Matter Mass and Chemical Components: A Synthesis of Findings from EPA’s
664 Particulate Matter Supersites Program and Related Studies. *Journal of the Air & Waste*
665 *Management Association* 58:164–195; doi:10.3155/1047-3289.58.2.164.
- 666 Subramanian R, Ellis A, Torres-Delgado E, Tanzer R, Malings C, Rivera F, et al. 2018. Air Quality
667 in Puerto Rico in the Aftermath of Hurricane Maria: A Case Study on the Use of Lower
668 Cost Air Quality Monitors. *ACS Earth and Space Chemistry* 2:1179–1186;
669 doi:10.1021/acsearthspacechem.8b00079.
- 670 Tan Y, Lipsky EM, Saleh R, Robinson AL, Presto AA. 2014. Characterizing the Spatial Variation
671 of Air Pollutants and the Contributions of High Emitting Vehicles in Pittsburgh, PA.
672 *Environmental Science & Technology* 48:14186–14194; doi:10.1021/es5034074.
- 673 Tsujita W, Yoshino A, Ishida H, Moriizumi T. 2005. Gas sensor network for air-pollution
674 monitoring. *Sensors and Actuators B: Chemical* 110:304–311;
675 doi:10.1016/j.snb.2005.02.008.
- 676 US EPA. 2016a. EPA NAAQS Table.
- 677 US EPA. 2016b. Quality Assurance Guidance Document 2.12: Monitoring PM_{2.5} in Ambient Air
678 Using Designated Reference or Class I Equivalent Methods.
- 679 Wang E. 2019. Plantower calibration (E-mail communication).

- 680 Wang Y, Li J, Jing H, Zhang Q, Jiang J, Biswas P. 2015. Laboratory Evaluation and Calibration
681 of Three Low-Cost Particle Sensors for Particulate Matter Measurement. *Aerosol Science*
682 and *Technology* 49:1063–1077; doi:10.1080/02786826.2015.1100710.
- 683 Watson JG, Chow JC, Moosmüller H, Green M, Frank N, Pitchford M. 1998. Guidance for using
684 continuous monitors in PM_{2.5} monitoring networks.
- 685 White RM, Paprotny I, Doering F, Cascio WE, Solomon PA, Gundel LA. 2012. Sensors and
686 “apps” for community-based: Atmospheric monitoring. *EM: Air and Waste Management*
687 Association’s Magazine for Environmental Managers 2012: 36–40.
- 688 Williams R, Vallano D, Polidori A, Garvey S. 2018. Spatial and Temporal Trends of Air Pollutants
689 in the South Coast Basin Using Low Cost Sensors.
- 690 Williams R, Vasu Kilaru, Snyder E, Kaufman A, Dye T, Rutter A, et al. 2014. Air Sensor
691 Guidebook.
- 692 Wilson WE, Chow JC, Claiborn C, Fusheng W, Engelbrecht J, Watson JG. 2002. Monitoring of
693 particulate matter outdoors. *Chemosphere* 49: 1009–1043.
- 694 Zheng T, Bergin MH, Johnson KK, Tripathi SN, Shirodkar S, Landis MS, et al. 2018. Field
695 evaluation of low-cost particulate matter sensors in high- and low-concentration
696 environments. *Atmospheric Measurement Techniques* 11:4823–4846; doi:10.5194/amt-
697 11-4823-2018.
- 698 Zhou Y, Zheng H. 2016. PMS5003 series data manual.
- 699 Zikova N, Hopke PK, Ferro AR. 2017a. Evaluation of new low-cost particle monitors for PM_{2.5}
700 concentrations measurements. *Journal of Aerosol Science* 105:24–34;
701 doi:10.1016/j.jaerosci.2016.11.010.
- 702 Zikova N, Masiol M, Chalupa D, Rich D, Ferro A, Hopke P. 2017b. Estimating Hourly
703 Concentrations of PM_{2.5} across a Metropolitan Area Using Low-Cost Particle Monitors.
704 *Sensors* 17:1922; doi:10.3390/s17081922.
- 705 Zimmerman N, Presto AA, Kumar SPN, Gu J, Hauryliuk A, Robinson ES, et al. 2018. A machine
706 learning calibration model using random forests to improve sensor performance for lower-
707 cost air quality monitoring. *Atmospheric Measurement Techniques* 11:291–313;
708 doi:10.5194/amt-11-291-2018.