Supporting Information

Laboratory evaluation of low-cost PurpleAir PM monitors and in-field correction using co-located portable filter samplers

Jessica Tryner[a,b], Christian L'Orange[a], John Mehaffy[a], Daniel Miller-Lionberg[b], Josephine C. Hofstetter[b], Ander Wilson[c], and John Volckens[a]

[a] *Department of Mechanical Engineering, Colorado State University, Fort Collins, Colorado, USA*
[b] *Access Sensor Technologies, Fort Collins, Colorado, USA*
[c] *Department of Statistics, Colorado State University, Fort Collins, Colorado, USA*

S1 Methods

S1.1 Laboratory Evaluations

During each laboratory evaluation, seven unique steady-state concentrations of NIST Urban Particulate Matter were generated inside the aerosol chamber (Figure S1). The concentration in the chamber was changed approximately once per hour, and each experiment lasted approximately 10 hours. A zero point was repeated at the start and end of the experiment. For each 1-hour long concentration point, the transient TEOM and PurpleAir data recorded during the first 15 minutes (before the PM concentration in the chamber had reached a steady-state value) were ignored. The steady-state PM concentrations recorded by each instrument during the remaining ∼45 minutes were then time-averaged over the entire segment.

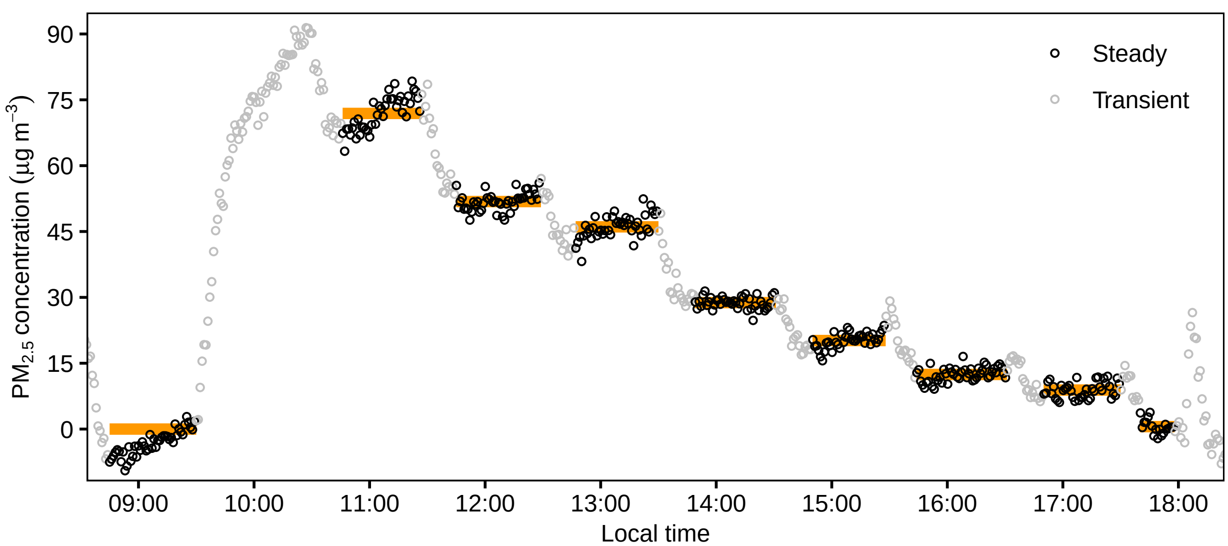

Figure S1. Concentrations recorded at 60-s intervals by the TEOM during the pre-deployment laboratory evaluation. Gray and black markers denote transient and steady-state data points, respectively. Gold lines represent the time-averaged concentrations calculated from the steady-state data points. Time-averaged concentrations below zero were replaced with zero.

During the third laboratory evaluation, the NIST Urban PM particle size distribution was measured using a scanning mobility particle sizer (SMPS) (Model 3082 Electrostatic Classifier and Model 2787 Condensation Particle Counter, TSI Incorporated, Shoreview, MN, USA) and an aerodynamic particle sizer (APS) (3321, TSI Incorporated, Shoreview, MN, USA). Particles passed through a Polonium-210 aerosol charge neutralizer before reaching the SMPS inlet. The SMPS reported data for particle diameters between 13.6 and 527.5 nm. The APS reported data for particle diameters between 542 and 18,430 nm. SMPS and APS data were collected for the entire duration of the laboratory evaluation. Each individual SMPS and APS scan lasted 120 seconds.

S1.2 Break-in Period

All 19 PurpleAir monitors were co-located on a single roof in Fort Collins, CO from 22 August through 12 October 2018 (Figure S2). The goal of this initial co-location was to "break-in" the monitors so that the performance captured during network field deployment would be representative of monitors that had been deployed outside for several weeks, as opposed to brand-new monitors.

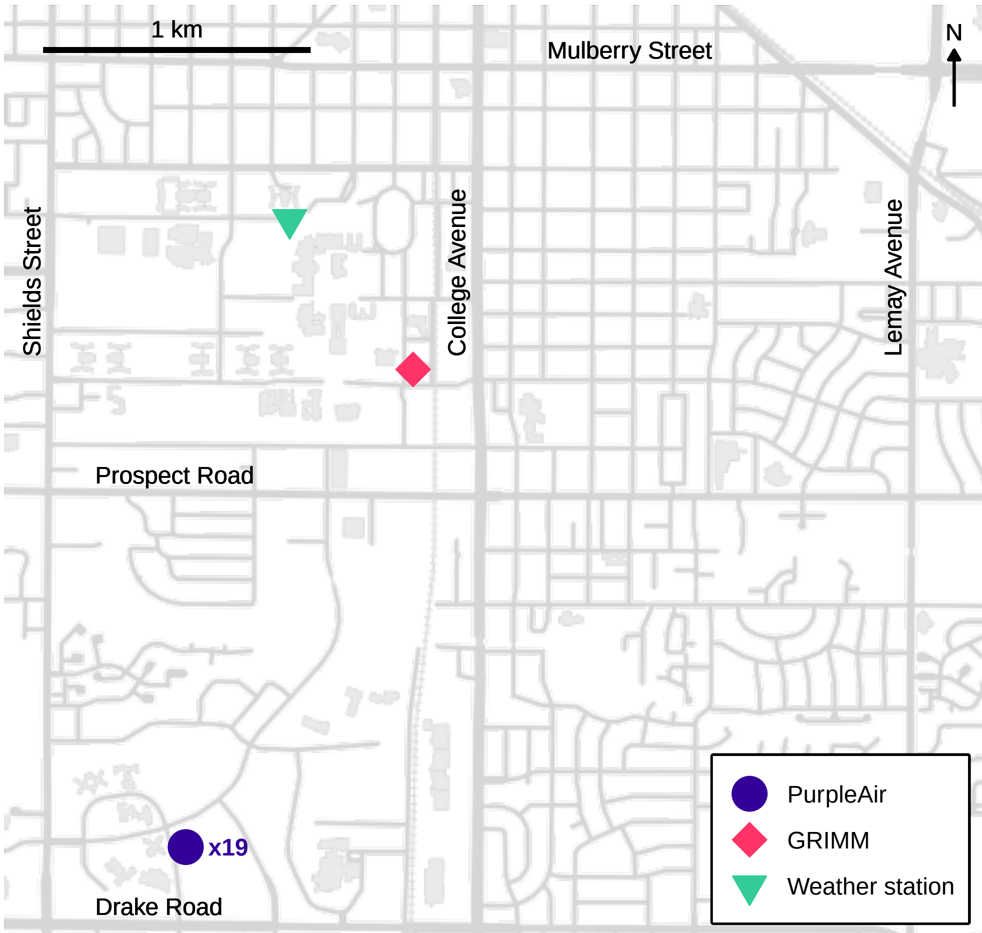

Figure S2. A map illustrating the location of the 19 PurpleAir monitors, relative to the GRIMM EDM 180 and the weather station, during the break-in period that took place between 22 August and 12 October 2018.

S1.3 Network Deployment

During the network field deployment, duplicate ASPEN boxes were co-located with a conventional 16.7 $\mathrm{L\,min^{-1}}$ $PM_{2.5}$ filter sampler at sites A-D (see Figure S3).

Figure S3. Two ASPEN boxes and one 16.7 $\mathrm{L\,min^{-1}}$ filter sampler co-located at site C.

We estimated the limit of detection (LOD) for "$PM_{2.5}$ ATM" concentrations reported by the PurpleAir monitors deployed in the field as:

$$LOD = \lceil 3s_{blank} \rceil \tag{S1}$$

where s_{blank} was the standard deviation of 131 hourly "$PM_{2.5}$ ATM" concentrations reported by two PurpleAir monitors co-located with the GRIMM EDM 180 during the 68 hours during network field deployment when the GRIMM EDM 180 reported a concentration below 1 $\mu\mathrm{g\,m^{-3}}$. The value of $3s_{blank}$ (0.31 $\mu\mathrm{g\,m^{-3}}$) was rounded up to the nearest integer because the Plantower PMS5003 sensor used in the PurpleAir monitor only reports $PM_{2.5}$ concentrations as integer values (non-integer concentrations reported in PurpleAir log files result from time averaging). Using this procedure, we estimated the LOD for hourly "$PM_{2.5}$ ATM" concentrations reported by PurpleAir monitors as 1 $\mu\mathrm{g\,m^{-3}}$. Our estimated LOD was lower than the LOD of 3 to 6 $\mu\mathrm{g\,m^{-3}}$ estimated by Sayahi et al. (2019) for PurpleAir monitors using PMS5003 sensors.

UPAS filter samples were used to generate monthly, weekly, and concurrent gravimetric correction factors for the PurpleAir monitors. These correction factors (CF) were calculated as shown in Equation S2 with c_{filter} equal to the $PM_{2.5}$ concentration derived from the UPAS filter sample and $\bar{c}_{PA}$ equal to the sample-averaged $PM_{2.5}$ ATM concentration reported by the paired PurpleAir (both in $\mu\mathrm{g\,m^{-3}}$).

$$CF = \frac{c_{filter}}{\bar{c}_{PA}} \tag{S2}$$

The correction factor calculated for a given paired PurpleAir monitor and UPAS filter sample was applied only to that PurpleAir monitor.

Monthly correction factors were calculated using UPAS filter samples that ended on 01 November and 29 November, respectively, and extended prospectively in time. Correction factors calculated from paired samples ending on 01 November were used to correct PurpleAir data collected between 02 and 29 November. Correction factors calculated from paired samples ending on 29 November were used to correct PurpleAir data collected between 30 November and 06 December (see Figure S4). A sensitivity study was conducted to investigate how results were affected by excluding data collected between 30 November and 06 December, since the monthly correction factor calculated from samples ending on 29 November was only applied to a single week at the end of the field deployment (instead of an entire month).

Weekly correction factors were calculated using UPAS filter samples that ended every Thursday between 01 and 29 November (except 22 November). The samples that ended each Thursday were used to calculate correction factors for PurpleAir data collected over the following 7 days. UPAS filter samples that ended on Wednesday 21 November were used to calculate correction factors for PurpleAir data collected between 23 and 29 November (see Figure S4).

Concurrent correction factors were calculated using all UPAS filter samples. For example, filter samples collected between 02 and 05 November were used to calculate correction factors for PurpleAir data collected between 02 and 05 November (see Figure S4).

To estimate the fraction of the variance in the log-transformed concurrent gravimetric correction factor that was explained by (a) between-date differences and (b) between-monitor differences, the two-way random effects model shown in Equation 8 was fit to the full dataset (n=175). Using this model, the ICC for sample date was defined as shown in Equation S3 and the ICC for PurpleAir monitor was defined as shown in Equation S4 (Lüdecke, 2019; McGraw and Wong, 1996). The 95% confidence intervals for these ICCs were calculated using parametric bootstrap resampling (n=1000).

$$ICC = \frac{\sigma_r^2}{\sigma_r^2 + \sigma_c^2 + \sigma_e^2} \tag{S3}$$

$$ICC = \frac{\sigma_c^2}{\sigma_r^2 + \sigma_c^2 + \sigma_e^2} \tag{S4}$$

Because two of the residuals associated with the model fit to the full dataset were extreme outliers, the model shown in Equation 8 was also fit to a dataset with those two data points removed (n=173). The two data points that were removed corresponded to the highest and lowest correction factors in the full dataset. The ICCs defined in Equations S3 and S4 were then recalculated to assess whether or not the ICCs were sensitive to the outliers.

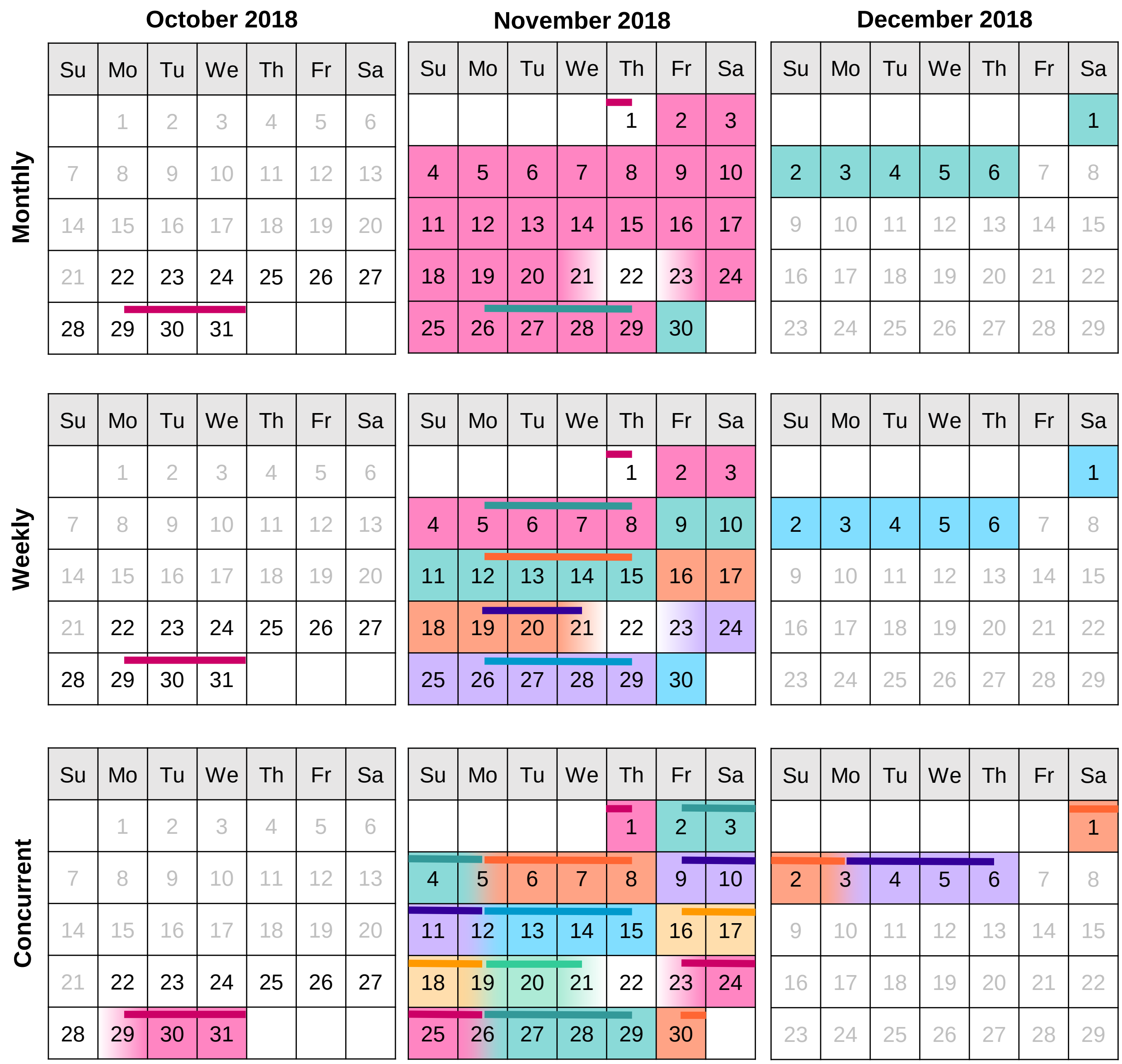

Figure S4. A graphical representation of the schedule used to develop monthly, weekly, and concurrent correction factors. Each bar represents a set of samples that was used to generate correction factors. Correction factors generated from a given set of samples were applied on dates shaded the same color. Samples that started on 19 November only ran for 48 hours due to the holiday on 22 November; all other samples ran for 72 hours.

S2 Results

S2.1 Laboratory Evaluations

At concentrations below 30 $\mu g\,m^{-3}$, the $PM_{2.5}$ CF=1 and $PM_{2.5}$ ATM values were equal (Figure S5).

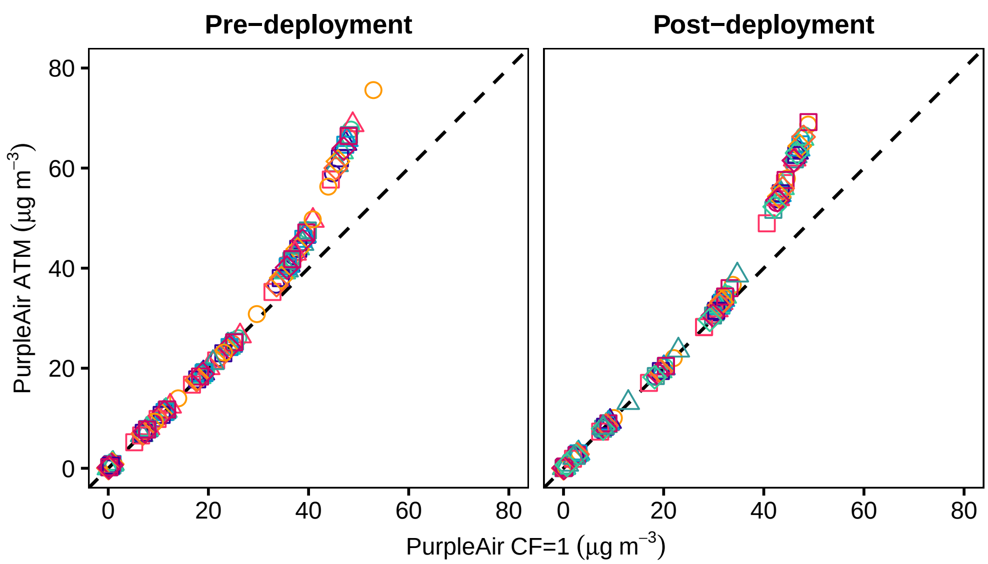

Figure S5. A comparison of the values labeled "$PM_{2.5}$ CF=1" and "$PM_{2.5}$ ATM" in the PurpleAir log files. Each data point corresponds to one of the time-averaged $PM_{2.5}$ concentrations shown in Figure 3.

Both before and after the field deployment, a linear model (Equation 3) explained 99% of the variance in the PurpleAir-reported "$PM_{2.5}$ ATM" response; however, the F-test identified a significant lack of fit between the model and the data (Pre-deployment: $F_{(I-2)/(IJ-I)} = 3.09$, $F_c = 2.17$; Post-deployment: $F_{(I-2)/(IJ-I)} = 42.4$, $F_c = 2.16$). The weighted residuals are shown in Figure S6.

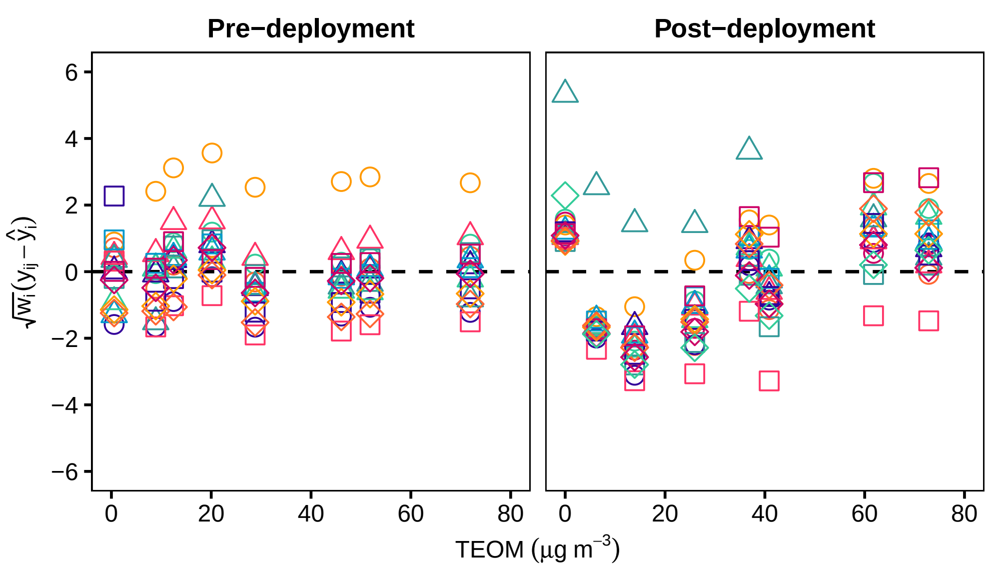

Figure S6. Weighted residuals for the linear calibration function fit to the time-averaged PM concentrations measured by the TEOM and the time-averaged "$PM_{2.5}$ ATM" concentrations reported by the PurpleAir monitors during the laboratory evaluations conducted before (left) and after (right) the field deployment.

S2.1.1 "CF=1" and "ATM" values: Plantower sensor output versus PurpleAir log data

In Figure S7, the raw $PM_{2.5}$ CF=1 and $PM_{2.5}$ ATM concentrations output by Plantower PMS5003 sensors are compared to the concentrations labeled "$PM_{2.5}$ CF=1" and "$PM_{2.5}$ ATM" in the PurpleAir log files. The $PM_{2.5}$ CF=1 concentrations output by the PMS5003 sensors and the concentrations labeled "$PM_{2.5}$ ATM" in the PurpleAir log files both appeared to increase linearly with the TEOM-reported concentration of NIST Urban PM, while the $PM_{2.5}$ ATM concentrations output by the PMS5003 sensors and the concentrations labeled "$PM_{2.5}$ CF=1" in the PurpleAir log files increased nonlinearly at concentrations above 30 $\mu g\,m^{-3}$. These results indicate that the $PM_{2.5}$ CF=1 values output by the PMS5003 sensors were labeled "$PM_{2.5}$ ATM" in the PurpleAir log files. Similarly, the $PM_{2.5}$ ATM values output by the PMS5003 sensors were labeled "$PM_{2.5}$ CF=1" in the PurpleAir log files (PurpleAir firmware version 2.50i). Concentrations referred to as "$PM_{2.5}$ ATM" throughout this manuscript refer to values labeled"$PM_{2.5}$ ATM" in the PurpleAir log files. In other words, concentrations referred to as "$PM_{2.5}$ ATM" throughout this manuscript actually refer to the $PM_{2.5}$ CF=1 values output by Plantower PMS5003 sensors.

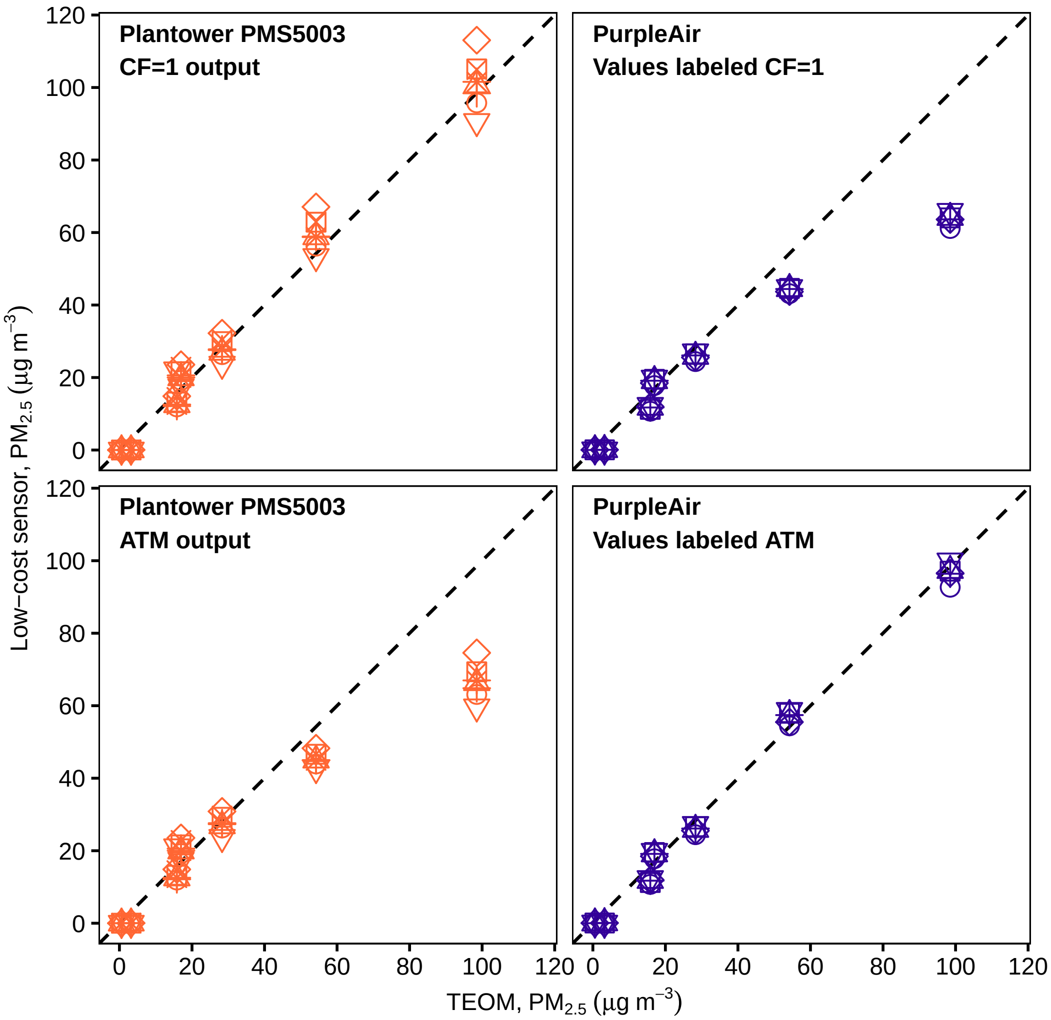

Figure S7. A comparison of time-averaged NIST Urban PM concentrations reported by the TEOM, eight unmodified Plantower PMS5003 sensors, and six PurpleAir monitors. The $PM_{2.5}$ CF=1 concentrations output by the PMS5003 sensors and the concentrations labeled "$PM_{2.5}$ CF=1" in the PurpleAir log files (firmware version 2.50i) are shown on the top. The $PM_{2.5}$ ATM concentrations output by the PMS5003 sensors and the concentrations labeled "$PM_{2.5}$ ATM" in the PurpleAir log files (firmware version 2.50i) are shown on the bottom. Dashed lines represent y=x.

S2.1.2 NIST Urban PM particle size distribution

Particle number size distributions (PNSD) measured for nebulized NIST Urban PM concentrations < 100 $\mu g\, m^{-3}$ during the third laboratory evaluation are shown in Figure S8. For these five concentration points, the average geometric mean diameter was 0.042 μm and the average geometric standard deviation was 2.05.

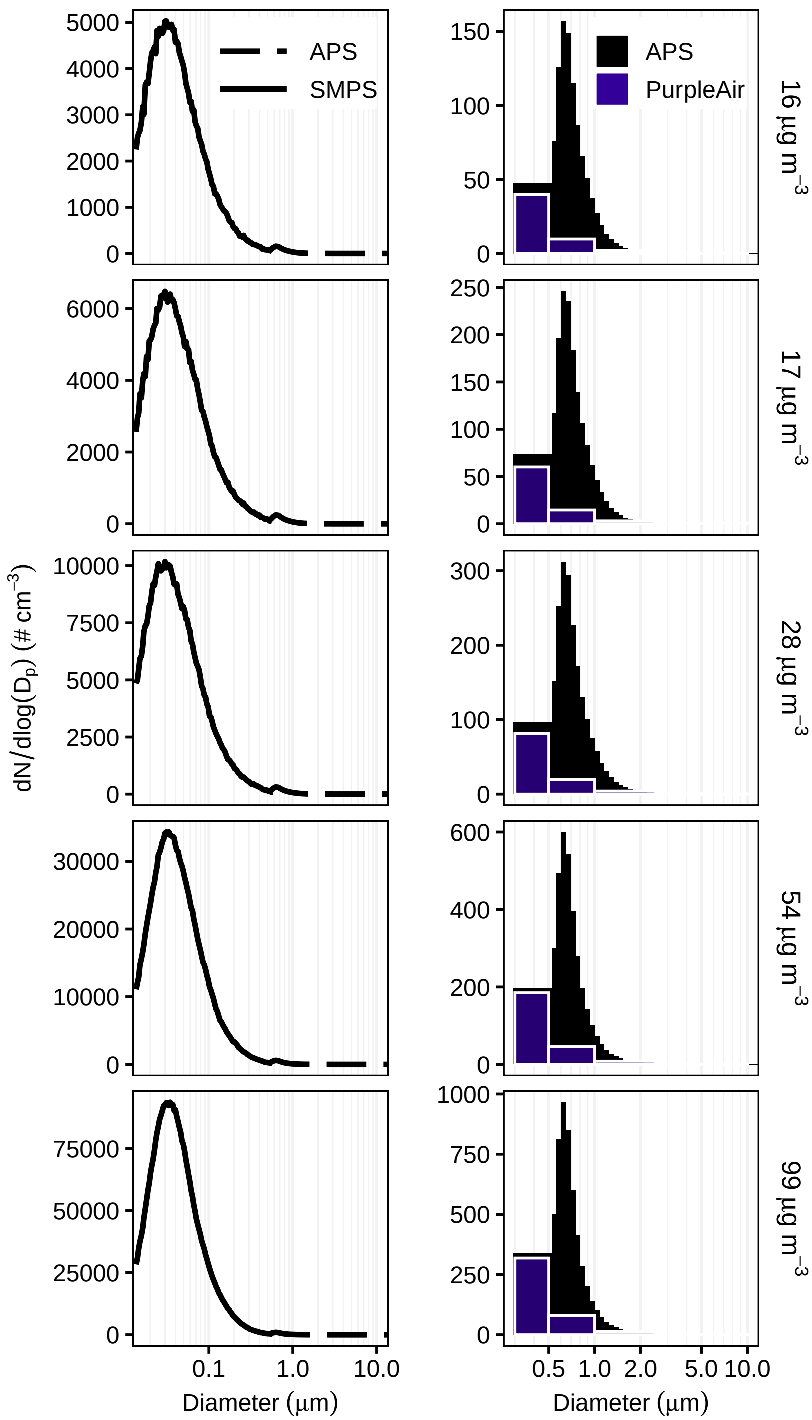

Figure S8. Particle number size distributions (PNSD) for nebulized NIST Urban PM concentrations < 100 $\mu g\, m^{-3}$. On the right, PNSDs measured by the APS are compared to particle count data reported by the PurpleAir monitors. The leftmost bin of APS data represents particles reported as "<0.523 μm".

S2.2 Break-in Period

The data recorded during the break-in period are shown in Figure S9. The median hourly "$PM_{2.5}$ ATM" concentration reported by the PurpleAirs typically exhibited the same trends as the hourly $PM_{2.5}$ concentration reported by the GRIMM EDM 180. We observed no obvious degradation in PurpleAir performance over this 52-day period. The RSD of the hourly "$PM_{2.5}$ ATM" concentrations reported by the 19 PurpleAir monitors increased during the last 12 days; however, these higher RSDs might be due to the very low $PM_{2.5}$ concentrations reported during that time (rather than degradation of the PurpleAir monitors over time).

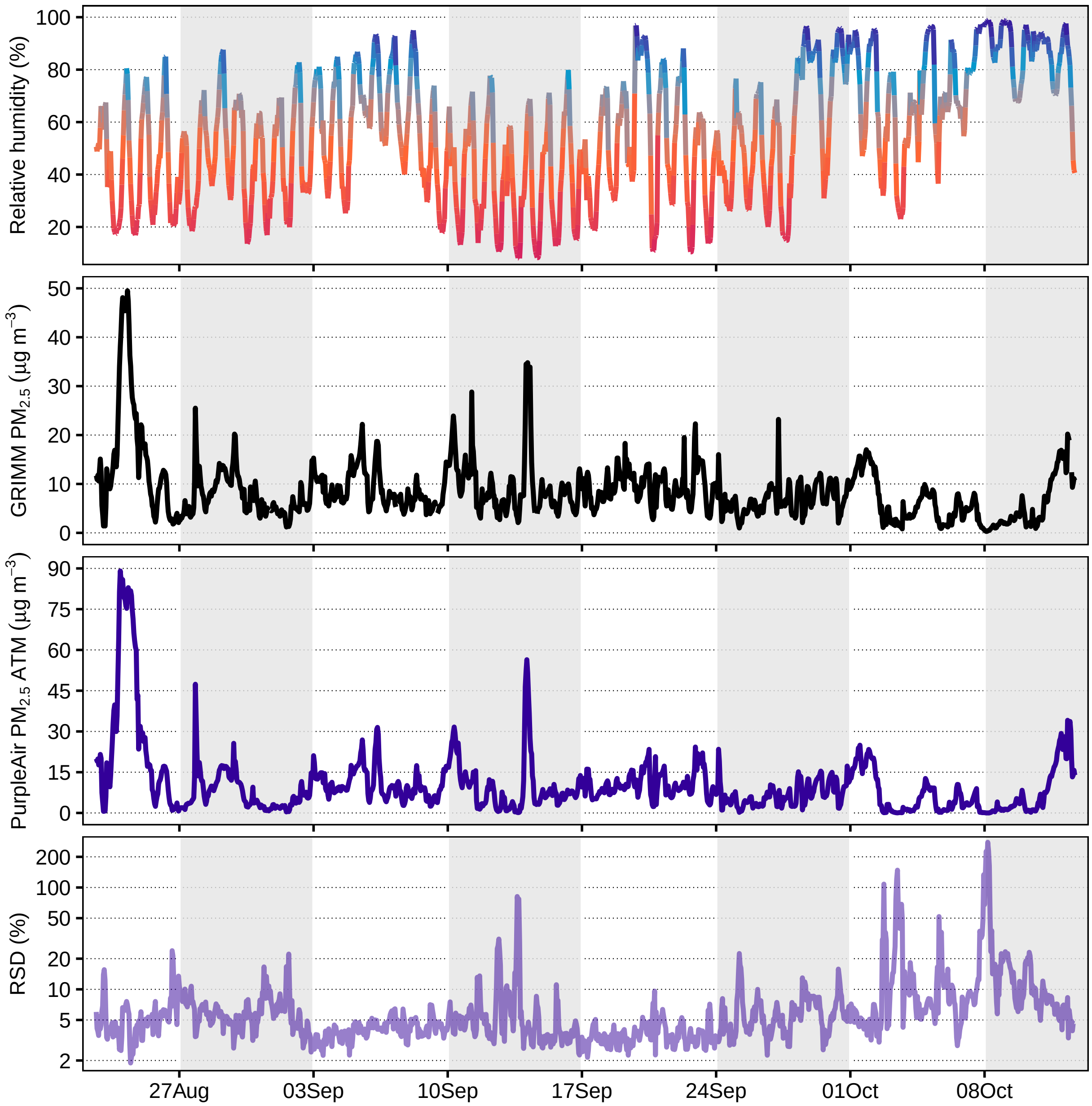

Figure S9. Top to bottom: Hourly ambient relative humidity reported at the main campus weather station, $PM_{2.5}$ concentration reported by the GRIMM EDM 180, median "$PM_{2.5}$ ATM" concentration reported by the 19 PurpleAir monitors, and relative standard deviation of the "$PM_{2.5}$ ATM" concentrations reported by the PurpleAir monitors during the break-in period. Note the difference in scales between the panel with the GRIMM EDM 180 data and the panel with the PurpleAir data.

S2.3 Network Deployment

The time-averaged "$PM_{2.5}$ ATM" concentrations reported by sensors A and B in each PurpleAir monitor over the duration of each UPAS filter sample are compared in Figure S10. Before the sample-averaged values shown in Figure S10 were calculated, 80-s readings for which the "$PM_{2.5}$ ATM" concentrations reported by sensors A and B differed by more than 5 $\mu g\,m^{-3}$ were eliminated. Sample averages were then eliminated if data were not available for at least 80% of the filter sample period (either because the monitor stopped logging or because sensors A and B disagreed for a large number of readings). Frequent disagreement between sensors A and B was observed for PA-003, PA-010, and PA-018, which is why few sample-averaged values were retained for those monitors.

Weather data collected during the network field deployment are shown in Figures S11–S13. While the PurpleAir monitors at Sites A–D were co-located with 16.7 $L\,min^{-1}$ filter samples (02 November and 06 December), the hourly temperature reported at the weather station ranged from -10 to 18 °C (median = 1.1 °C). The hourly RH reported at the weather station ranged from 10% to 100% (median = 61%). The hourly RH was below 50% for 36% of this period and below 75% for 69% of this period (Figure S11). Over the entire duration of the network field deployment (22 October to 06 December), the temperatures reported by the PurpleAir monitors were higher than those measured at the weather station (Figure S12) and the relative humidities reported by the PurpleAir monitors were lower than those measured at the weather station (Figure S13). Similarly, Magi et al. (2019) reported that PurpleAir monitors reported higher temperatures and lower relative humidities than a co-located weather station.

The bias (in $\mu g\,m^{-3}$) of the sample-averaged uncorrected PurpleAir "$PM_{2.5}$ ATM" concentrations over time is shown in Figure S14. Almost all of the uncorrected sample-averaged "$PM_{2.5}$ ATM" concentrations overestimated the concentration derived from the filter samples (bias > 0 $\mu g\,m^{-3}$ for 55/58 samples). Histograms illustrating the bias (in $\mu g\,m^{-3}$) of the sample-averaged $PM_{2.5}$ concentrations reported by the PurpleAir monitors after applying different correction factors are shown in Figure S15.

The effect of excluding data collected between 30 November and 06 December (i.e., samples that ended in December) when applying monthly correction factors is shown in Table S1 and Figure S17. This sensitivity analysis was conducted because the December "monthly" correction factors were applied to a single week at the beginning of December (instead of to an entire month's worth of data). When all data collected between 02 November and 06 December were considered, monthly correction factors increased the fraction of samples with absolute biases of ≤ 2 $\mu g\,m^{-3}$ and $\leq 20\%$ to 77% and 46%, respectively. When only data collected between 02 and 29 November were considered, monthly correction factors increased the fraction of samples with absolute biases of ≤ 2 $\mu g\,m^{-3}$ and $\leq 20\%$ to 72% and 38%, respectively.

The intraclass correlation coefficients calculated from the two-way random effects model shown in Equation 8 were 0.44 (95% CI = 0.18–0.63) for sample date and 0.03 (95% CI = 0.00–0.11) for PurpleAir monitor (using the full dataset, n=175). Clustering of correction factors by sample date is illustrated in Figure S18. For the dataset with the lowest and highest correction factors excluded (n=173), the ICCs were 0.49 (95% CI = 0.24–0.67) for sample date and 0.06 (95% CI = 0.00–0.14) for PurpleAir monitor.

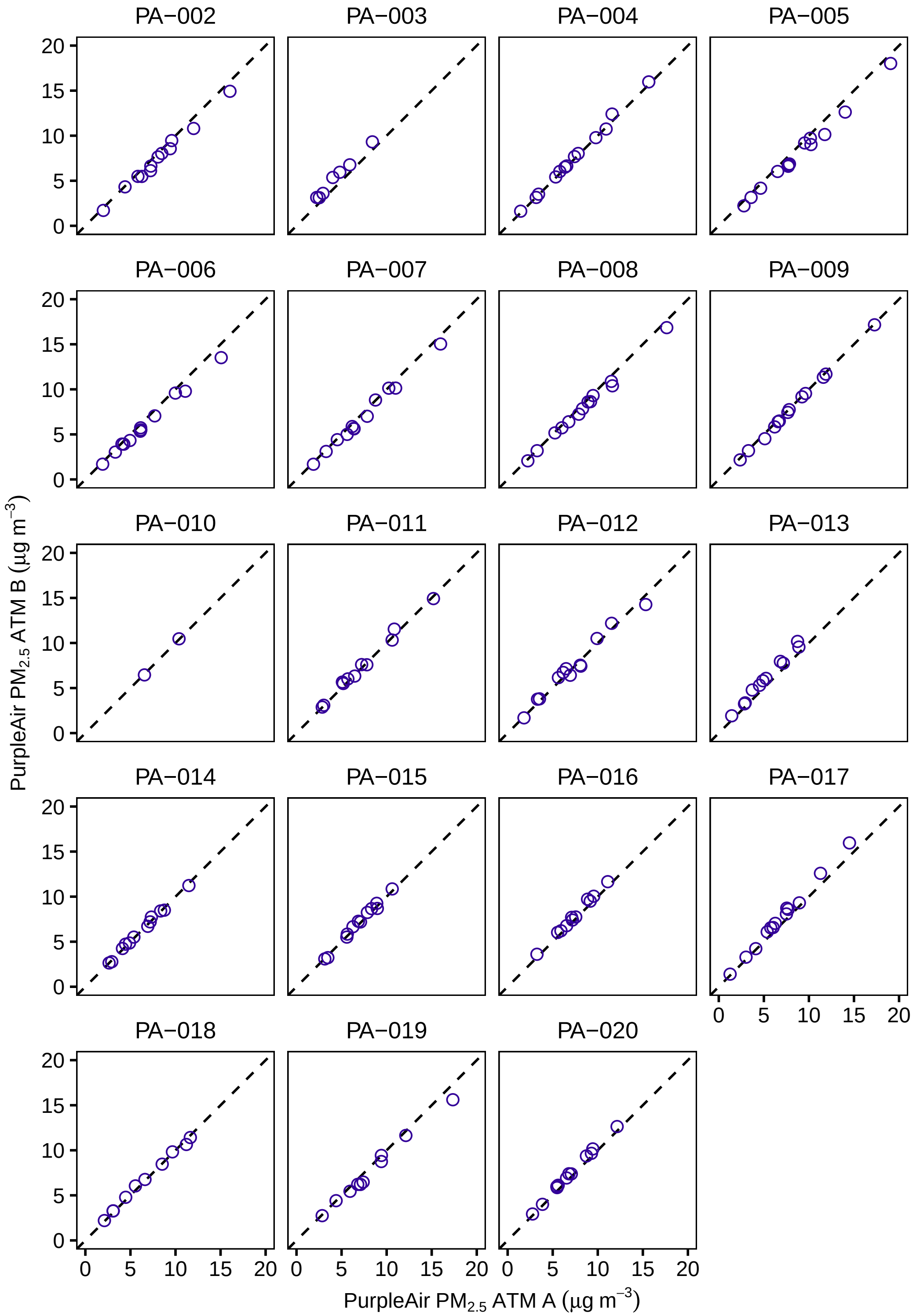

Figure S10. A comparison of the sample-averaged "$PM_{2.5}$ ATM" concentrations reported by sensors A and B in each PurpleAir monitor. Only the 211/242 samples retained after applying quality assurance criteria are shown.

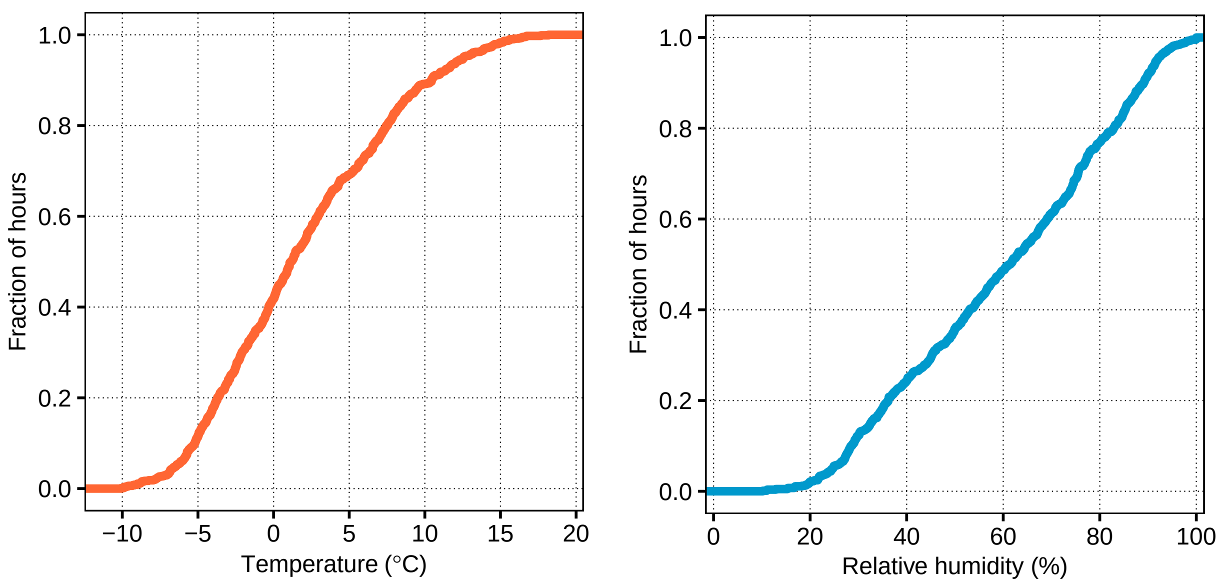

Figure S11. Empirical cumulative distribution functions for the hourly temperature and relative humidity (RH) values measured at the CSU main campus weather station between 02 November and 06 December. The value on the y-axis represents the fraction of hours for which the RH or temperature was below the value on the x-axis.

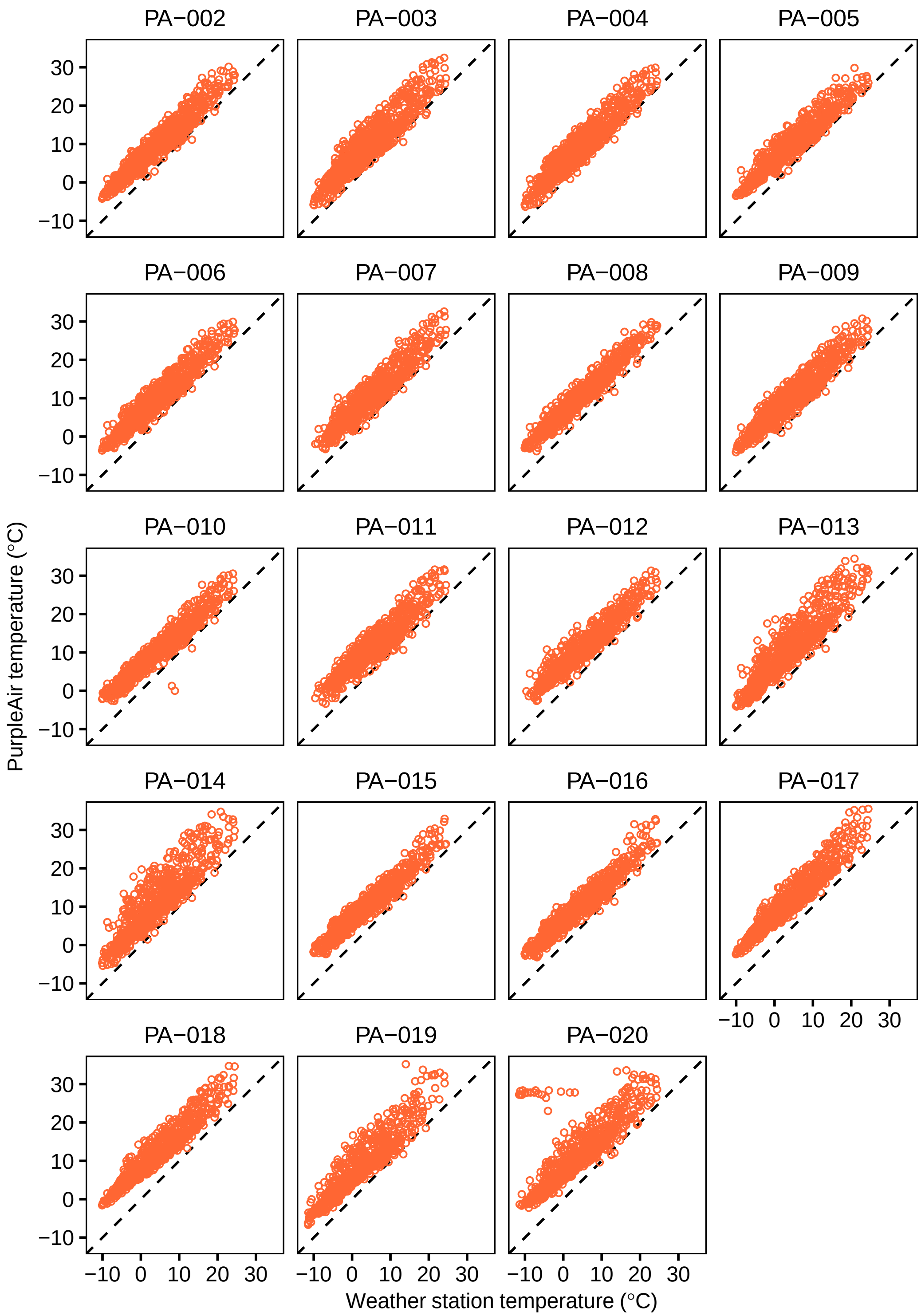

Figure S12. A comparison of the hourly temperature reported at the weather station on the CSU main campus and by each PurpleAir monitor between 22 October and 06 December. The diagonal line is y = x.

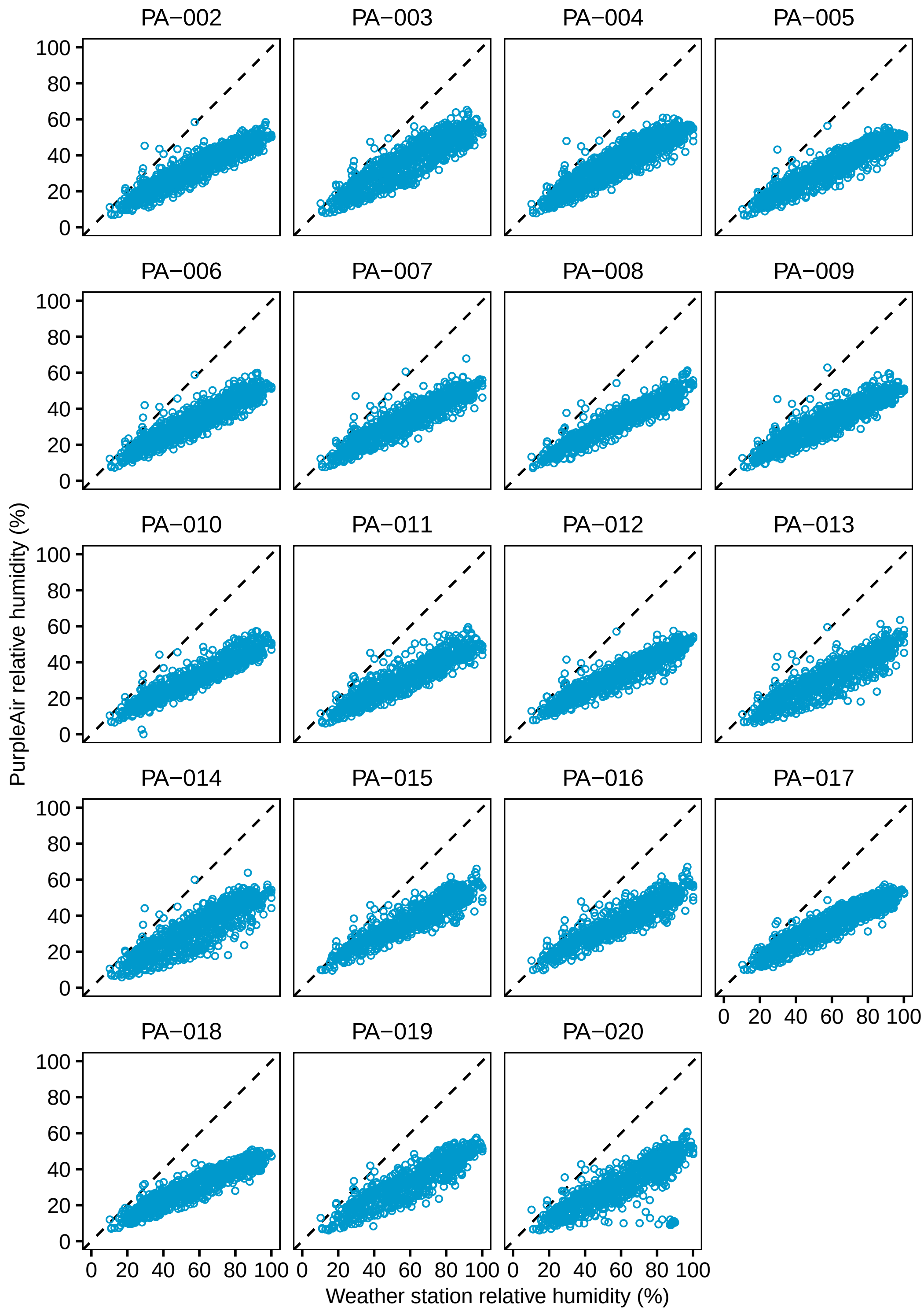

Figure S13. A comparison of the hourly relative humidity reported at the weather station on the CSU main campus and by each PurpleAir monitor between 22 October and 06 December. The diagonal line is y = x.

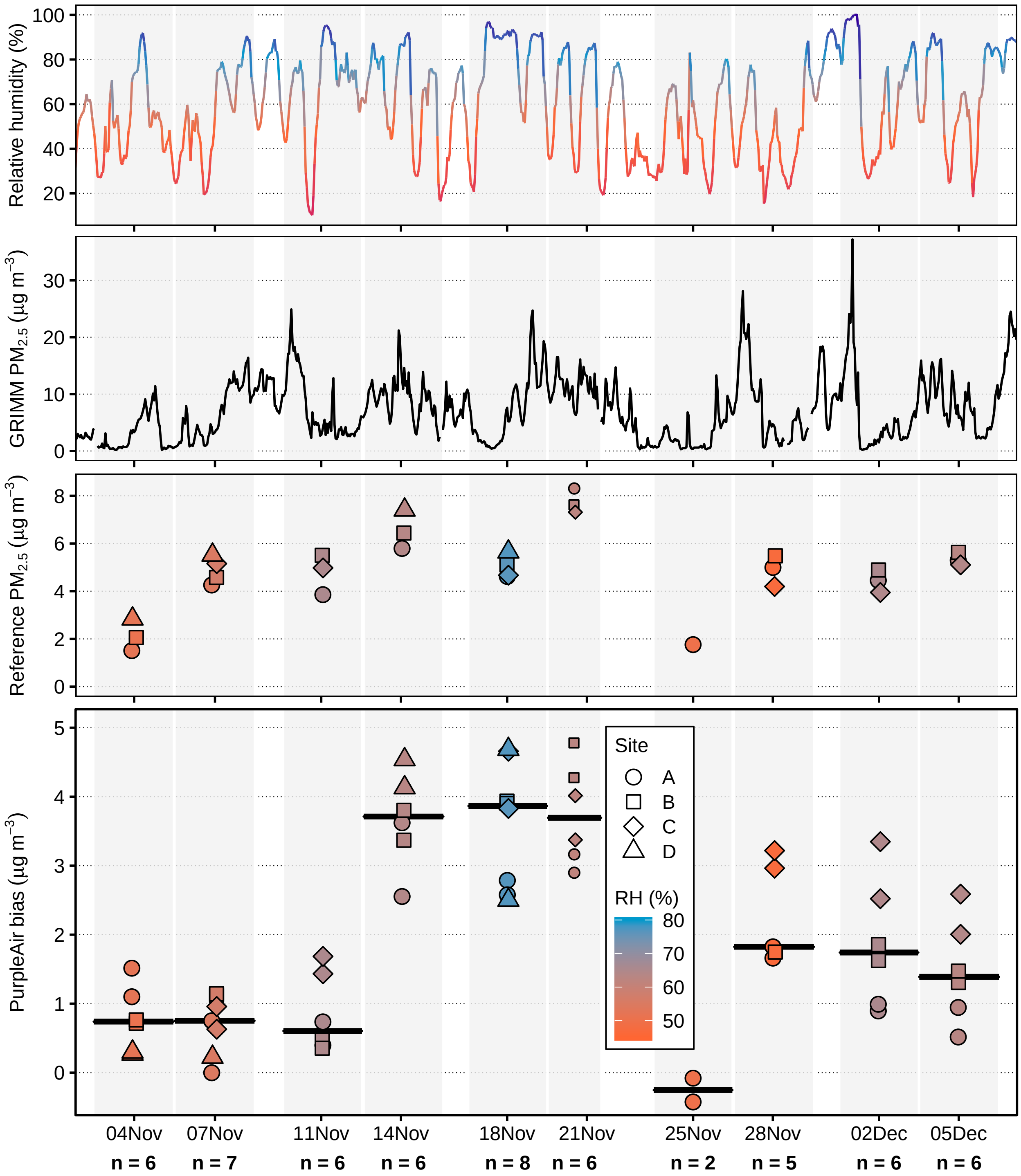

Figure S14. Top to bottom: Hourly relative humidity (RH) reported at the weather station, one-hour average $PM_{2.5}$ concentrations reported by the GRIMM EDM 180 at Site B, $PM_{2.5}$ concentrations derived from 16.7 L min^{-1} "reference" filter samples, and bias of uncorrected sample-averaged "$PM_{2.5}$ ATM" concentrations reported by PurpleAir monitors (relative to concentrations derived from co-located 16.7 L min^{-1} filter samples). Gray boxes represent the duration of each sample period. Horizontal black lines represent the median bias for each period. The number of valid comparisons obtained during each period is listed below the x-axis. Smaller markers are used for the 48-hour samples collected on 19-21 November. The marker color represents the sample-averaged RH reported at the weather station. Note that the color scale for the hourly RH extends beyond the range shown in the legend.

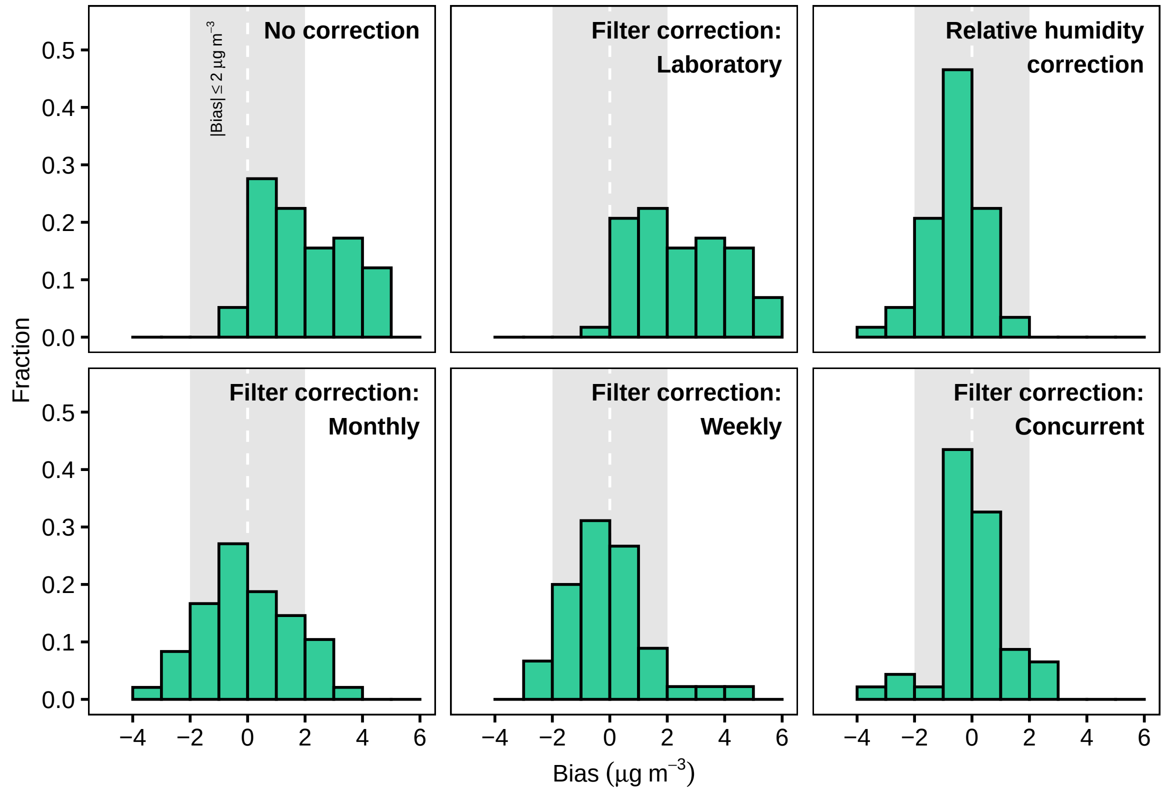

Figure S15. Bias (in $\mu g\,m^{-3}$) of sample-averaged "$PM_{2.5}$ ATM" concentrations reported by PurpleAir monitors, relative to the concentrations derived from co-located 16.7 $L\,min^{-1}$ filter samples, after applying different corrections. The y-axis represents the fraction of samples that fell into each bin.

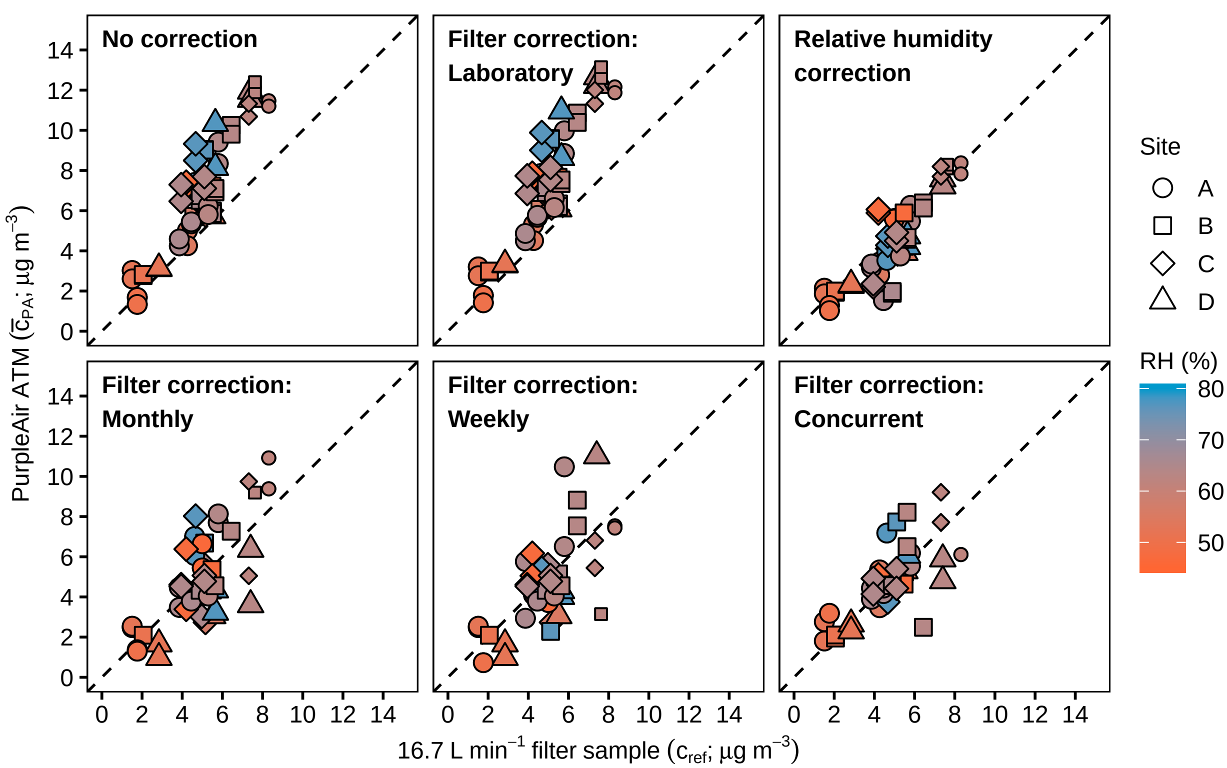

Figure S16. A comparison of sample-averaged $PM_{2.5}$ concentrations derived from 16.7 $L\,min^{-1}$ filter samples and reported by PurpleAir monitors ("$PM_{2.5}$ ATM") after applying different corrections. The diagonal line is y = x.

Table S1. Bias of sample-averaged "$PM_{2.5}$ ATM" concentrations reported by PurpleAir monitors, relative to concentrations derived from co-located 16.7 $L\,min^{-1}$ filter samples, after applying monthly correction factors. For the columns labeled "Nov. & Dec." all data collected between 02 November and 06 December were considered. For the columns labeled "Nov. only" only data collected between 02 and 29 November were considered.

	Concentration ($\mu g\,m^{-3}$)		Percent (%)	
Bias metrics	Nov. & Dec.	Nov. only	Nov. & Dec.	Nov. only
Median bias	-0.12	-0.04	-2	0
Min., max. bias	-3.8, 3.4	-3.8, 3.4	-65, 72	-65, 72
25^{th}, 75^{th} percentile	-1.1, 1.0	-1.1, 1.5	-21, 23	-22, 31
Fraction (percentage) of samples with $\lvert bias \rvert \leq 2\ \mu g\,m^{-3}$	37/48 (77%)	29/40 (72%)	-	-
Fraction (percentage) of samples with $\lvert bias \rvert \leq 20\%$	-	-	22/48 (46%)	15/40 (38%)

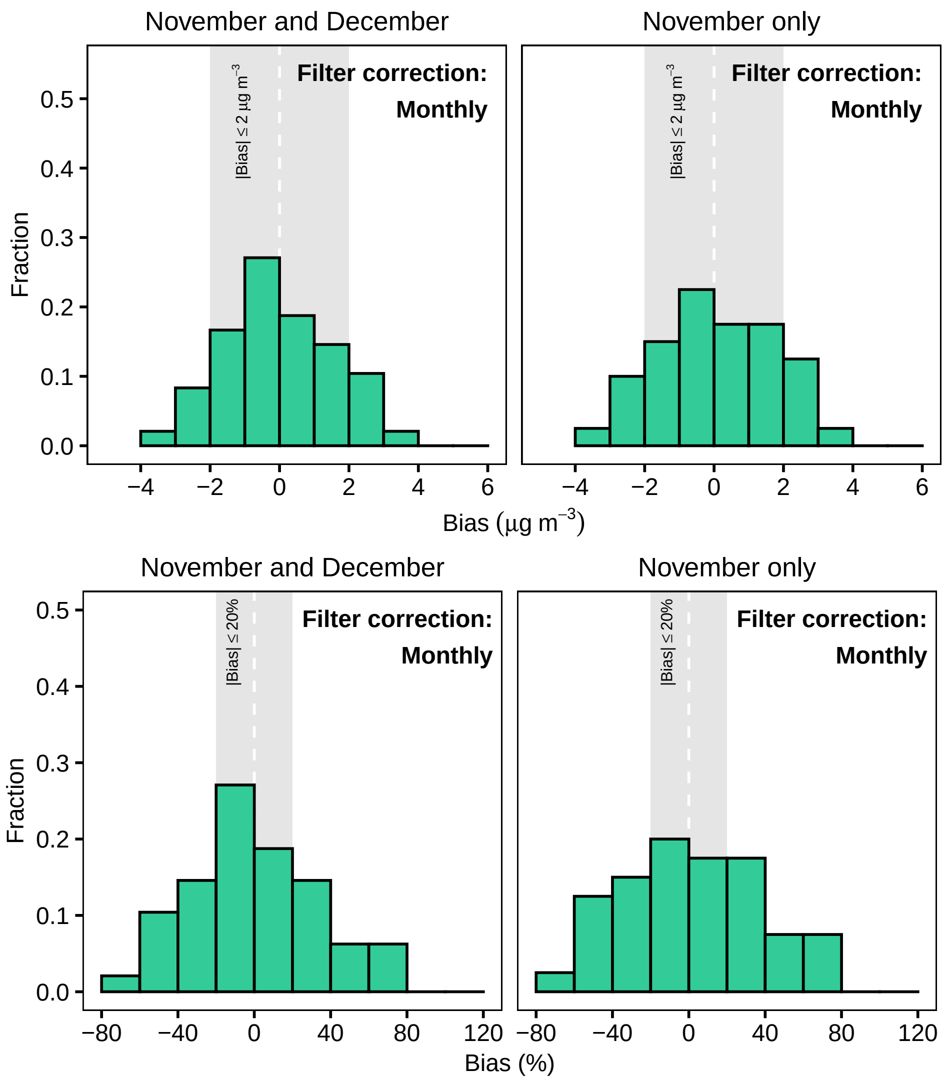

Figure S17. Bias (top: in $\mu g\,m^{-3}$; bottom: percent) of sample-averaged $PM_{2.5}$ concentrations reported by the PurpleAir monitors, relative to the concentrations derived from co-located 16.7 $L\,min^{-1}$ filter samples, after applying monthly correction factors. The y-axis represents the fraction of samples that fell into each bin. In the panels labeled "November and December", all data collected between 02 November and 06 December are shown. In the panels labeled "November only", only data collected between 02 and 29 November are shown.

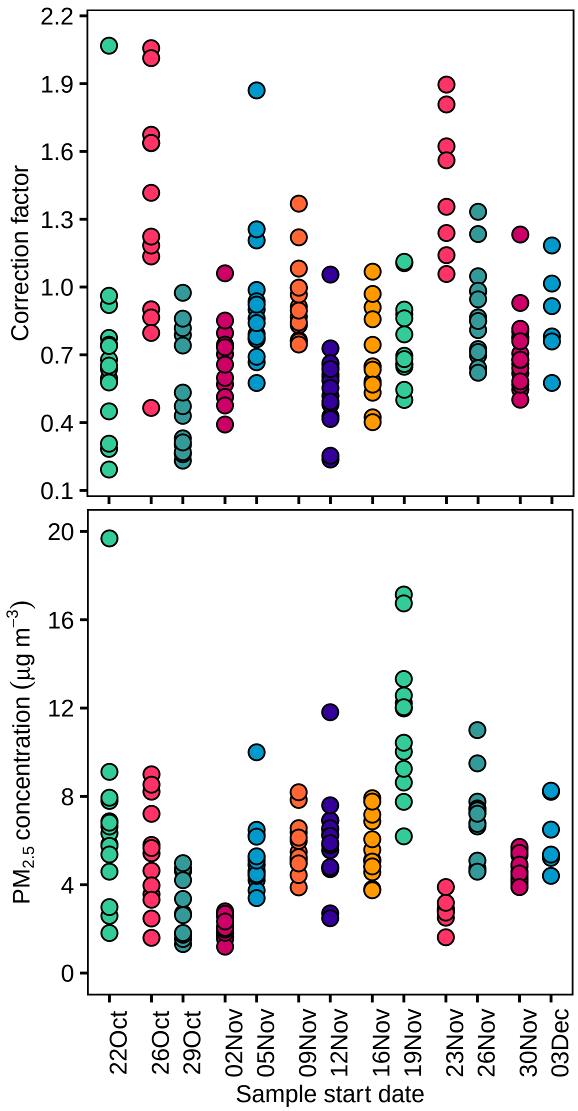

Figure S18. Concurrent gravimetric correction factors (top) and $PM_{2.5}$ concentrations (bottom) derived from UPAS filter samples collected during each sample period. Different sample periods are shown with different colors to improve readability. The sample period that started on 19 November lasted 48 hours; all other sample periods lasted 72 hours.

References

Lüdecke, D., 2019. *Package 'sjstats': Collection of Convenient Functions for Common Statistical Computations*. Version 0.17.4. URL: `https://CRAN.R-project.org/package=sjstats`.

Magi, B. I., Cupini, C., Francis, J., Green, M., and Hauser, C., 2019. Evaluation of $PM_{2.5}$ measured in an urban setting using a low-cost optical particle counter and a Federal Equivalent Method Beta Attenuation Monitor. Aerosol Science and Technology. DOI: `10.1080/02786826.2019.1619915`.

McGraw, K. O. and Wong, S. P., 1996. Forming inferences about some intraclass correlation coefficients. Psychological Methods 1(1), 30–46. DOI: `10.1037/1082-989X.1.1.30`.

Sayahi, T., Butterfield, A., and Kelly, K. E., 2019. Long-term field evaluation of the Plantower PMS low-cost particulate matter sensors. Environmental Pollution 245 932–940. DOI: `10.1016/j.envpol.2018.11.065`.