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

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

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

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, which must be accounted for to improve their in-field accuracy. We conduct long-term tests of two low-cost $PM_{2.5}$ 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 $\mu g/m^3$) 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 is reduced below 1 $\mu g/m^3$. 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.

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 environments, can have substantial benefits (Apte et al. 2015). Accurate monitoring of $PM_{2.5}$ is thus important for a variety of applications, including long-term health studies, assessing the impacts of technology and/or regulatory changes on emissions, and supporting decision-making for future regulatory efforts or to alter individual behavior in real-time. Monitoring is especially of interest in urban areas where the high density of exposed populations is coupled with higher

variability in particulate concentrations due to the large number and variety of sources (Jerrett et al. 2005; Karner et al. 2010; Eeftens et al. 2012); thus, a sparse monitoring network can lead to an incomplete understanding of $PM_{2.5}$ spatial variability and its subsequent health impacts. Recent advances in low-cost air quality sensing technologies have made it feasible for dense networks of monitors to be deployed in urban areas, providing a neighborhood-scale understanding of air pollution (Snyder et al. 2013). Several pilot programs for monitoring air quality at such high spatial resolution using these technologies are underway (Jiao et al. 2016; English et al. 2017; Williams et al. 2018; Zimmerman et al. 2018).

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

Assessments of these low-cost sensors must also account for different use-cases (Rai et al. 2017); we consider two in this work. First, sensors may be used, e.g. by community monitoring groups, to provide information on local air quality in real-time to support individual decisions, for example about where to go for a walk in a city to avoid highly polluted areas. In this case, exact quantitative results are less important than providing accurate indicators, e.g. that PM concentrations are currently higher in one part of a city than in another. Second, sensors may be used to determine long-term trends, e.g. for quantifying the exposure of a population or the impacts of a new

pollution-mitigation policy. In this case, quantitatively accurate long-term performance is important, while short-term performance is less so. Knowledge of the capabilities and limitations of these low-cost sensors with respect to these use-cases is especially relevant considering that products such as the PurpleAir sensor are already used by citizen scientists worldwide (www.purpleair.com).

In this paper, we provide evaluations of the long-term performance of two types of relatively low-cost (under $2000 for the NPM and $250 for the PurpleAir) $PM_{2.5}$ sensors in field conditions in the city of Pittsburgh, Pennsylvania and its surroundings. The ambient hourly $PM_{2.5}$ concentrations for this study are low (typically below 20 $\mu g/m^3$) compared to previous field evaluations of these sensors (e.g. Kelly et al. 2017; Jayaratne et al. 2018). We also propose and evaluate both physically-based and fully-empirical methods to correct for the influence of humidity and temperature on sensor readings, thereby making them more comparable to BAM instrument data. We have focused our attention on field studies due to the importance of assessing sensors in a similar environment to that in which they are to be used (White et al. 2012; Piedrahita et al. 2014). In Pittsburgh, like in other urban areas, $PM_{2.5}$ is composed of regionally transported (aged) aerosol and fresh vehicular emissions (Tan et al. 2014). Additionally, a metallurgical coke producing facility is a major local point source. Hence, we develop a calibration equation through collocation with a reference monitor at an urban background site that represents aged background PM and a source-oriented site near the major point source. We further evaluate these models across multiple seasons (January 2017 to May 2018) at both locations, as well as at a roadside location where vehicular contribution to $PM_{2.5}$ below the sensor detection limit should be highest.

2. Methods

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

The Real-time Affordable Multi-Pollutant (RAMP) monitor is a low-cost sensing system collaboratively developed by SenSevere and the Center for Atmospheric Particle Studies at Carnegie Mellon University (Zimmerman et al. 2018). It incorporates five gas sensors, electronics, batteries, and wireless communication hardware. In addition to its internal sensors, the RAMP can be connected to external instruments for measuring $PM_{2.5}$. One such instrument is the Met-One Neighborhood Particulate Monitor (NPM) sensor, which uses a forward light scattering laser. The unit is also equipped with an inlet heater and $PM_{2.5}$ cyclone. Previous research has assessed the performance of two of these instruments over a two-month period in southern California, and found only moderate correlations (R^2 between 0.5 and 0.7) with regulatory-grade instruments (AQ-SPEC 2015). The NPM is available for about $2000 or about one tenth the price of regulatory-grade instruments measuring $PM_{2.5}$. A total of 50 NPM units have been deployed alongside RAMPs.

The PurpleAir $PM_{2.5}$ monitor (PPA) is also employed along with the RAMPs. This sensor incorporates a pair of Plantower PMS 5003 laser sensors, which provide measures of $PM_{2.5}$ as well

as of $PM_{1.0}$ and PM_{10}. Previous testing of three of these units over a two-month period in southern California showed good correlation (R^2 above 0.9) with regulatory-grade instruments (AQ-SPEC 2017). This sensor is available for about $250, or about one hundredth of the price of a regulatory-grade instrument. A total of 20 PurpleAir units have been deployed with RAMPs.

2.2. Data Collection

Sensor performance was assessed using data collected at three field sites (one corresponding to an "urban background", one impacted by industrial emissions, and one by vehicle emissions) coincident with monitoring stations operated by the Allegheny County Health Department (ACHD), at which BAM instruments provided hourly concentration measurements for comparison (Hacker 2017). Although these instruments are not used for regulatory reporting, they are recognized federal equivalent methods and provide hourly data for Air Quality Index calculations.

The "Lincoln" site (AQS#42-003-7004), located at 40.308ºN by 79.869ºW, is a "source-dominated" site within 1 km of a facility producing coke for steel manufacturing that is the largest primary $PM_{2.5}$ point source in Allegheny county. This part of Allegheny County exceeded the annual Environmental Protection Agency (EPA) $PM_{2.5}$ standard over 2015-2017 (ACHD 2017). This site is illustrative of a "fence line" monitoring application, where monitors are placed in proximity to a known emission source. Average $PM_{2.5}$ concentration at this site was 14.5 $\mu g/m^3$ in 2017, with a one-hour maximum of 162 $\mu g/m^3$. Here, one NPM sensor was operated for a total of 294 days from April 24, 2017 until the end of data collection for this study on June 1, 2018. Additionally, between October 26, 2017 and February 12, 2018 (109 days), a total of 12 NPM and 2 PurpleAir sensors were collocated at the site (although not all instruments were active for the entire period); between October and February, temperature varied between -20 and +31°C and relative humidity varied from 22% to 97%.

The second deployment site, denoted as "Lawrenceville" (AQS#42-003-0008, 40.465ºN by 79.961ºW), is an urban background site located in an urban residential and commercial neighborhood, and part of the EPA's NCore monitoring network (Hacker 2017). Average $PM_{2.5}$ concentration at this site (based 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 this site, one NPM sensor was operated for a total of 380 days between January 13, 2017 and May 6, 2018. In addition, a total of 25 NPM and 9 PurpleAir sensors were collocated at the site between March 30, 2018 and June 4, 2018 (66 days, although again, not all instruments were present for the entire period); temperature varied from -3 to +43°C and humidity varied between 17% and 97% (both of which are measured by the RAMPs' onboard sensors). Five NPM sensors were collocated at both Lincoln and Lawrenceville at different times; none of the PurpleAir sensors were collocated at both sites. Finally, the "Parkway East" site (AQS#42-003-1376, 40.437ºN by 79.864ºW) represents a roadside location (Hacker 2017). Between September 6 and 27, 2018 (21 days), two PurpleAir sensors were collocated at this site. Data from this site were only used for assessing correction methods, not for developing them.

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

2.3. Hygroscopic Growth Correction Methods

Figure 1 compares the as-reported data from the NPM and PurpleAir sensors to the BAM instrument at the Lawrenceville site. There are sizeable discrepancies (up to 20 µg/m^3 in some cases) in the values, with humidity clearly having an effect. To allow for more direct comparability between these instruments, a method was sought to correct the readings of the low-cost sensors to better match those of the federal equivalent BAM instruments. As a starting point, the hygroscopic growth factor is the ratio of particulate mass at a given humidity and temperature to that at 22ºC and 35% relative humidity, and is calculated using the average measured summer and winter aerosol composition in the Pittsburgh area (Gu et al. 2018) and literature κ-values for the major non-refractory aerosol components sulfate, nitrate, ammonium, and organic matter (Cerully et al. 2015; Petters and Kreidenweis 2007).

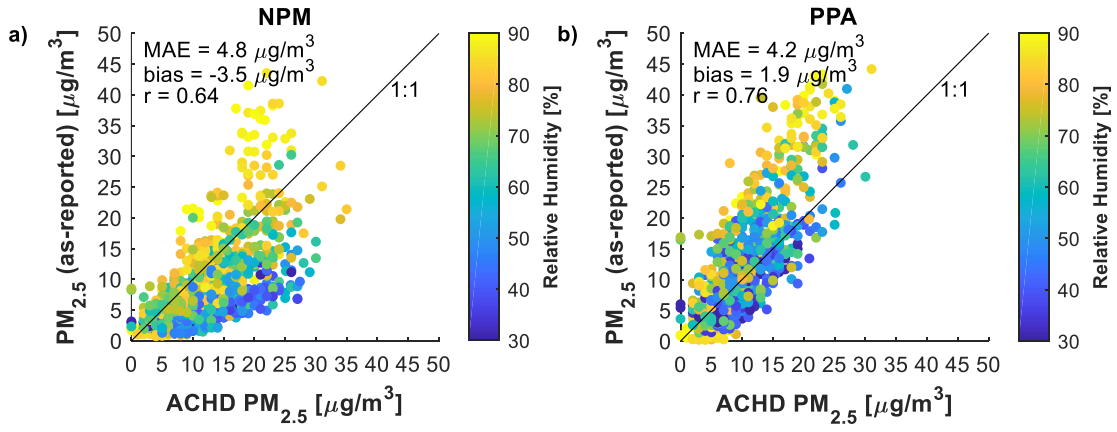

Figure 1: Comparison of one-hour-average NPM (a) and PurpleAir (b) as-reported sensor readings to the BAM instrument during collocation at the Lawrenceville site. Each point indicates the median across all sensors of the given type present at the site. Colors indicate relative humidity at the time of the measurements.

Correction of low-cost sensor readings using the hygroscopic growth factor alone was found to be insufficient (see supplementary materials). Therefore, the hygroscopic growth correction was combined with an additional linear correction:

$$[\text{corrected } PM_{2.5}] = \theta_1 \left(\frac{[\text{as-reported } PM_{2.5}]}{\text{fRH}(T,RH)}\right) + \theta_0 \quad (1)$$

The hygroscopic growth factor is calculated as follows:

$$\text{fRH}(T,RH) = 1 + \kappa_{\text{bulk}} \frac{a_w(T,RH)}{1-a_w(T,RH)} \quad (2)$$

where:

$$a_w(T,RH) = RH \exp\left(\frac{4\sigma_w M_w}{\rho_w RT D_p}\right)^{-1} \quad (3)$$

Details on the parameters of these equations are provided in the supplemental materials. The hygroscopicity of bulk aerosol (κ_{bulk}) is evaluated considering seasonal changes in particle composition observed in Pittsburgh; these are accounted for by dividing the year into summer (May to September inclusive), winter (November to March inclusive), and other periods (with the "other" period using an average of the summer and winter compositions). Within each period, it is assumed that the aerosol composition and size distribution are constant over time and throughout the urban area.

The coefficients θ_0 and θ_1 are estimated using a combination of data collected at both the urban-background Lawrenceville and source-oriented 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 the Parkway East site. Coefficients are set using typical linear regression techniques, minimizing the error between the corrected sensor measurements and the collocated BAM instrument at each site. These coefficients are estimated separately for the different time periods (summer, winter, other) for each of the low-cost sensor types (NPM, PurpleAir). This is necessary to account for the different responses of each type of sensor. For example, seasonal changes in particle size distributions will lead to changes in the θ_0 term as more or less of the particulate matter mass falls below the 300nm detection size threshold 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 more readily available, other more generalizable, empirical correction equations were developed using these data. Dewpoint (DP) was considered as a factor related to condensation that might serve in place of the hygroscopic growth factor; temperature (T) and relative humidity (RH) were also considered.

Various combinations of the as-reported sensor readings and the above environmental parameters were fit using linear and quadratic regression models to correct the data. For selecting an empirical

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; for NPM sensors, a quadratic function of the sensor reading, temperature, and humidity was thus selected:

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

The form selected for PurpleAir sensors was a two-piece linear function of the sensor reading, temperature, humidity, and dewpoint, with a threshold at 20 μg/m^3:

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

Coefficients calibrated for these equations (using standard regression techniques) along with their uncertainties are provided in the supplemental information.

2.5. In-field Drift-adjustment

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

The first adjustment method, known as the "Deployment Records" (DR) method, involves using a log of sensor deployment history to account for biases against a reference instrument. In this case, the relative bias of a deployed sensor versus a "benchmark" sensor collocated with a regulatory-grade instrument is determined by computing the relative difference in readings from these sensors for the last period during which they were collocated. The relative bias between this benchmark sensor and the collocated regulatory-grade instrument is also assessed. Then, bias of the deployed sensor to the regulatory-grade instrument is adjusted for, using the benchmark sensor as an intermediate step. The second method, known as the "Fifth Percentiles" (5P) method, involves computing the monthly 5th percentile of readings at a given deployment site, and then comparing to the 5th percentile recorded at the nearest regulatory monitoring station. Readings from the deployed sensor are then adjusted so that these percentiles match. This is done with the assumption that the 5th percentile represents a "background" level to which all sites in the region are subject. The third method is a variation of the 5P method, known as the "Average of Low readings" (AL) method, which uses the average of all readings in a month below 5 μg/m^3 as the target value to be matched. All three methods rely on the availability of relatively frequent (e.g. hourly) data from

regulatory-grade instruments, and the first method relies on historical collocation data with these instruments. Diagrams depicting each of these proposed methods are provided in the supplemental materials. The latter two methods of rectifying drift by matching distribution parameters over time are similar to those proposed by Moltchanov et al. (2015).

2.6. Assessment metrics

To evaluate the performance of a sensor as compared to a reference (typically a regulatory-grade instrument), the bias, mean absolute error, and correlation coefficient statistics are used (details are provided in the supplemental information). Performance of the instruments was also assessed from a classification perspective, using the EPA's National Ambient Air Quality Standards 24-hour standard of 35 μg/m^3 (www.epa.gov/criteria-air-pollutants/naaqs-table) as a representative threshold, by assessing how often the sensor agreed with a reference instrument as to whether this concentration was surpassed. This determination is made on an hourly basis for this assessment, while the regulation cited above applies to daily averages. This comparison is therefore conservative, and we would expect better performance for daily averages based on the results of Section 3.4. Classification precision indicates the fraction of values of concentration c above threshold τ detected by the sensor which were also detected by the reference:

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

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

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

Therefore, classification precision describes how often an event detected by the sensor actually occurred (assuming the reference instrument reading is the "true" concentration) while recall describes the fraction of actual events which were detected by the sensor. Values of these metrics 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. 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 are strong (typically $r > 0.9$) and are higher at the Lincoln site likely due to the higher concentrations. Absolute differences in readings were typically about 2 µg/m^3 or less, which includes systematic biases between sensors generally on the order of ±1 µg/m^3. These results indicate that, using as-reported information from sensors of the same type distributed over a city, robust qualitative conclusions can be made about the relative concentrations of $PM_{2.5}$ in different neighborhoods, as differences greater than about 2 µg/m^3 can be distinguished at one-hour resolution between neighborhoods in a city. This is consistent with prior results for another type (Alphasense OPC-N2, which is more than twice the price of a PurpleAir unit) of optical $PM_{2.5}$ sensor (Crilley et al. 2018).

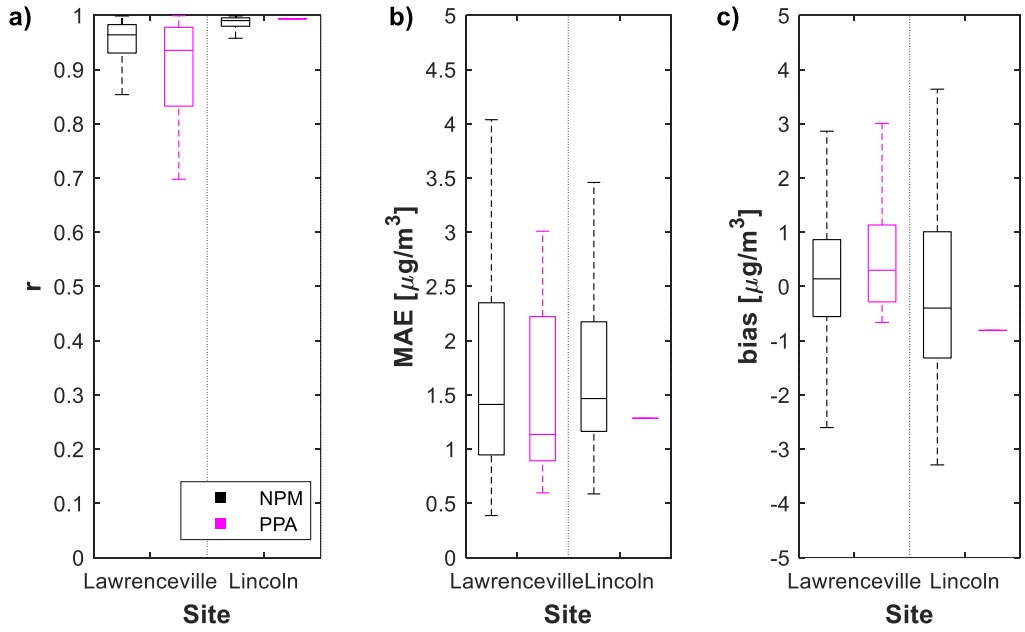

Figure 2: Inter-comparison of as-reported one-hour-average data between sensors during collocation periods at both sites. Black boxplots indicate metric ranges for pairs of NPM sensors, and purple boxplots indicate ranges for pairs of PurpleAir sensors. This represents 114 NPM pairs at Lawrenceville,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

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 are about twice that of the NPM, while at high humidity the ratio of readings approaches

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; this changes the water content and therefore the size and optical properties of the particles measured by the NPM as compared to those measured by the PurpleAir, which lacks such a heater. Second, these instruments, coming from different manufacturers, are calibrated differently, likely using calibration aerosols with different size distributions and compositions. They will therefore respond differently when exposed to a common aerosol which differs from their calibration aerosols.

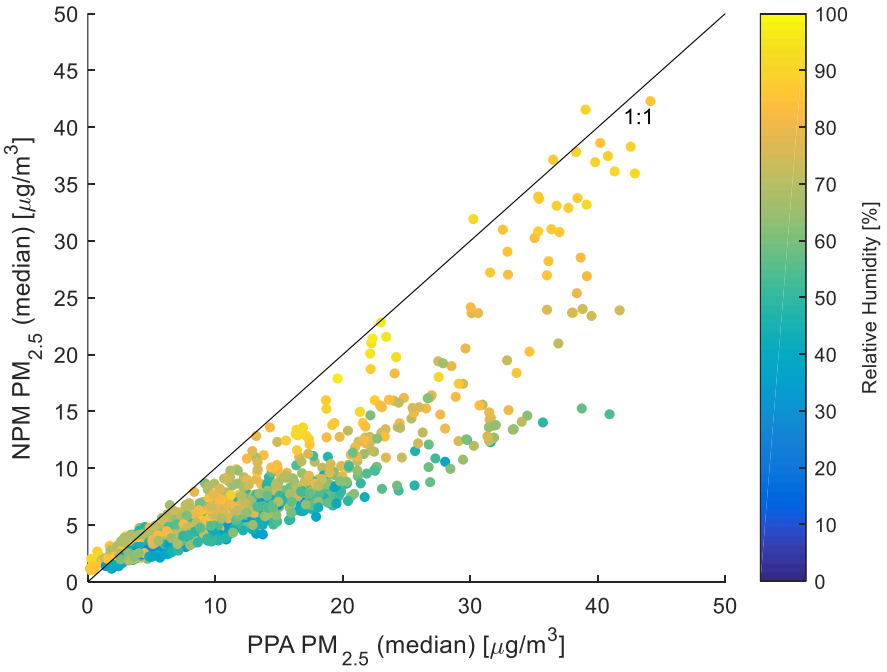

Figure 3: Comparison between medians of as-reported one-hour-average data of 25 NPM and 9 PurpleAir sensors during collocation at the Lawrenceville site. Colors indicate relative humidity at the time of the measurements.

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. (1) against the ACHD regulatory-grade (BAM) instrument readings. This correction decreases MAE by about 40% for both NPM and PurpleAir sensors with respect to their as-reported values and reduces bias significantly, but there is still noticeable measurement noise (r ~0.75) about the identity line.

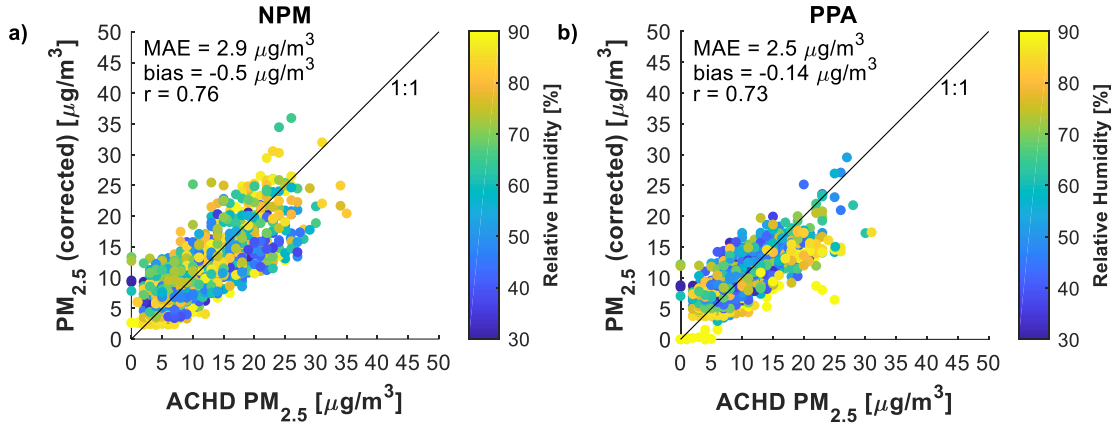

Figure 4: Comparison of one-hour-average NPM (a) and PurpleAir (b) sensor readings to the BAM instrument during collocation at the Lawrenceville site after correction using Eq. (1), with appropriate coefficients for NPM and PurpleAir. Each point indicates the median across all sensors of the given type present at the site (including both "training" and "testing" sensors). Colors indicate relative humidity at the time of the measurements.

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 physics-based approach of Eq. (1) (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 are used. In terms of correlation (Figure 5a), no improvement is made for PurpleAir sensors, while only a modest improvement results from correction of the NPM sensors. In terms of MAE (Figure 5b) and bias (Figure 5c), however, both correction approaches result in noticeable improvements. For NPM sensors, both the physically-based Eq. (1) and fully-empirical Eq. (4) give comparable performance. For PurpleAir sensors, the fully-empirical approach of Eq. (5) provides a smaller spread of MAE and bias results as compared to Eq. (1), while the median MAE of both approaches are almost the same, and the median bias of Eq. (5) is slightly worse. An independent assessment of the performance of PurpleAir sensors deployed at the Parkway East site and corrected using Eq. (1) provided comparable results (median R of 0.71, median MAE of 2.7 μg/m^3, median bias of 0.36 μg/m^3). Overall, while both correction approaches improve upon the as-reported data, there is no strong reason to prefer one correction method to the other based on their performance.

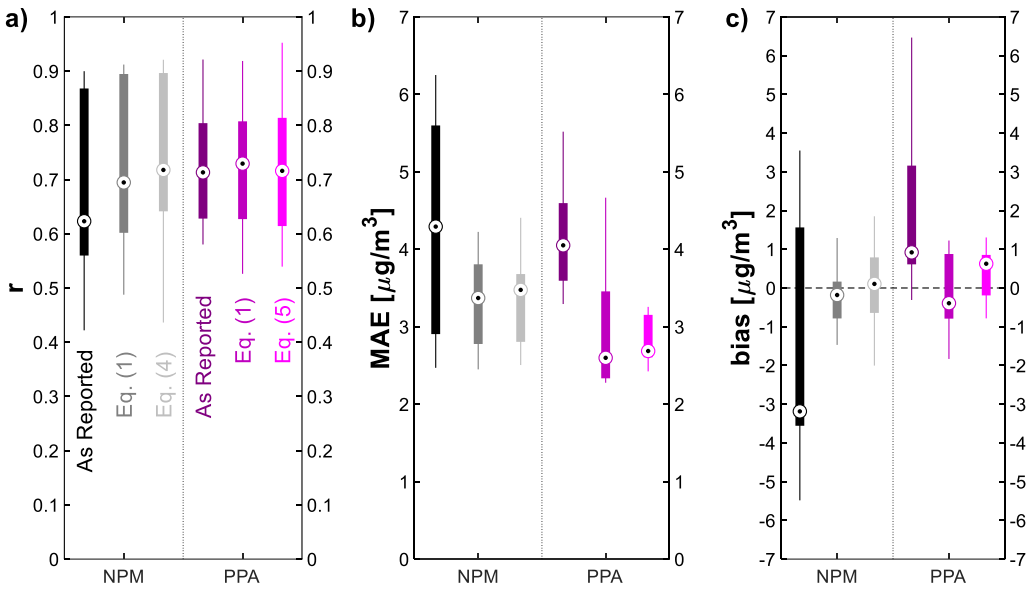

Figure 5: Performance metrics of one-hour-average as-reported and corrected sensor data compared to BAM instruments during collocation at both the Lawrenceville and Lincoln sites. Results shown relate to a total of 17 NPM and 5 PurpleAir sensors of the "testing" set. Corrections are performed using either the approach of Eq. (1), 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. (1) 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 supplementary materials). Note that for both NPM and PurpleAir sensors, the value of θ_0 (the linear intercept term) is 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 information), and thus will not be detected by these optical sensors, necessitating a larger correction. For θ_1 (the linear slope term), while the values for summer and winter are the same for NPM sensors, for PurpleAir sensors the value is higher in the winter. However, the hygroscopic growth factor (for the same temperature and relative humidity) is also higher in winter, as winter-time aerosol has a larger contribution from more hygroscopic inorganic aerosol. Thus, the net result is 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.

Table 1: Calibrated coefficients for Eq. (1). 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. Short-Term Performance

Figure 6 assesses the ability of the sensors to correctly identify times when $PM_{2.5}$ is higher than a threshold. The timeline charts the number of hours per week with average concentrations above 35 μg/m^3 identified by the NPM sensor (corrected using Eq. (1)) and the BAM instrument at the Lincoln site. Results are not reported for the Lawrenceville site since hourly concentrations there surpassed the threshold less than 1% of the time. True positives occurred when both instruments detected an event; false positives are when only the NPM measured the event, and false negatives when the NPM failed to detect an event seen by the BAM. The classification precision (Eq. (6)) of the sensor was 85% and its classification recall (Eq. (7)) was 71%; for comparison, these values are 61% and 78% respectively for the un-corrected, as-reported NPM data. Of the misclassifications, 15% occurred when the regulatory-grade instrument measured average concentrations between 30 and 40 μg/m^3; the rest represented larger discrepancies between the BAM and the NPM. A one hour "grace period" was also considered, i.e., if an event detection by one instrument leads or trails the other by up to an hour, this was still counted as a true positive. With this grace period, the classification precision was 90% and classification recall was 97% (versus 73% and 97% respectively for the un-corrected data).

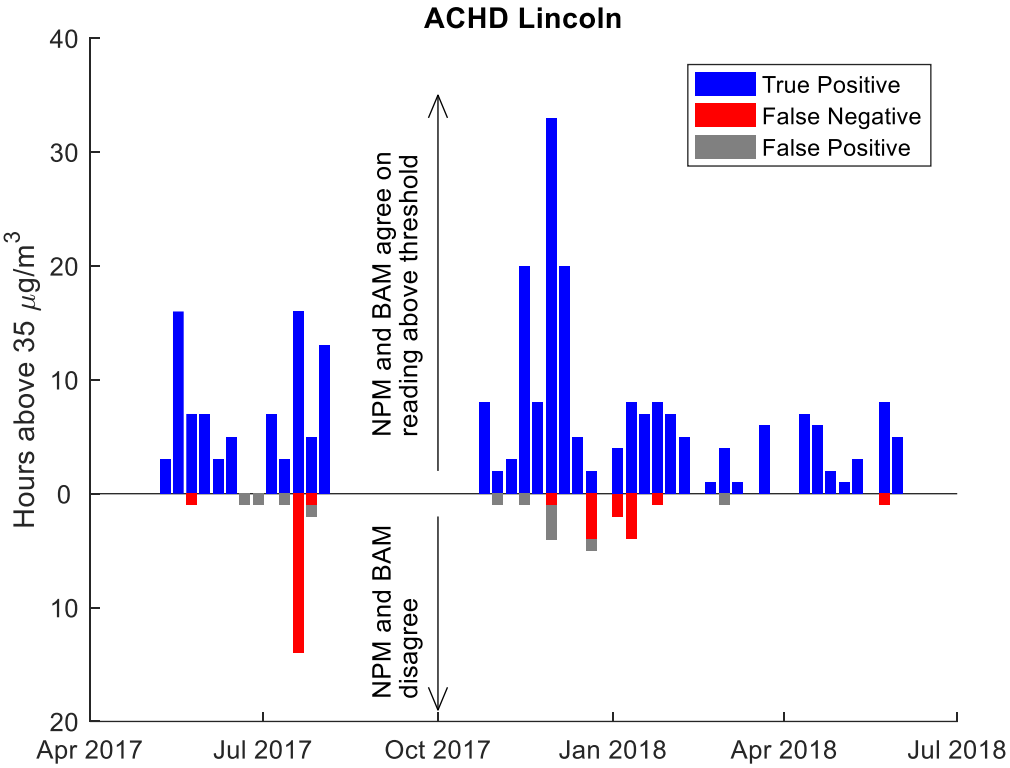

Figure 6: Detection of hourly high $PM_{2.5}$ events by NPM sensor at Lincoln. True positives (correct detections) are counted for each hour on a weekly basis, along with false positives (NPM falsely indicated high PM) and false negatives (NPM missed high PM), with a grace period of ±1 hour.

3.4. Long-Term Performance

Long-term assessment is necessary to categorize bias and assess data quality after extensive field use of sensors. Previous studies of lower-cost optical particle counters operating for up to four months report no evidence of significant drift (Crilley et al. 2018). The long-term performance of NPM sensors was assessed using data collected by the two sensors deployed at the Lawrenceville and Lincoln sites for a more extended period (e.g. more than a year of data at Lawrenceville collected over a 16-month span). First, data corrected using Eq. (4) were used to assess the in-field drift-adjustment methods proposed in Section 2.5 to eliminate the "random walk" behavior observed when this correction approach is used over long periods. Figure 7 shows the spread in monthly biases (difference between the monthly average readings of the corrected sensors and the BAM instruments) for both sites, both without drift-adjustment and with the three proposed drift-adjustment methods. Note that these biases are for the single long-term-deployment sensor at each site, whereas Figure 5 presents results for the entire "testing" set of sensors over a shorter period. Based on these results, the "average of low readings" method is best, reducing both the median

bias and spread in biases at the Lawrenceville site. However, there are no clear improvements at the Lincoln site.

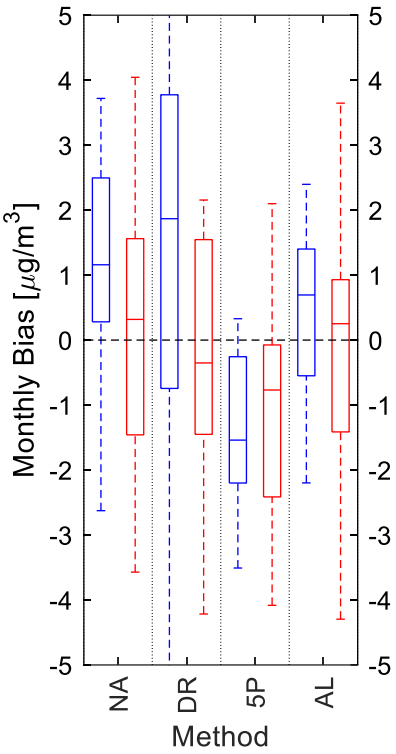

Figure 7: Performance of various drift-adjustment methods in reducing the bias in monthly averages; NA – no adjustment applied; DR – drift-adjusted using deployment records; 5P – drift-adjusted using percentiles of nearest reference site; AL – drift-adjusted using averages of low readings at nearest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

Figure 8 plots the MAE of the corrected sensor data with and without drift-adjustment (using the AL method) compared to the associated regulatory-grade instrument, as a function of averaging period. For weekly averages error is below 2 µg/m^3. For quarterly or longer averages, errors are about or below 1 µg/m^3, which is about 10% of the annual average concentrations for Pittsburgh. Drift-adjustment of measurements corrected with the fully-empirical Eq. (4) improves the performance at the Lawrenceville site (where concentrations are typically lower) to match that of Eq. (1). At the Lincoln site the drift-adjustment method tends to slightly increase errors; this indicates that drift adjustment may not be required (or even suitable) for all locations.

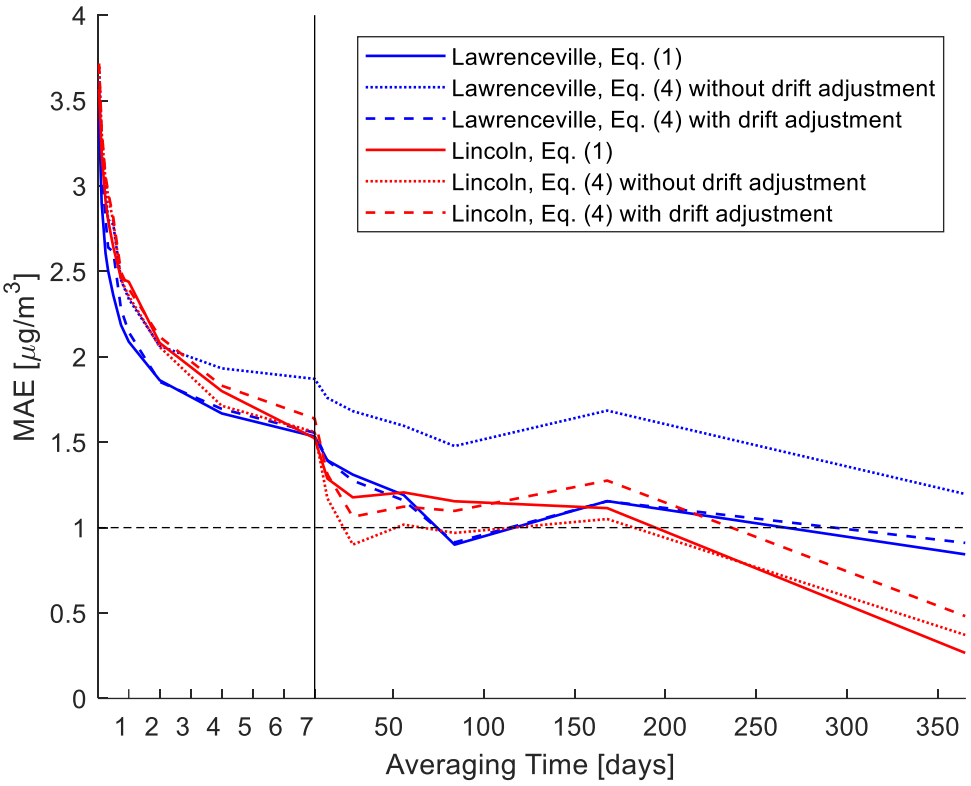

Figure 8: Mean absolute error in $PM_{2.5}$ measurements for two NPM sensors during long-term deployments as a function of averaging period (note the differing horizontal axis scale on either side of the vertical black line). Solid lines represent measurements corrected using Eq. (1); dotted lines indicate measures corrected using Eq. (4) but not drift-adjusted; dashed lines indicate measures corrected using Eq. (4) and drift-adjusted using the AL method.

4. Discussion

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

The first proposed correction equation is designed to account for two of the main factors contributing to differences between optical measurements and the BAM instrument readings. First, a hygroscopic growth factor is used to account for the increase in measured particle mass due to ambient humidity. Second, a linear correction is applied to account for the different optical response characteristics of different sensors, as related to mismatches between the size distribution

and chemical composition of the factory calibration aerosol and the (ambient) aerosol to be measured. We also evaluated alternative empirical correction equations which did not rely on the assumptions necessary for estimating hygroscopic growth. For both NPM and PurpleAir sensors, both correction approaches achieve similar performance, although even following correction, relatively large differences in hourly averages (MAE of 3 to 4 $\mu g/m^3$) are observed with respect to the BAM regulatory-grade instruments. This lack of consistency with BAM instruments has also been observed previously (e.g. Zheng et al. 2018) and may not be reconcilable with low-cost optical sensors. However, as data are averaged over longer periods, accuracy is improved, such that long-term (1 year or more) averages are likely to have errors below 1 $\mu g/m^3$.

The efficacy of several proposed in-field drift-adjustment methods are also evaluated, although these methods were only seen to have noticeable benefits at Lawrenceville, and so their utility is still unclear. One possible explanation is that the higher $PM_{2.5}$ at the source-dominated Lincoln site may still be largely composed of less-hygroscopic organic compounds, and so the fully-empirical correction, which does not account for seasonal changes, works well here. However, the same non-seasonally-adjusted empirical approach needs seasonal drift adjustments at the more regionally-dominated (including more hygroscopic inorganic PM fraction) Lawrenceville site.

The NPM (with or without correction) detected 97% of occurrences when the BAM recorded $PM_{2.5}$ higher than 35 $\mu g/m^3$. However, in the uncorrected data, 27% of such high values were not observed in the BAM; the corrections presented here (Eq. (1)), which account for aerosol hygroscopic growth, reduce this error to 10%. This indicates the potential for this sensor to be used to identify pollution hotspots after accounting for humidity effects.

In terms of use cases, the high level of mutual consistency and ability (with suitable corrections) to provide accurate long-term averages makes these low-cost sensors useful for large-scale mapping campaigns to determine long-term spatial patterns and temporal trends in $PM_{2.5}$. For real-time monitoring, although these sensors can detect hourly "spikes" reasonably well, concentration values are only accurate within about ± 3 $\mu g/m^3$. Nevertheless, this is sufficient for qualitative indications of relative short-term air quality, as indicated by the high concentration detection performance (Section 3.3). The small size and ease of deployment of these units make them well suited to urban monitoring. The low-cost (sub-$250 each) PurpleAir sensors also incorporate a pair of optical sensors, allowing for internal self-consistency checks to flag possible erroneous data. The cyclone and inlet heater of the (sub-$2,000 each) NPM sensors can protect the units from excessive dust and humidity (to which PurpleAir sensors, which lack these features, may be more susceptible during longer deployments). Finally, we note that while these results are determined for the specific environment of Pittsburgh, Pennsylvania, we believe they will generalize to other areas which are characterized by annual $PM_{2.5}$ mass concentrations less than 20 $\mu g/m^3$, which covers about half the world's population (Apte et al. 2015), and across both urban background (e.g. Lawrenceville) and source-impacted (e.g. Lincoln) sites.

Considering future low-cost $PM_{2.5}$ sensor deployments, it is recommended that sensor data be used as-is when only qualitative or comparative information is of interest within a region with similar weather conditions, and only when the same type of low-cost sensor is being used throughout, as uncorrected data from different sensors may not be comparable. For situations where more accurate quantitative data are needed, the use of correction Eq. (1) is recommended where detailed information on particle composition is available; otherwise, Eqs. (4) or (5) can be used. Regardless, coefficients should be determined via collocation with regulatory-grade instruments to account for local conditions, and depending on local conditions, further drift adjustments using the techniques presented here (or others) may be necessary.

Acknowledgements

Funding for this study was provided by the Environmental Protection Agency (Assistance Agreement Nos. RD83587301 and 83628601), and the Heinz Endowment Fund (Grants E2375 and E3145). The authors would like to thank Eric Lipsky, Naomi Zimmerman, and S. Rose Eilenberg for assistance with instrument setup and operation, as well as Ellis Robinson for providing chemical composition data from the AMS. Finally, the authors would like to thank Spyros Pandis for helpful discussions.

References

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

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

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

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

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

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

Burkart J, Steiner G, Reischl G, Moshammer H, Neuberger M, Hitzenberger R. 2010. Characterizing the performance of two optical particle counters (Grimm OPC1.108 and OPC1.109) under urban aerosol conditions. Journal of Aerosol Science 41:953–962; doi:10.1016/j.jaerosci.2010.07.007.

Cabada JC, Khlystov A, Wittig AE, Pilinis C, Pandis SN. 2004. Light scattering by fine particles during the Pittsburgh Air Quality Study: Measurements and modeling. Journal of Geophysical Research 109; doi:10.1029/2003JD004155.

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

Chow JC, Doraiswamy P, Watson JG, Chen L-WA, Ho SSH, Sodeman DA. 2008. Advances in Integrated and Continuous Measurements for Particle Mass and Chemical Composition. Journal of the Air & Waste Management Association 58:141–163; doi:10.3155/1047-3289.58.2.141.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Rai AC, Kumar P, Pilla F, Skouloudis AN, Di Sabatino S, Ratti C, et al. 2017. End-user perspective of low-cost sensors for outdoor air pollution monitoring. Science of The Total Environment 607–608:691–705; doi:10.1016/j.scitotenv.2017.06.266.

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

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

Solomon PA, Sioutas C. 2008. Continuous and Semicontinuous Monitoring Techniques for Particulate Matter Mass and Chemical Components: A Synthesis of Findings from EPA's Particulate Matter Supersites Program and Related Studies. Journal of the Air & Waste Management Association 58:164–195; doi:10.3155/1047-3289.58.2.164.

Tan Y, Lipsky EM, Saleh R, Robinson AL, Presto AA. 2014. Characterizing the Spatial Variation of Air Pollutants and the Contributions of High Emitting Vehicles in Pittsburgh, PA. Environmental Science & Technology 48:14186–14194; doi:10.1021/es5034074.

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

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

Watson JG, Chow JC, Moosmüller H, Green M, Frank N, Pitchford M. 1998. Guidance for using continuous monitors in PM2.5 monitoring networks.

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

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

Wilson WE, Chow JC, Claiborn C, Fusheng W, Engelbrecht J, Watson JG. 2002. Monitoring of particulate matter outdoors. Chemosphere 49: 1009–1043.

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 Discussions 1–40; doi:10.5194/amt-2018-111.

Zhou Y, Zheng H. 2016. PMS5003 series data manual.

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

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

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