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

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

[1]Carnegie Mellon University
[2]isni—0000000120970344—Carnegie Mellon University

November 30, 2022

Abstract

Low-cost fine particle mass (PM2.5) 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, which must be accounted for to improve their in-field accuracy. We conduct long-term tests of two low-cost PM2.5 sensors: Met-One NPM and PurpleAir PA-II units. We find a high level of self-consistency within each sensor type after testing 25 NPM and 9 PurpleAir units. We develop corrections for the low-cost sensor measurements to better match regulatory-grade data through collocation with Beta Attenuation Monitors (BAM). The first correction based on a physical model accounts for hygroscopic growth using particle composition and corrects for particle mass below the optical sensor detection limit by collocation with a BAM. A second fully-empirical correction uses linear or quadratic functions of environmental variables. Either model yields comparable improvements over raw measurements. Sensor performance is 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 is reduced below 1 µg/m3. These sensors are thus suitable for supplementing regulatory-grade instruments in sparsely monitored regions and for conducting hotspot mapping to understand air quality variability in urban areas.

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

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

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

[2]OSU-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 $\mu g/m^3$) 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 $\mu g/m^3$. 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 (20-23ºC) 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 $\mu g/m^3$) 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.

2. Methods

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 PM_{10}. 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, 40.308ºN by 79.869ºW) 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    $\mu g/m^3$ in 2017, with a one-hour maximum of 162 $\mu g/m^3$. 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, 40.465ºN by 79.961ºW) 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 $\mu g/m^3$ in 2017, with a maximum one-hour concentration of 67 $\mu g/m^3$. 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 µg/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 µg/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 22ºC and 35% relative humidity (the conditions at which
168 regulatory data are reported), and is calculated as follows:

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

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

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
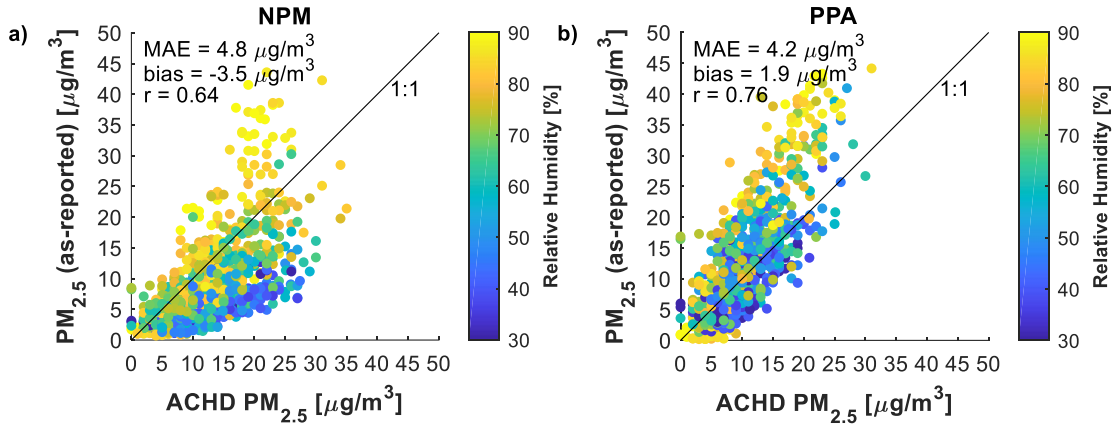183 humidity; and D_p is the particle diameter (see Table S.1 for details).

5

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

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

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

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

2.4. Empirical Correction Methods

The hygroscopic growth factor correction method described above is based on information about the specific aerosol chemical composition of the sensor deployment area, which may not be 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    $[\text{corrected PM}_{2.5}]_{\text{NPM}} = \alpha_0 + \alpha_1[\text{PM}_{2.5}]_{\text{NPM}} + \alpha_2 T + \alpha_3 RH + \alpha_4[\text{PM}_{2.5}]^2_{\text{NPM}} +$
220    $\alpha_5[\text{PM}_{2.5}]_{\text{NPM}}T + \alpha_6[\text{PM}_{2.5}]_{\text{NPM}}RH + \alpha_7 T^2 + \alpha_8 TRH + \alpha_9 RH^2$ (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 µg/m³:

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

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 PM$_{2.5}$ readings by up to 4 µg/m³ 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/m^3 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/m^3 (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
$$\text{classification precision} = \frac{\sum_{i=1}^{n} \mathbb{I}(c_i > \tau)\mathbb{I}(\hat{c}_i > \tau)}{\sum_{i=1}^{n} \mathbb{I}(c_i > \tau)} \cdot 100\% \tag{6}$$

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

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.

3. Results

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

3.1. Consistency between Sensors

To determine the consistency between sensors, pairwise comparisons of 1-hour-averaged data were made among NPM and PurpleAir sensors (i.e. NPM with NPM and PurpleAir with PurpleAir) collocated at either the Lawrenceville of Lincoln site during the same period. At Lawrenceville, during the RAMP collocations, temperature varied between -20 and +31°C and relative humidity varied from 22% to 97%; at Lincoln, temperature varied from -3 to +43°C and humidity varied between 17% and 97% (as measured by the RAMPs' onboard sensors). Figure 2 presents the results of these inter-comparisons; only results for sensors collocated for at least 3 days (36 1-hour averages) are presented. Overall, mutual correlations were strong (typically r > 0.9) and were higher at the Lincoln site likely due to the wider range of concentrations. Absolute differences in as-reported readings were typically about 2 μg/m^3 or less, which includes systematic biases between sensors generally on the order of $\pm$1 μg/m^3. This is similar to prior results for Alphasense OPC-N2 optical PM$_{2.5}$ sensors, which are more than twice the price of PurpleAir units (Crilley et al. 2018).
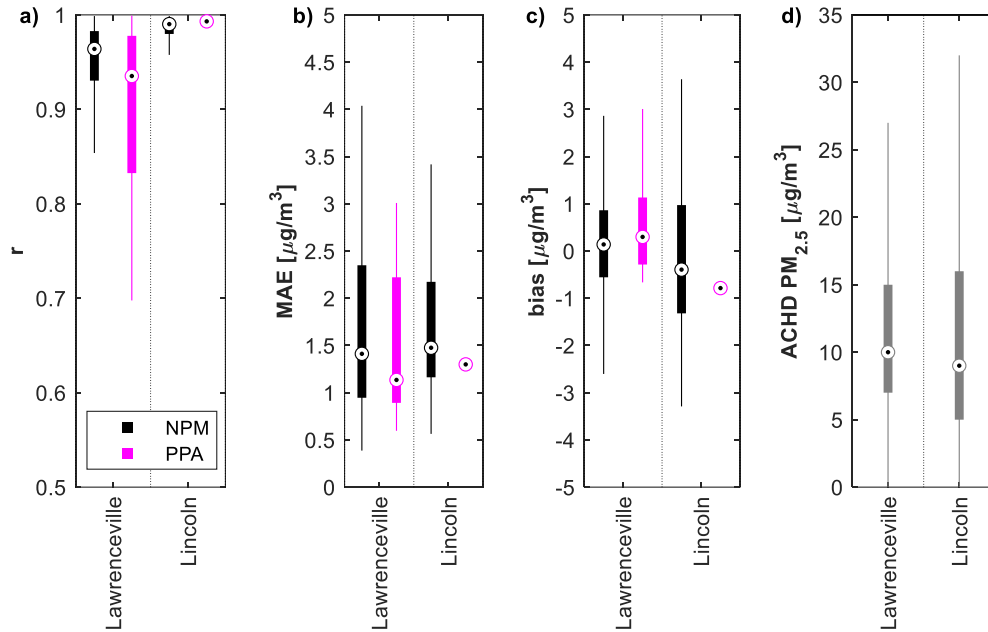
Figure 2: Inter-comparison of as-reported one-hour-average data between sensors during collocation periods at both sites. In the boxplots, circles with dots denote the median, thick bars denote the interquartile range, and thin bars denote the 95% confidence range. Black boxplots indicate metric ranges for pairs of NPM sensors, and purple boxplots indicate ranges for pairs of PurpleAir sensors. This represents 114 NPM pairs at Lawrenceville, 66 NPM pairs at Lincoln, 16 PurpleAir pairs at Lawrenceville and 1 PurpleAir pair at Lincoln. For reference, the ranges of concentrations measured by BAM instruments at the sites during the same time are depicted in panel d.

Figure 3 compares hourly averages of as-reported data from NPM sensors at Lawrenceville to those collected by PurpleAir sensors at Lawrenceville as a function of humidity (the median readings of all sensors active at the site at the same time are shown). At low humidity, PurpleAir readings were about twice that of the NPM, while at high humidity the ratio of readings approached one; comparisons made between raw readings of the two sensor types would therefore be heavily humidity-dependent. There are several likely causes for these differences. First, the NPM possesses an inlet heater with a 4-second residence time which activates when relative humidity exceeds 40%. However, this residence time may not be sufficient to totally remove humidity effects on the NPM. Second, these instruments are calibrated differently. The NPM is calibrated with 0.6µm polystyrene latex spheres (Met One 2018), while PurpleAir Plantowers are calibrated with ambient aerosol across several cities in China (Wang 2019). They therefore respond differently when exposed to a common aerosol which differs from their calibration aerosols.
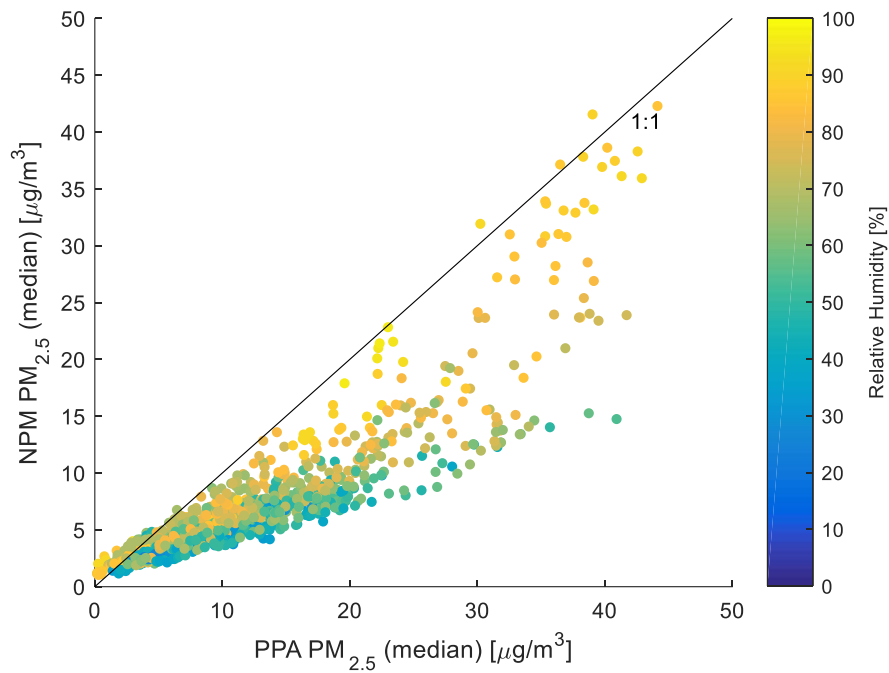
10

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

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

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

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

Figure 5 assesses the performance of the designated "testing" set of low-cost sensors deployed to the Lawrenceville and Lincoln sites during the March to June (at Lawrenceville) and October to February (at Lincoln) collocation periods. The figure compares as-reported data to data corrected using the hygroscopic-growth-based approach of Eq. (3) (with appropriate coefficients for NPM or PurpleAir sensors) and data corrected using the fully-empirical approaches of Eq. (4) for NPM or Eq. (5) for PurpleAir. In all cases hourly-averaged data were used. In terms of correlation (Figure 5a), no improvement was made for PurpleAir sensors, while only a modest improvement resulted from correction of the NPM sensors. In terms of MAE (Figure 5b) and bias (Figure 5c), however, both correction approaches resulted in noticeable improvements. For NPM sensors, both the physics-based Eq. (3) and fully-empirical Eq. (4) gave comparable performance. For PurpleAir sensors, the fully-empirical approach of Eq. (5) provided a smaller spread of MAE and bias results as compared to Eq. (3), while the median MAE of both approaches were almost the same, and the median bias of Eq. (5) was slightly worse. Overall both correction approaches improved upon the as-reported data and there was no significant difference between their performance.
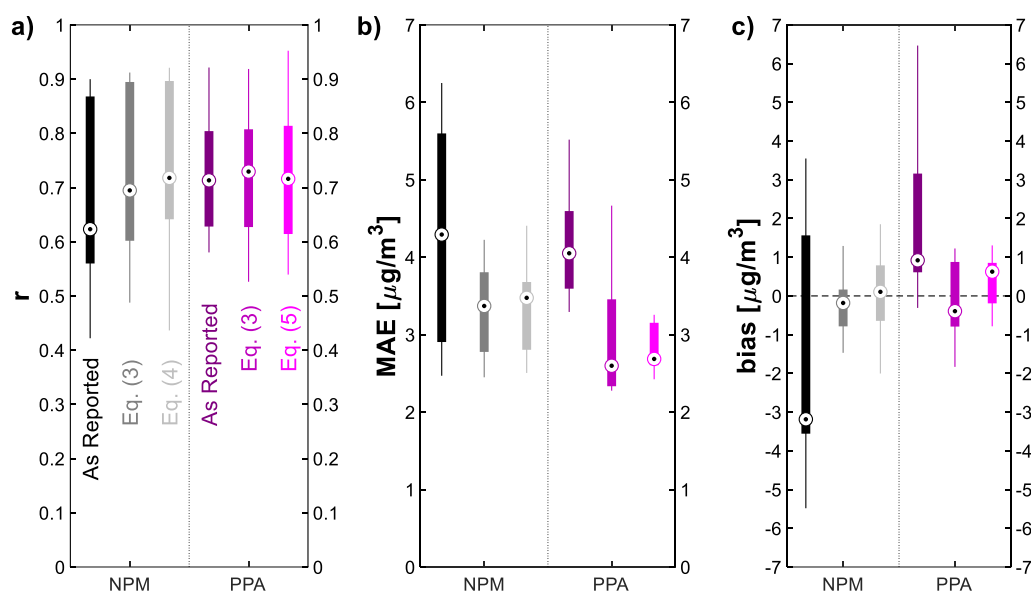
Figure 5: Performance metrics of one-hour-average as-reported and corrected sensor data compared to BAM instruments during collocation at both the Lawrenceville and Lincoln sites. Results shown relate to a total of 17 NPM and 5 PurpleAir sensors of the "testing" set. Corrections are performed using either the approach of Eq. (3), with appropriate coefficients for NPM or PurpleAir, or the approaches of Eq. (4) for NPM and Eq. (5) for PurpleAir.

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

13

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

		Met-One NPM		PurpleAir PPA	
θ_0	Summer	5.28 ± 0.09	$\mu g/m^3$	5.4 ± 0.4	$\mu g/m^3$
	Winter	2.03 ± 0.08	$\mu g/m^3$	-0.3 ± 0.2	$\mu g/m^3$
	Other	1.68 ± 0.13	$\mu g/m^3$	3.7 ± 0.1	$\mu g/m^3$
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	

3.3. Performance Assessment at Other Regulatory Sites

To further assess the performance of these sensor corrections at locations independent of where they were developed, several sensors were tested at two additional sites. The "Parkway East" site (AQS#42-003-1376, 40.437ºN by 79.864ºW) represents a roadside location (Hacker 2017), and thus may have a different, vehicular traffic-influenced particle composition and size distribution than either the urban background or coke oven-impacted sites at which the corrections were developed. Between September 6 and 27, 2018 (21 days), two PurpleAir sensors were collocated at this site. Data from these sensors were corrected using Eq. (3). These provided comparable results to testing at the Lincoln and Lawrenceville sites (median r of 0.71, median MAE of 2.7 µg/m³, median bias of 0.36 µg/m³). For reference, the average concentration at this site during the same time was 10.6 µg/m³.

The "DEP Johnstown" site (AQS#42-021-0011, 40.310ºN by 78.915ºW) is in Cambria county, about 90 kilometers east of Pittsburgh (McDonnell 2017). While possessing a similar overall climate to Pittsburgh, it represents a more rural site. From April 3 to 6, 2017 (3 days), a single NPM sensor was deployed at this site. Data from this sensor was corrected using Eq. (3), and gave performance within the ranges observed at the other sites (r of 0.62, MAE of 1.9 µg/m³, bias of -0.99 µg/m³). For reference, the average concentration at this site during the same time was 6.2 µg/m³.

3.4. Sensitivity Analysis to Aerosol Composition

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

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 µg/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 µg/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 µg/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).

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 µg/m^3. For annual averages, errors were
452   about or below 1 µg/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 µg/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.
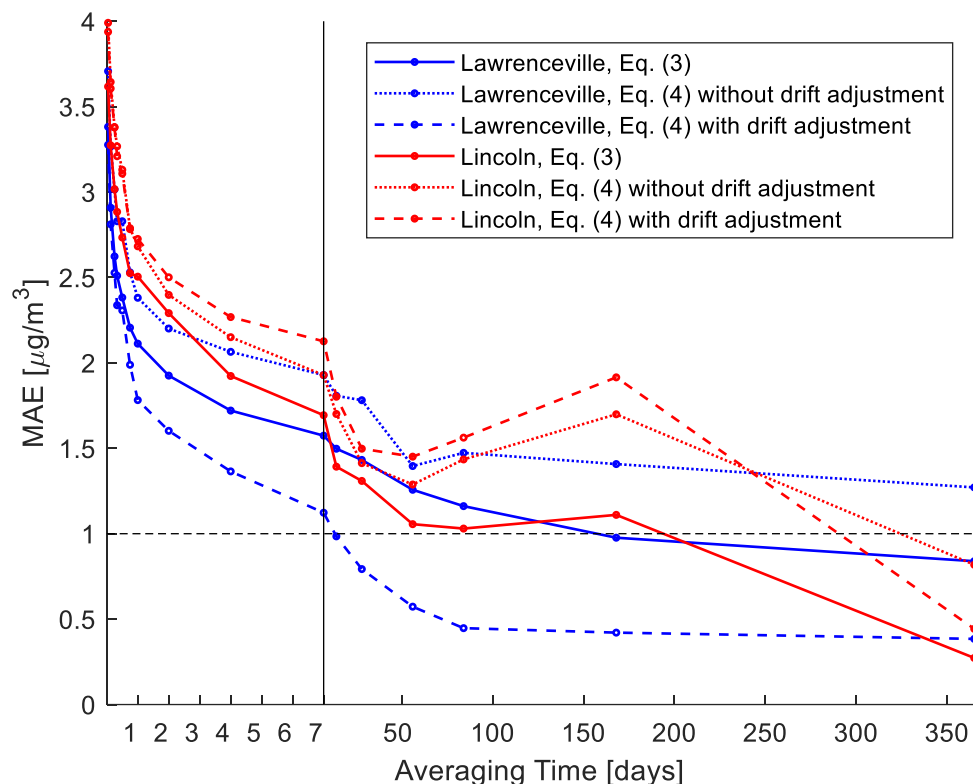

458

459   Figure 6: Mean absolute error in 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

16

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.

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 µg/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 µg/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 µg/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).

17

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 $\mu g/m^3$. However, in the uncorrected low-cost sensor data, 27% of values above 35
512 $\mu g/m^3$ 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 ±4 $\mu g/m^3$. 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 $\mu g/m^3$ 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

18

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.

Acknowledgements

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   PM2.5. 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.

19

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.0b013e3181dbece1.

579   Burkart J, Steiner G, Reischl G, Moshammer H, Neuberger M, Hitzenberger 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 PM2.5, PM10, PM2.5 absorbance and PMcoarse 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   2016Annual 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 PM2.5 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 PM2.5 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 PM2.5
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 PM2.5 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.

Supplemental ~~Information~~ Materials For:

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

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

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

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

This document contains information meant to supplement and support the information presented in the paper referenced above. Section S.1 provides pictures of the RAMP sensor and associated PM sensors. Section S.2 describes the method for computing hygroscopic growth factors and investigates the sensitivity of these factors to changes in aerosol composition. Section S.3 provides details on how empirical correction methods were selected. Section S.4 outlines the methods proposed for sensor drift adjustment, and provides results relating to these methods. Section S.5 provides formulae for the assessment metrics presented in this paper. Section S.6 presents data collected on particle size distributions in Pittsburgh. Section S.7 presents various results providing further details about the performance of various correction approaches applied to low-cost PM sensor data. Finally, Section S.8 provides a figure depicting the results related to the short-term use case assessment of the low-cost sensors.

1

S.1. RAMP and PM Sensor Picture

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

S.2. Correction Methods – Hygroscopic Growth Factor Computation

This hygroscopic growth factor is computed as:

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

where:

$$a_w(T, RH) = RH \exp\left(\frac{4\sigma_w M_w}{\rho_w RT D_p}\right)^{-1} \tag{S.2}$$

κ_{bulk} is the hygroscopicity of bulk aerosol; $\kappa_{\text{bulk}} = \sum_i x_i \kappa_i$ where. x_i and κ_i are the volume fraction hygroscopocity parameters of the ith component comprising the particle. Organic, sulfate, nitrate and ammonium are assumed as the main components comprising the particle. The fractional contributions of these chemical components to PM$_{2.5}$ during summer, winter, and as an annual average (applied to other periods) are obtained from recent AMS measurements in Pittsburgh (Gu et al. 2018) and their hygroscopocity parameters are adopted from literature (Cerully et al. 2015; Petters and Kreidenweis 2007). a_w is the water activity parameter, estimated using Eq. (S.2), where σ_w, M_w, and ρ_w represent the surface tension, molecular weight and density of water, respectively; T is the absolute temperature, R is the ideal gas constant, RH is ambient relative humidity; D_p is the particle diameter, adopted as volume median diameter from long-term size distribution measurements using SMPS in Pittsburgh. Table S.1 lists different parameter values used in hygroscopic growth factor calculation.

42   Table S.1: Parameters used in hygroscopic growth factor calculation

Parameter	Value			Unit	Source
	Summer	Winter	Other		
κ_{OA}	0.15	0.15	0.15	-	(Cerully et al. 2015)
κ_{SO4}	0.5	0.5	0.5	-	(Petters and Kreidenweis 2007)
κ_{NO3}	0.6	0.6	0.6	-	(Petters and Kreidenweis 2007)
κ_{NH4}	0.5	0.5	0.5	-	(Petters and Kreidenweis 2007)
x_{OA}	0.64	0.41	0.53	-	(Gu et al. 2018)
x_{SO_4}	0.24	0.16	0.20	-	(Gu et al. 2018)
x_{NO_3}	0.04	0.29	0.165	-	(Gu et al. 2018)
x_{NH_4}	0.08	0.15	0.115	-	(Gu et al. 2018)
κ_{bulk}	0.26	0.34	0.30	-	
σ_w	0.072	0.072	0.072	N/m	
M_w	0.018	0.018	0.018	kg/mol	
ρ_w	1000	1000	1000	kg/m^3	
R	8.314	8.314	8.314	J/mol K	
D_p	200	200	200	nm	

43

44

45   Figure S.2: Example of how the hygroscopic growth factor varies with humidity in summer,
46   winter, and otherwise.

47   To examine the sensitivity of the hygroscopic growth factor to different aerosol compositions, a
48   sensitivity analysis was conducted for differing aerosol compositions resulting in different κ_{bulk}
49   values. Using data from the EPA Chemical Speciation Network for 2018 (available online at
50   https://aqs.epa.gov/aqsweb/airdata/download_files.html), the fractional composition of PM$_{2.5}$ as
51   carbonaceous matter, sulfate, nitrate, and ammonium were determined, and annual average bulk
52   hygroscopicity factors were computed for each of 139 sites where these data are available.
53   Carbonaceous mass was computed using a sum of elemental carbon and organic mass (OM,
54   calculated as organic carbon multiplied by 1.8) (Turpin and Lim 2001). The κ value for EC was
55   assumed the same as for OM; EC was typically from 8% to 18% of OM, so errors due to this

3

56 assumption should be small. Histograms for the fractional composition of these components
57 across network sites are presented in Figure S.3.

58

59 Figure S.3: Histograms representing the ranges in fractional compositions for carbonaceous,
60 sulfate, nitrate, and ammonium components of $PM_{2.5}$ measured at 139 sites in the US EPA
61 Chemical Speciation Network.

62 Figure S.4 presents the results as a function of relative humidity (in five percentage point
63 increments), for a base concentration of 10 µg/m^3 at an ambient temperature of 22°C. The
64 boxplots indicate the spread (across the speciation network sites) of the percent difference
65 between PM readings corrected using each of the 139 speciation sites and PM readings corrected
66 using the Pittsburgh values of κ_{bulk}, as determined from the AMS data and presented in Table
67 S.1. The solid black line indicates results when using only the nearest speciation site to
68 Pittsburgh outside of Allegheny county (in Washington county, about 35 km away). Overall, the
69 failure to use an appropriate local κ_{bulk} factor typically (i.e. for the interquartile range of site
70 compositions) causes less than 10% errors and may lead to up to 25% errors in extreme cases.
71 However, using a nearby local factor, errors can be reduced below 1%. Therefore, it is
72 recommended to use speciation information from the closest available station if specific local
73 information is not available. It should further be noted that these results all employ the same
74 linear correction coefficients from Eq. (3) as were determined for Pittsburgh, as presented in
75 Table 1; if local collocations are performed to determine appropriate coefficients for each area,
76 the resulting errors are likely to be further reduced or eliminated. Furthermore, while PM
77 composition and size distribution at a given location may change significantly from day-to-day
78 (Saha et al. 2019), the settings used in the proposed corrections reflect long-term averages. Thus,
79 while they cannot capture such short-term fluctuations (as is reflected by the residual uncertainty
80 in the presented results), they provide more robust performance in the long-term without the

81 need for simultaneous composition and size distribution information to be collected alongside
82 low-cost sensor data.

PM$_{2.5}$ = 10 μg/m^3, T = 22°C
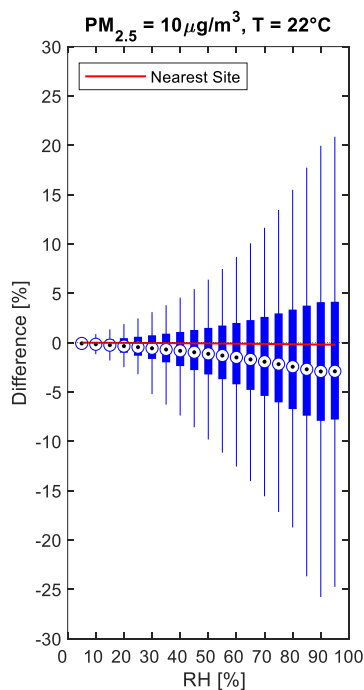
83

84 Figure S.4: Sensitivity analysis of hygroscopic growth rate corrections. Boxplots indicate the
85 range of percent differences between corrections performed using each of the chemical
86 compositions measured at sites in the EPA Chemical Speciation Network and corrections
87 performed using the Pittsburgh chemical composition (as described above). Results are binned
88 by relative humidity. The solid red line indicates the percent differences from using chemical
89 composition data at the nearest non-Pittsburgh site.

90

91 Several explanatory factors were considered for the empirical correction method. Dewpoint DP
92 was considered as a factor related to condensation that might serve as a proxy for the
93 hygroscopic growth factor which is independent of aerosol composition. Furthermore, humidity
94 is known to affect the performance of optical particle sensors directly (e.g. Jayaratne et al. 2018),
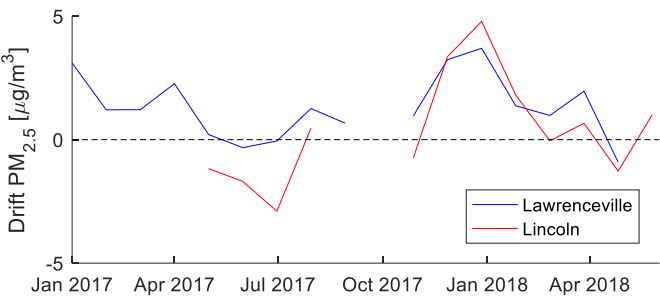95 and so relative humidity RH was included as a factor. Finally, temperature T was included as a

factor since it has been observed to affect the performance of optical sensor components (Johnson et al. 2016; Jayaratne et al. 2018; Zheng et al. 2018).
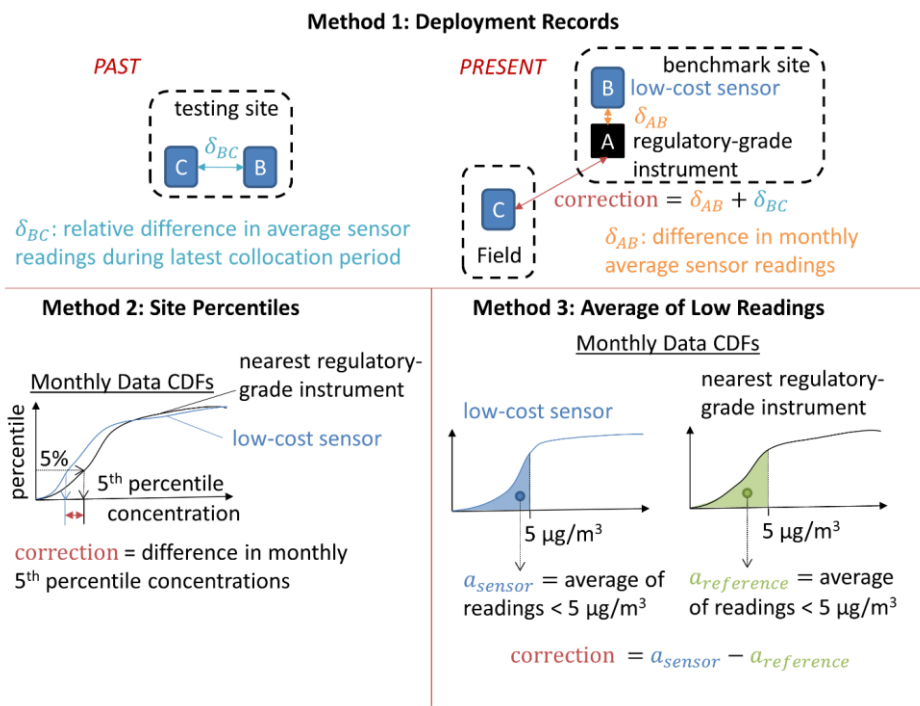
Various combinations of the as-reported sensor readings and the above inputs into various functional forms and with different application thresholds were applied to generate correction equations. Two functional forms were considered: linear and quadratic regression models. Thresholds were considered to define different subsets of the domain over which different functional parameters could be applied, allowing for piecewise-linear or piecewise-quadratic functions. Models without thresholds were considered, as well as models with single or multiple threshold values chosen from among 5, 10, 15, 20, 30, 40, and 50 µg/m^3 (as determined from the raw sensor reading). For reference, ambient concentrations in Pittsburgh typically range from 3 to 20 µg/m^3.

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 assessed for the various models are included as supplementary data. The performance of each correction model on the test sensor set was scored using a heuristic combining various performance metrics (bias, mean absolute error, r, and threshold classification score) across a range of concentrations experienced at both collocation sites and penalizing the complexity of the model (and therefore its propensity to overfit to training data). The format of this scoring system was inspired by the "Eureqa" equation discovery system of Schmidt and Lipson (2009), with modifications for the specific context of this problem (see the supplementary data for the resulting metrics). The resulting metrics are available in a table attached to the supplementary materials but separate from this document. 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).

6

122   **S.5.S.4.   Drift-Adjustment Methods**

123

124   Figure S.53: Illustration of observed NPM sensor drift at the Lincoln and Lawrenceville sites.
125   Drift is depicted as the difference in monthly average readings of the NPM sensor, corrected
126   using Eq. (4), versus the collocated regulatory-grade instrument at each site.

127

128   Figure S.64: Diagrams of the three proposed drift-adjustment methods.

7

129 Note that in the Average of Low Readings method, if no readings within a month are below 5
130 micrograms per cubic meter, the minimum reading for that month is instead used as the basis for
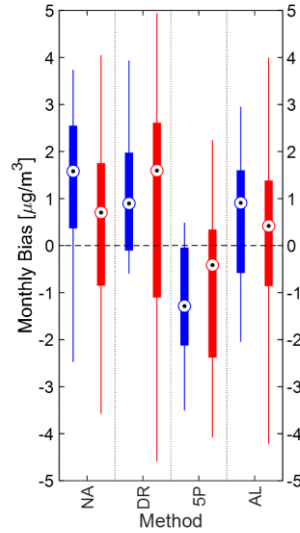131 the adjustment.

132 _____

133 Figure S.7: Performance of various drift-adjustment methods in reducing the bias in monthly
134 averages; NA – no adjustment applied; DR – drift-adjusted using deployment records; 5P –
135 drift-adjusted using percentiles of nearest reference site; AL – drift-adjusted using averages of
136 low readings at nearest reference site. Performance is determined separately for the NPM
137 instruments deployed for extended periods at the Lawrenceville (blue) and Lincoln (red) sites.
138 Corrections are performed using Eq. (4).

139 Figure S.7 shows the spread in monthly biases (difference between the monthly average readings
140 of the corrected sensors and the BAM instruments) for both long-term collocation sites, both
141 without drift-adjustment and with the three proposed drift-adjustment methods. Note that these
142 biases are for the single long-term-deployment sensor at each site, whereas Figure 5 in the main
143 paper presented results for the entire "testing" set of sensors over a shorter period.

144 ~~S.6.~~S.5.   **Assessment metrics**

145 For n measurements of concentration by the sensor (c) and reference ($\hat{c}$), bias is computed as:

146 $$\text{bias} = \frac{1}{n}\sum_{i=1}^{n}(c_i - \hat{c}_i) \qquad (S.3)$$

8

147 mean absolute error (MAE) is evaluated as:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|c_i - \hat{c}_i| \qquad (S.4)$$

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

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

151 These statistics assess, respectively, the systematic differences between the sensor and reference
152 measurements over time, the average absolute difference in measurements taken at the same
153 time, and the degree of linearity between the measurements. Lower absolute values of bias and
154 MAE denote better agreement, while a value of r close to 1 denotes stronger correlation.

155 Additionally, the following EPA bias and precision score metrics (Camalier et al.
156 2007)(Camalier et al., 2007) were used:

$$\text{Precision Score} = \sqrt{\frac{n\sum_{i=1}^{n}\delta_i^2 - \left(\sum_{i=1}^{n}\delta_i\right)^2}{n\chi_{0.1,n-1}^2}} \qquad (S.6)$$

158 where $\chi_{0.1,n-1}^2$ denotes the 10th percentile of the chi-squared distribution with $n-1$ degrees of
159 freedom, and:

$$\delta_i = 100\frac{c_i - \hat{c}_i}{\hat{c}_i} \qquad (S.7)$$

161 The bias score is:

$$\text{Bias Score} = \frac{1}{n}\sum_{i=1}^{n}|\delta_i| + \frac{t_{0.95,n-1}}{n}\sqrt{\frac{n\sum_{i=1}^{n}\delta_i^2 - \left(\sum_{i=1}^{n}|\delta_i|\right)^2}{n-1}} \qquad (S.8)$$

163 where $t_{0.95,n-1}$ is the 95th percentile of the t distribution with $n-1$ degrees of freedom. These
164 precision and bias scores can be compared to performance guidelines for various sensing
165 applications (Williams et al. 2014)(Williams et al., 2014). For PM$_{2.5}$, requirements for
166 educational monitoring (Tier I) are for precision and bias scores below 50%; for hotspot
167 identification and characterization (Tier II) or personal exposure monitoring (Tier IV), these
168 should be below 30%; for supplemental monitoring (Tier III), below 20%; and for regulatory
169 monitoring (Tier V), below 10%.

170 ## S.7.S.6. Seasonal Changes in PM$_{2.5}$ fraction below 300 nm in Pittsburgh

171 Aerosol size distributions over the 10-300 nm mobility size range were measured with a TSI
172 scanning mobility particle sizer (SMPS) at the CMU campus. PM$_{0.3}$ mass concentrations were
173 estimated assuming a mobility density of 1 gm/cm^3 and spherical particles, and then corrected to

174   the equivalent mass at 35% RH using the previously-discussed hygroscopic corrections. $PM_{2.5}$
175   mass concentrations were obtained from an NPM instrument attached to a RAMP co-located
176   with the SMPS. These values were corrected using Eq. (~~1~~3). For the winter months, the RAMP
177   RH was assumed to be the same as the conditions inside the SMPS. For the summer months, we
178   assumed that the SMPS RH was 15% higher (than the RAMP RH) inside the air-conditioned
179   trailer where the SMPS operated. The SMPS/NPM comparison is further complicated by the fact
180   that we are comparing an electrical mobility sizer to an optical sizer, but the overall result of
181   higher sub-300 nm aerosol mass is consistent with previously reported results. Stanier et al.
182   (2004) observed a larger aerosol volume in the 100-560 nm size range in the summer months
183   during the 2001-2002 Pittsburgh Air Quality Study. Saha et al. (2018) found that in 2016-2017,
184   though SO_2 concentrations have reduced compared to 2001-2002 resulting in fewer nucleation
185   events, the warmer months still see higher frequency of nucleation events and with higher
186   intensity compared to the winter months.
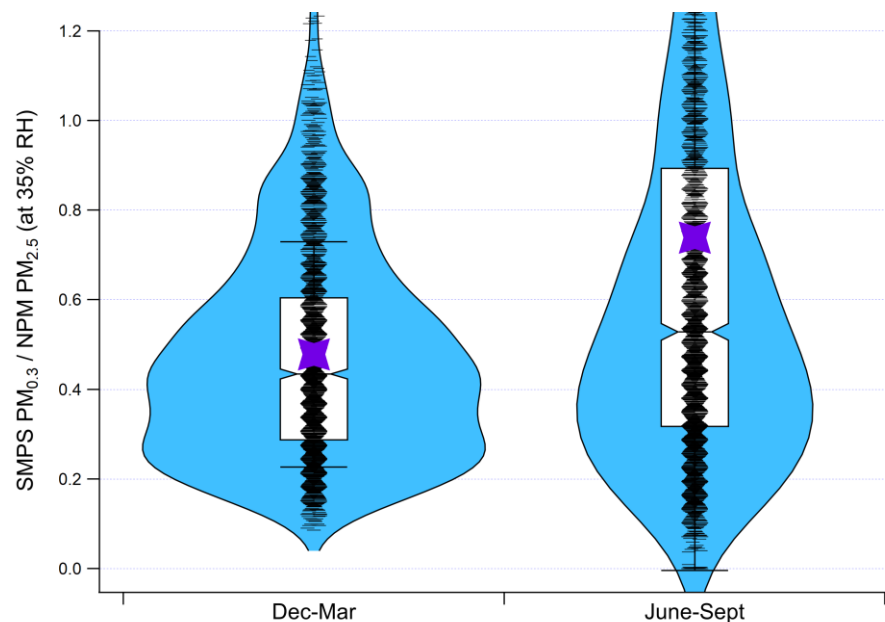
> **Formatted:** Subscript

187

188   Figure S.~~5~~8: Ratios of $PM_{0.3}$ to $PM_{2.5}$ based on summer and winter data collected in Pittsburgh.
189   Individual data points are jittered; means are shown by the purple stars; whiskers represent one
190   standard deviation of the data. Values greater than unity likely indicate data where our
191   assumptions are no longer valid, but these are <25% of the data. The median $PM_{0.3}/PM_{2.5}$ is 0.43
192   in the winter and 0.53 in the summer. For an annual average concentration of ~10 µg/m^3, this
193   represents a 1 µg/m^3 higher sub-300 nm fraction in the summer.

10

194 ## S.7.   Results for Correction Methods

195 Table S.2: Prior to the application of any corrections, this table presents the MAE, bias, and
196 correlation coefficients for the as-reported sensor data (the same data as shown in Figure 1)
197 broken down by relative humidity range.

RH	MET			PPA		
range	MAE	bias	r	MAE	bias	r
[%]	[μg/m^3]	[μg/m^3]		[μg/m^3]	[μg/m^3]	
30 - 35	6.0	-5.9	0.83	2.8	-0.91	0.70
35 - 40	7.1	-7.1	0.73	2.8	-0.85	0.78
40 - 45	6.2	-6.2	0.75	3.0	-0.25	0.71
45 - 50	5.5	-5.5	0.72	2.6	0.63	0.85
50 - 55	5.1	-5.1	0.67	3.3	1.2	0.74
55 - 60	5.2	-5.0	0.71	3.7	1.8	0.74
60 - 65	4.5	-4.2	0.77	3.4	1.6	0.87
65 - 70	3.8	-3.1	0.76	3.4	1.0	0.74
70 - 75	3.1	-2.1	0.80	5.2	3.5	0.75
75 - 80	3.9	-2.5	0.79	5.4	3.8	0.82
80 - 85	3.4	-0.6	0.85	6.2	4.7	0.89
85 - 90	5.4	2.5	0.87	7.9	6.1	0.95

198

199 Table S.3: This table presents the MAE, bias, and correlation coefficients for the sensor data
200 after correction with Eq. (3) (the same data as shown in Figure 4) broken down by relative
201 humidity range.

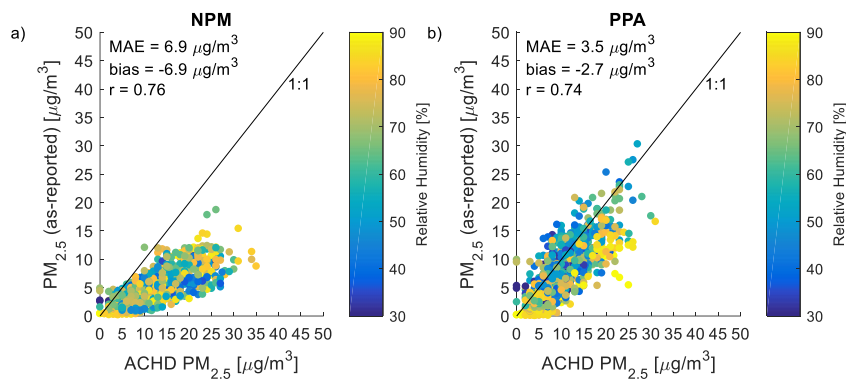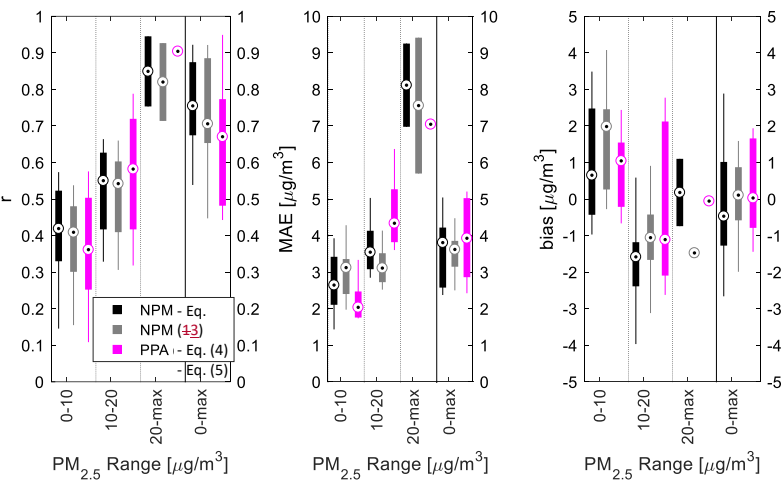| RH	MET			PPA		
range	MAE	bias	r	MAE	bias	r
[%]	[μg/m^3]	[μg/m^3]		[μg/m^3]	[μg/m^3]	
30 - 35	2.4	-0.6	0.81	2.2	0.45	0.71
35 - 40	3.3	-1.9	0.75	2.2	0.09	0.79
40 - 45	2.7	-1.2	0.77	2.3	0.43	0.72
45 - 50	2.7	-0.9	0.75	2.2	0.54	0.86
50 - 55	2.7	-0.4	0.72	2.5	0.53	0.75
55 - 60	3.0	-0.8	0.75	2.5	0.48	0.73
60 - 65	3.0	-0.3	0.78	2.1	0.33	0.86
65 - 70	2.8	0.5	0.76	2.0	0.22	0.75
70 - 75	2.7	0.9	0.80	2.6	0.49	0.76
75 - 80	3.0	-0.7	0.81	2.6	-0.36	0.79
80 - 85	2.8	-0.1	0.86	2.2	-0.34	0.85
85 - 90	2.9	-0.2	0.90	3.7	-2.7	0.92

202 ## S.8.

Figure S.96: Comparison of median one-hour-average NPM (a) and PPA (b) sensor readings to the BAM instrument during collocation at the Lawrenceville site after correction using a hygroscopic growth factor only (i.e. corrected measurement is raw measurement divide by fRH). Colors indicate relative humidity at the time of the measurements. Note that the NPM measurement corrected in this manner severely underestimates $PM_{2.5}$ concentration. For PPA sensors, while absolute errors are decreased relative to those of using the as-reported values directly, bias is also increased and correlation is reduced.

211

Table S.~~42~~: Coefficients for empirical correction equations

Coefficient	Value Estimate	Standard Deviation	Unit
α_0	0	2.9	$\mu g/m^3$
α_1	2.93	0.08	N/A
α_2	-0.11	0.08	$\mu g/{}^{\circ}Cm^3$
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}	${}^{\circ}C^{-1}$
α_6	-2.7×10^{-2}	0.11×10^{-2}	$\%^{-1}$
α_7	2.9×10^{-3}	0.8×10^{-3}	$\mu g/{}^{\circ}C^2m^3$
α_8	5.0×10^{-3}	1.0×10^{-3}	$\mu g/{}^{\circ}C\%m^3$
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	$\mu g/{}^{\circ}Cm^3$
β_3	-0.82	0.11	$\mu g/\%m^3$
β_4	2.9	0.53	$\mu g/{}^{\circ}Cm^3$
γ_0	21	2.1	$\mu g/m^3$
γ_1	0.43	0.013	N/A
γ_2	-0.58	0.090	$\mu g/{}^{\circ}Cm^3$
γ_3	-0.22	0.023	$\mu g/\%m^3$
γ_4	0.73	0.098	$\mu g/{}^{\circ}Cm^3$

212

213 The following figure summarizes the medians and ranges in performance of the corrected NPM
214 and PPA hourly averaged data across both collocation sites, using all sensors deployed to both
215 sites (as opposed to only the testing set), as well as specifying performance by different
216 concentration ranges (0 to 10, 10 to 20, and higher than 20 $\mu g/m^3$). Correlation is typically better
217 for NPM sensors (using either empirical correction equation), with r between 0.7 and 0.9, while
218 for PPA sensors it ranges down to 0.5. Correlations also improve at higher concentrations. The
219 MAE for both sensors are between 3 and 5 $\mu g/m^3$. MAE also tends to increase as concentrations
220 increase, but the PPA sensors appear to be less affected than NPM at concentrations above 20
221 $\mu g/m^3$; however, considering there were only two PPA sensors at the Lincoln site (where these
222 higher concentrations were more common) this may be a sample size artefact. Although unbiased
223 over the full range, the corrected sensor readings tend to be positively biased at low
224 concentrations and negatively biased at moderate concentrations. This is opposite to the trend
225 seen before correction and may be due to overcorrections at the extremes.

13

226

227

Figure S.10̶7̶: Comparison of one-hour-average corrected 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). Results
shown relate to a total of 32 NPM and 11 PPA sensors, and only consider sensors with at least
five samples in the relevant range.

The following figures illustrate how the performance of the proposed correction approaches is
affected if data from just one of the sites (Lincoln or Lawrenceville) is used to train the model,
and it is then tested on data from the other site.

14

Figure S.118: Comparison of sensor performance compared to the BAM instrument during collocation at the Lawrenceville site, using correction models calibrated using only data collected at the Lincoln site. Performance is comparable in terms of correlation and MAE to models trained using data from both sites, although bias, especially using Eq. (13) for NPM sensors, is generally worse.
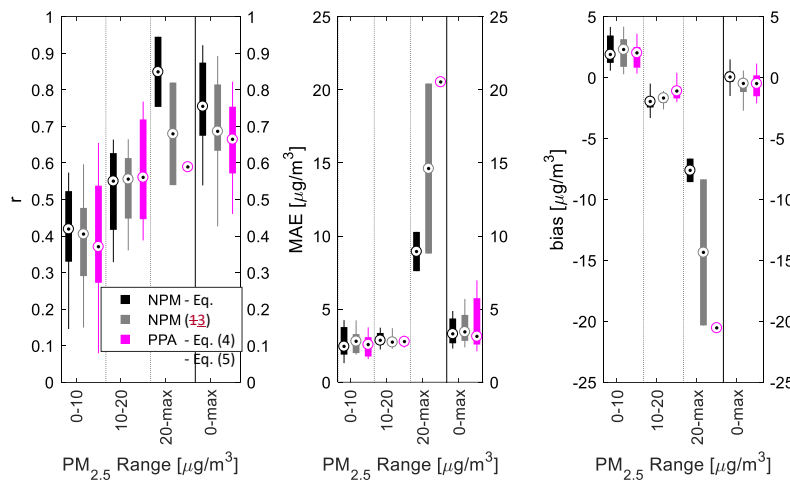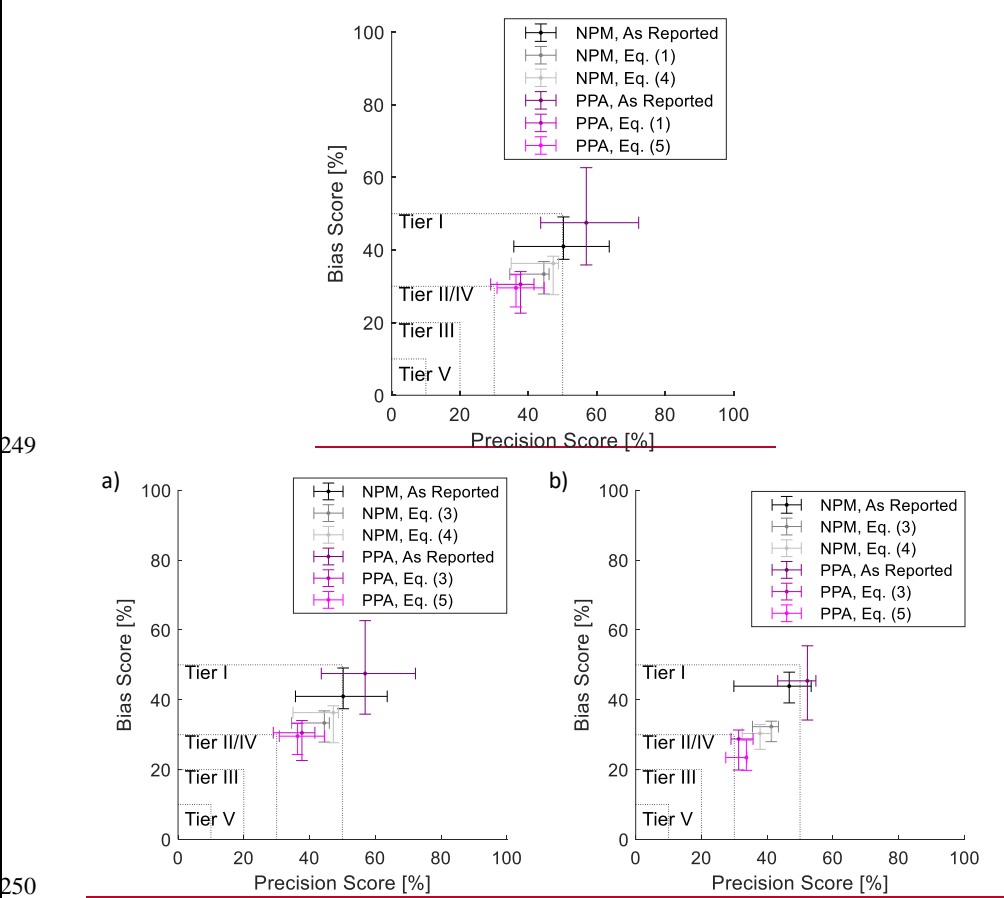

Figure S.12: Comparison of sensor performance compared to the BAM instrument during collocation at the Lincoln site, using correction models calibrated using only data collected at the Lawrenceville site. Performance is comparable except in the 20-max range, where performance

15

246 is significantly worse than for models calibrated using data from both sites. This illustrates the
247 importance of calibrating correction equations across the entire range of concentrations which
248   might be expected during field deployments.

Figure S.1310: Evaluation of EPA precision and bias score metrics for hourly-averaged (a) or
daily-averaged (b) data from NPM and PurpleAir sensors. Center-points of crosses indicate
median performance, with arms indicating 25%-75% range. Following corrections, both
instruments meet Tier I requirements for educational and informational monitoring.
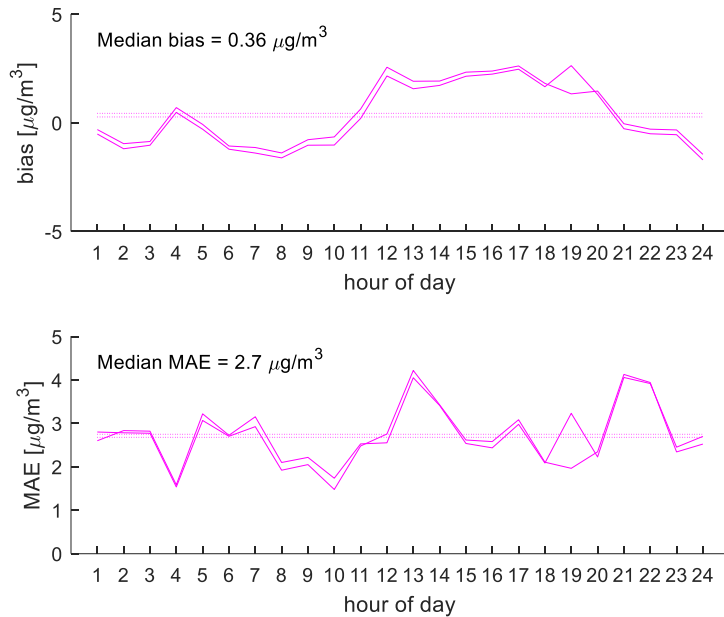
16

Figure S.1411: Results of a performance evaluation of a pair of PurpleAir sensors at the Parkway East site. Corrections are performed using Eq. (3). Results cover a data collection period of three weeks. Hourly-average bias and MAE are plotted as a function of time of day in the solid lines for the two sensors; dotted lines indicate the median performance throughout the day for each sensor. Median bias and MAE for both sensors are also listed in the figure. Corrections are performed using Eq. (1).
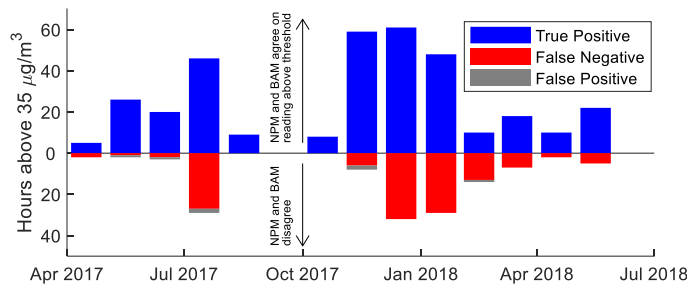
S.8.   Short-Term Performance Assessment

Figure S.15: Detection of hourly high PM$_{2.5}$ events by NPM sensor at Lincoln. True positives (correct detections) are counted for each hour on a monthly basis, along with false positives

267   (NPM falsely indicated high PM) and false negatives (NPM missed high PM), with a grace
268   period of ±1 hour.

270 Camalier L, Eberly S, Miller J, Papp M. 2007. Guideline on the Meaning and the Use of
271   Precision and Bias Data Required by 40 CFR Part 58 Appendix A.

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

276 Gu P, Li HZ, Ye Q, Robinson ES, Apte JS, Robinson AL, et al. 2018. Intra-city variability of PM
277   exposure is driven by carbonaceous sources and correlated with land use variables.
278   Environmental Science & Technology; doi:10.1021/acs.est.8b03833.

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

282 Johnson KK, Bergin MH, Russell AG, Hagler GSW. 2016. Using Low Cost Sensors to Measure
283   Ambient Particulate Matter Concentrations and On-Road Emissions Factors.
284   Atmospheric Measurement Techniques Discussions 1–22; doi:10.5194/amt-2015-331.

285 Petters MD, Kreidenweis SM. 2007. A single parameter representation of hygroscopic growth
286   and cloud condensation nucleus activity. Atmospheric Chemistry and Physics 7:1961–
287   1971; doi:10.5194/acp-7-1961-2007.

288 Saha PK, Robinson ES, Shah RU, Zimmerman N, Apte JS, Robinson AL, et al. 2018. Reduced
289   Ultrafine Particle Concentration in Urban Air: Changes in Nucleation and Anthropogenic
290   Emissions. Environmental Science & Technology 52:6798–6806;
291   doi:10.1021/acs.est.8b00910.

292 Saha PK, Zimmerman N, Malings C, Hauryliuk A, Li Z, Snell L, et al. 2019. Quantifying high-
293   resolution spatial variations and local source impacts of urban ultrafine particle
294   concentrations. Science of The Total Environment 655:473–481;
295   doi:10.1016/j.scitotenv.2018.11.197.

296 Schmidt M, Lipson H. 2009. Distilling Free-Form Natural Laws from Experimental Data.
297   Science 324:81–85; doi:10.1126/science.1165893.

298 Stanier CO, Khlystov AY, Pandis SN. 2004. Ambient aerosol size distributions and number
299   concentrations measured during the Pittsburgh Air Quality Study (PAQS). Atmospheric
300   Environment 38:3275–3284; doi:10.1016/j.atmosenv.2004.03.020.

Turpin BJ, Lim H-J. 2001. Species Contributions to PM2.5 Mass Concentrations: Revisiting Common Assumptions for Estimating Organic Mass. Aerosol Science and Technology 35:602–610; doi:10.1080/02786820152051454.

Williams R, Vasu Kilaru, Snyder E, Kaufman A, Dye T, Rutter A, et al. 2014. Air Sensor Guidebook.

Zheng T, Bergin MH, Johnson KK, Tripathi SN, Shirodkar S, Landis MS, et al. 2018. Field evaluation of low-cost particulate matter sensors in high- and low-concentration environments. Atmospheric Measurement Techniques 11:4823–4846; doi:10.5194/amt-11-4823-2018.