Electronic Supplementary Material (ESI) for Environmental Science: Atmospheres.

Supplementary Information:

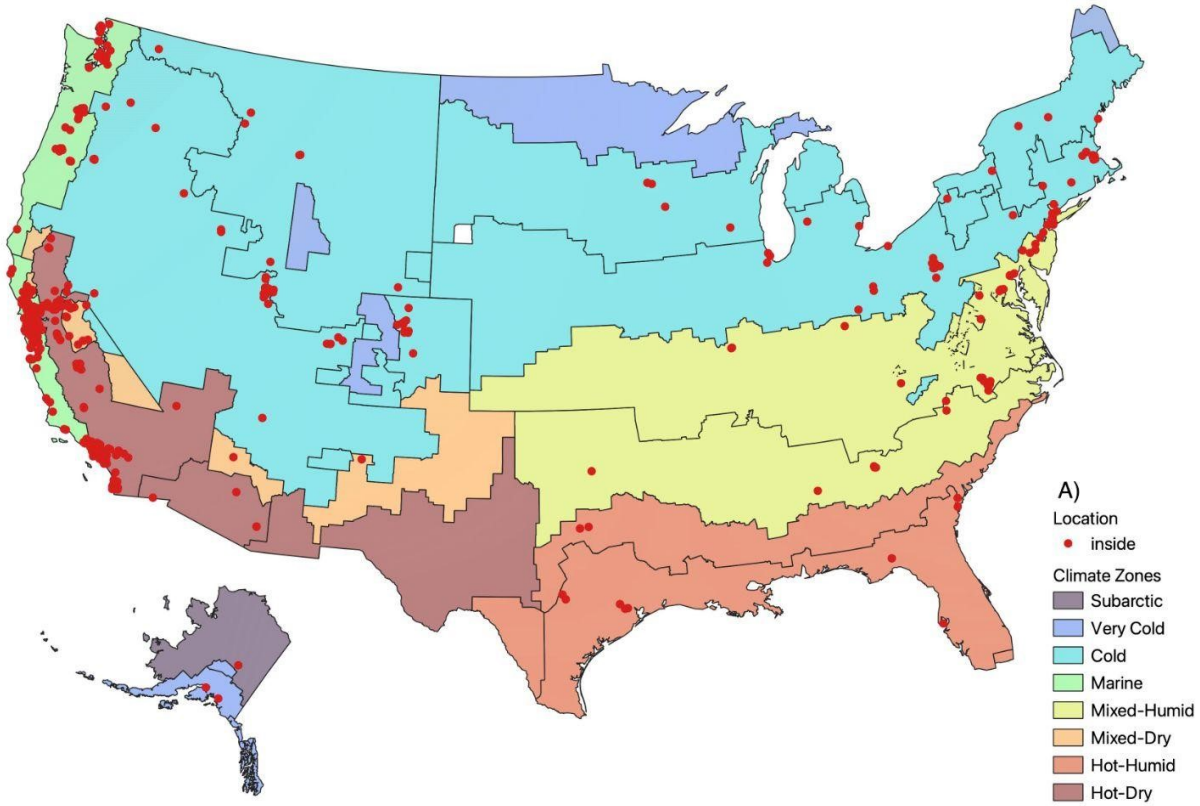
An analysis of degradation in low-cost particulate matter sensors

Priyanka deSouza[1,2*], Karoline Barkjohn[3], Andrea Clements[3], Jenny Lee[4], Ralph Kahn[5], Ben Crawford[6], Patrick Kinney[7]

1: Department of Urban and Regional Planning, University of Colorado Denver, Denver CO, 80202, USA
2: CU Population Center, University of Colorado Boulder, Boulder CO, 80302, USA
3: Office of Research and Development, US Environmental Protection Agency, 109 T.W. Alexander Drive, Research Triangle Park, NC 27711, USA
4: Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA
5: NASA Goddard Space Flight Center, Greenbelt, MA, 20771, USA
6: Department of Geography and Environmental Sciences, University of Colorado Denver, 80202, USA
7: Boston University School of Public Health, Boston, MA, 02118 USA

*: Corresponding author priyanka.desouza@ucdenver.edu

Figures


A)
Location
inside
Climate Zones
Subarctic
Very Cold
Cold
Marine
Mixed-Humid
Mixed-Dry
Hot-Humid
Hot-Dry

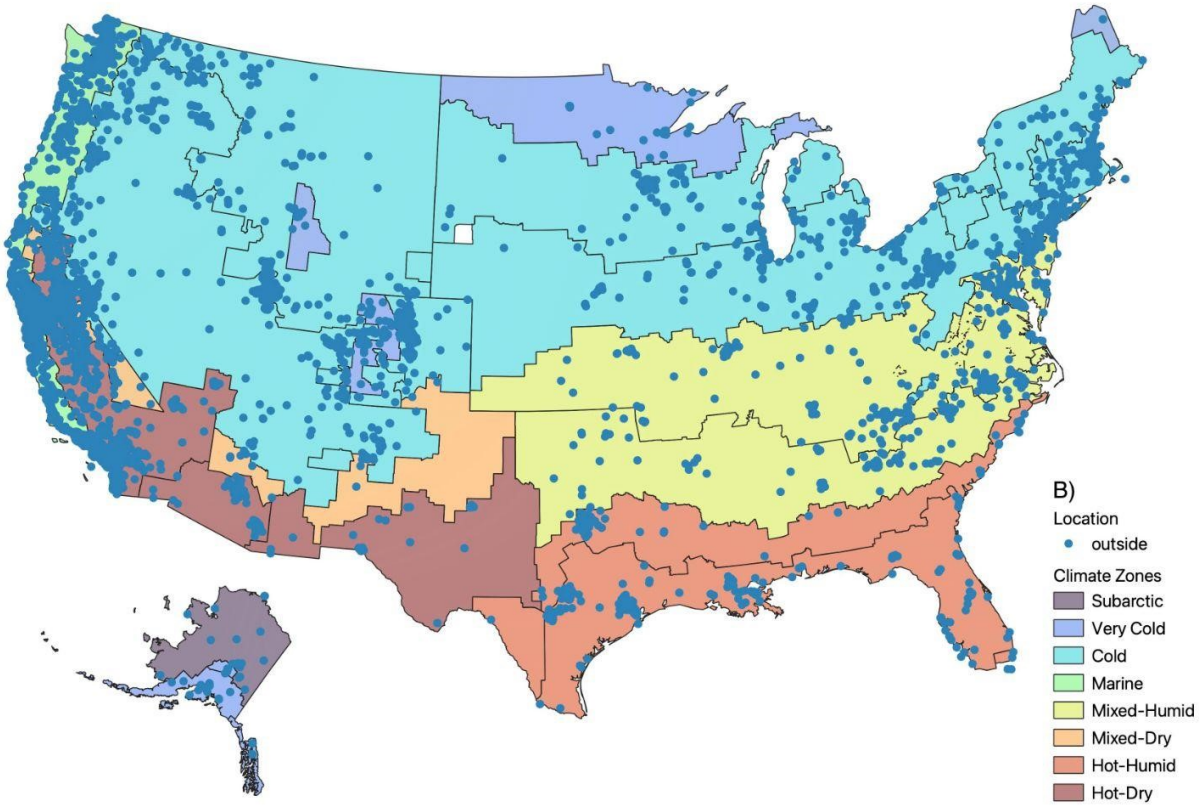

B)
Location
outside
Climate Zones
Subarctic
Very Cold
Cold
Marine
Mixed-Humid
Mixed-Dry
Hot-Humid
Hot-Dry

Figure S1: *A) Indoor and B) Outdoor PurpleAir monitors in the United States (Hawaii is not displayed) considered in this study*

Figure S2*: Sensors for which we did not have climate information*

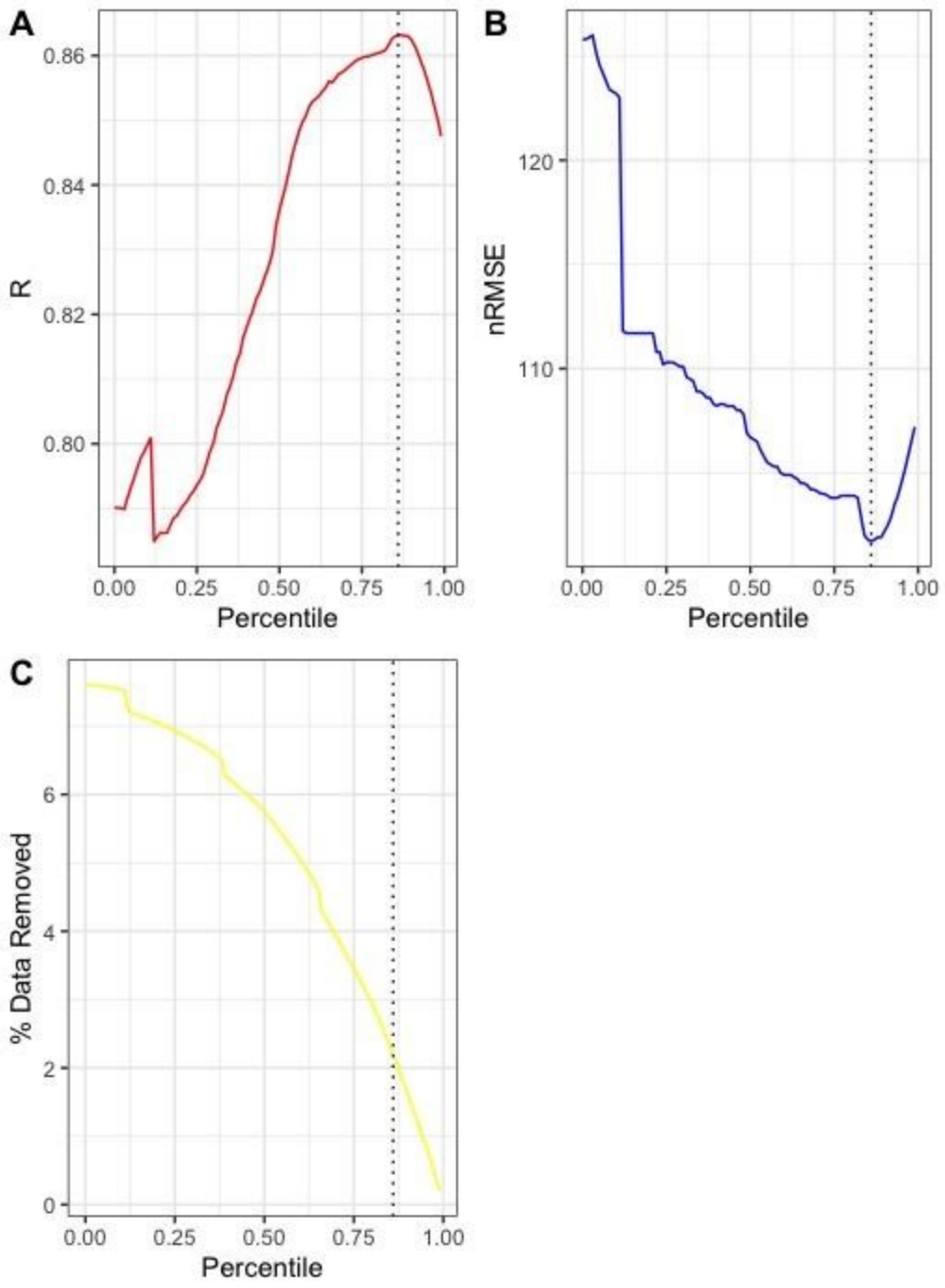


***Figure S3**: Agreement between the hourly PurpleAir measurements and the corresponding reference measurements from instruments not relying on the principles of light scattering, where measurements are flagged and removed based on the criterion" | channel A - channel B | < 5 µg/m³ and the % difference between channels A and B:* $\frac{abs(A - B) \times 2}{(A + B)} > x$ *th percentile of the percentage difference between A and B for each PurpleAir sensor, where we vary* x *between 0 - 0.99, captured by: A) Pearson*

correlation coefficient (R), and B) normalized root mean square error (nRMSE) metrics comparing unflagged measurements and the corresponding reference data based on different threshold percentile values. C) The % of measurements that were removed (because they were flagged) when evaluating R and nRMSE, for different percentile thresholds applied to the data are also displayed. The dotted vertical line represents the 86th percentile which corresponds to the lowest nRMSE and the highest R.

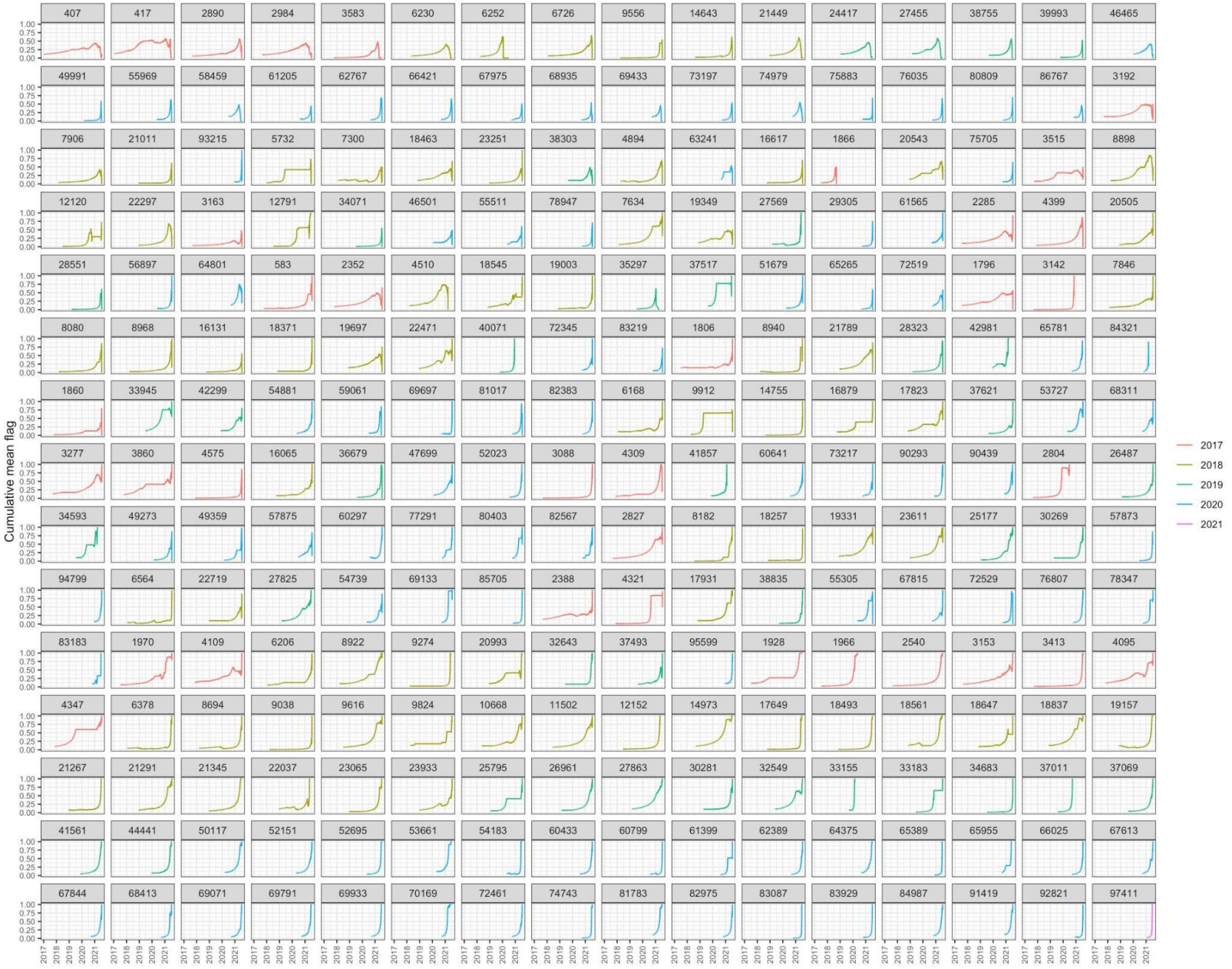
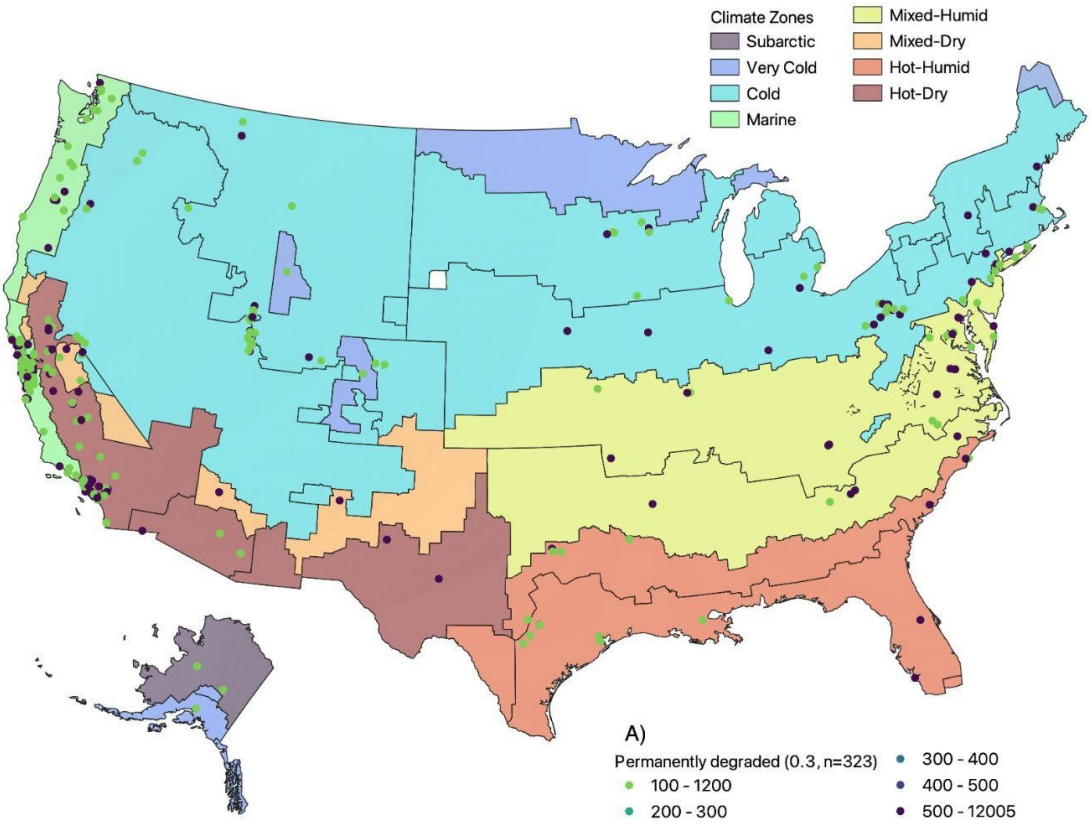


***Figure S4**: Cumulative mean of the flagged indicator at every hour of measurement calculated by averaging the flag indicator at the hour of measurement and for all subsequent hours for permanently degraded monitors (defined as sensors that record a*

cumulative mean flagged indicator > 0.4 for at least 100 hours of operation). The title of each subplot corresponds to the PurpleAir monitor ID


Climate Zones
Subarctic
Very Cold
Cold
Marine
Mixed-Humid
Mixed-Dry
Hot-Humid
Hot-Dry
A)
Permanently degraded (0.3, n=323)
100 - 1200
200 - 300
300 - 400
400 - 500
500 - 12005

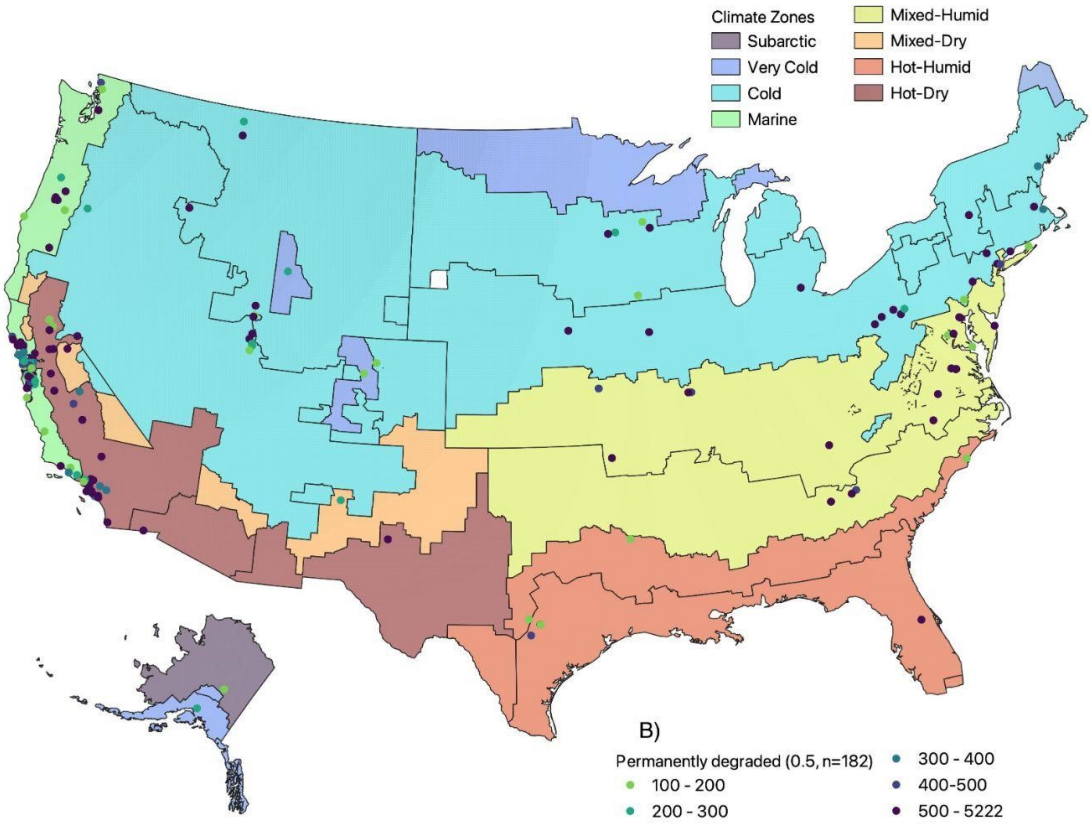

Climate Zones
Subarctic
Very Cold
Cold
Marine
Mixed-Humid
Mixed-Dry
Hot-Humid
Hot-Dry
B)
Permanently degraded (0.5, n=182)
100 - 200
200 - 300
300 - 400
400-500
500 - 5222

***Figure S5**: Map of permanently degraded PurpleAir monitors with at least 100 measurements for which the cumulative mean of the flag indicator is A) ≥ 0.3, B) ≥ 0.5. The number of hours of operation for which the cumulative mean of the flag indicator is above the threshold is displayed for each monitor*

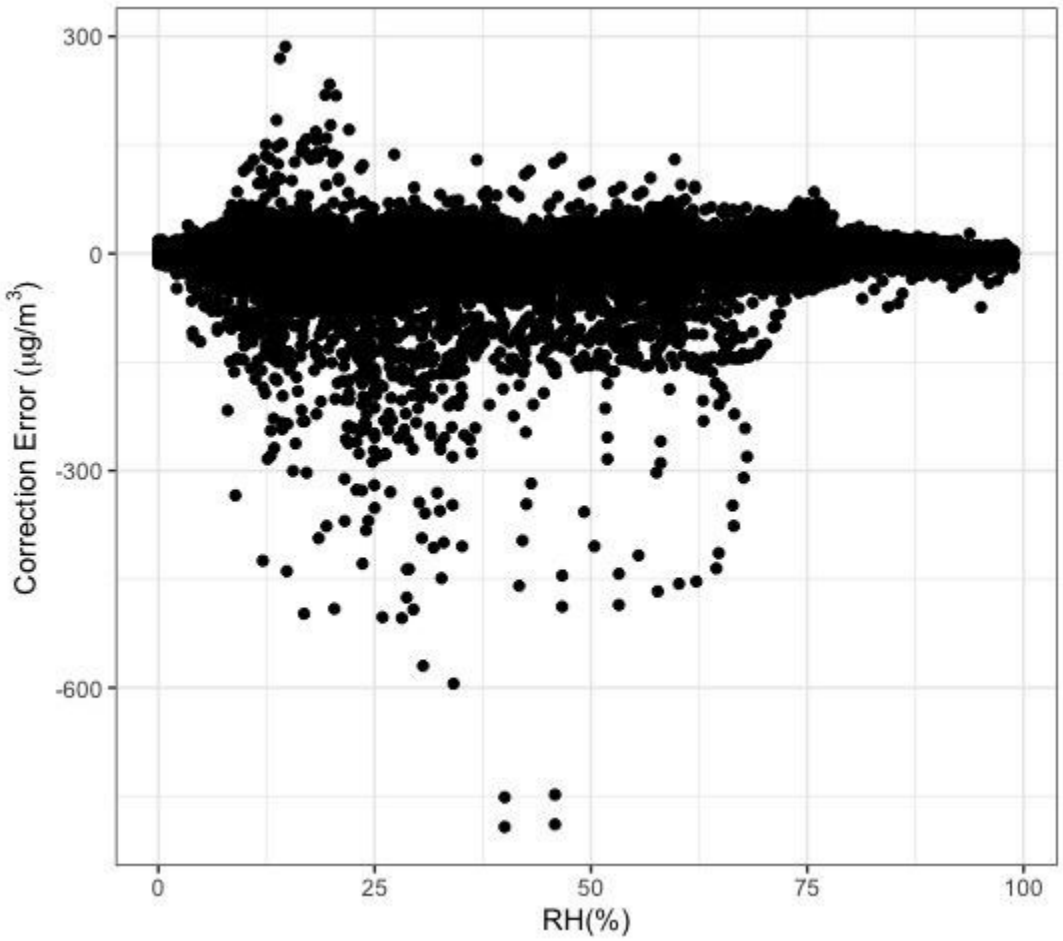


***Figure S6**: Correction error versus RH*

Tables

***Table S1**: Summary statistics of measurements and period of operation of PurpleAir monitors considered in this study by state*

State	Number of monitors	Number of measurements	Hours of operation per monitor (Min/Max, Mean, Median)
Alabama	18	253,307	Min/Max: 0/36,582 Mean: 15,780 Median: 15,228
Alaska	89	1,237,656	Min/Max:0/30,515 Mean: 11,041 Median:10,585
Arizona	100	809,779	Min/Max:0/34,237 Mean:10,252 Median:7,753
Arkansas	12	179,239	Min/Max:0/28,202 Mean:10,995 Median:10,436
California	7303	66,788,259	Min/Max:0/40,077 Mean:9,798 Median:6,998
Colorado	441	3,966,682	Min/Max:0/34,946 Mean:10,971 Median: 8,847
Connecticut	17	194,088	Min/Max:0/33,610 Mean:12,509 Median:11,306
Delaware	26	124,833	Min/Max:0/34,052 Mean:12,043 Median:13,322
D.C	10	90,458	Min/Max: 0/26,484 Mean: 8,887 Median: 6,861
Florida	49	525,499	Min/Max:0/34,237 Mean: 12,468 Median:11,309
Georgia	60	540,393	Min/Max:0/34,453 Mean: 9,284 Median: 6,912
Hawaii	41	637,328	Min/Max:0/35,119

			Mean: 13,687 Median:13,061
Idaho	75	599,536	Min/Max:0/33,223 Mean: 11,968 Median:10,336
Illinois	63	586,783	Min/Max:0/34,453 Mean: 11,787 Median:9,886
Indiana	40	591,389	Min/Max:0/34,786 Mean:12,642 Median:12,067
Iowa	34	393,034	Min/Max:0/39,633 Mean:14,215 Median:13,721
Kansas	16	586,783	Min/Max:0/30,126 Mean:9,349 Median:6,815
Kentucky	8	591,389	Min/Max:0/34,089 Mean:12,514 Median:10,747
Louisiana	35	450,334	Min/Max:0/31,522 Mean:11,766 Median:11,218
Maine	33	203,152	Min/Max:0/31,871 Mean:9,158 Median:7,361
Maryland	143	860,122	Min/Max:0/35,097 Mean:11,237 Median:8,856
Massachusetts	123	775,583	Min/Max:0/34,688 Mean:10,350 Median:7,988
Michigan	71	637,766	Min/Max:0/37,905 Mean:12,011 Median:10,986
Minnesota	66	749,580	Min/Max:0/34,452 Mean:12,324 Median:11,255
Mississippi	3	21,728	Min/Max:0/21,322 Mean:9,763 Median:9,183

Missouri	14	148,132	Min/Max:0/32,188 Mean: 10,075 Median:8,072
Montana	53	358,730	Min/Max:0/34,788 Mean:10,902 Median:8,160
Nebraska	8	131,600	Min/Max:1/34,592 Mean:13,739 Median:11,910
Nevada	124	1,042,406	Min/Max:0/28,064 Mean:9,045 Median:6,670
New Hampshire	18	220,990	Min/Max:0/37,804 Mean:15,690 Median:15,460
New Jersey	34	376,849	Min/Max:0/35,119 Mean:10,891 Median:9,235
New Mexico	53	594,760	Min/Max:0/34,669 Mean:12,485 Median:10,494
New York	113	1,026,533	Min/Max:0/34,090 Mean:11,097 Median:9,699
North Carolina	176	2,105,292	Min/Max:0/39,123 Mean:12,444 Median:10,699
North Dakota	4	9,824	Min/Max:1/10,742 Mean: 5,927 Median:5,831
Ohio	63	762,122	Min/Max:0/34,668 Mean:11,430 Median: 9,710
Oklahoma	16	198,650	Min/Max:0/35,097 Mean:12,808 Median:10,674
Oregon	490	4,731,279	Min/Max:0/35,245 Mean:10,924 Median:8,893
Pennyslvania	242	2,700,209	Min/Max:0/34,450 Mean:12,364 Median:11,210

Rhode Island	7	97,820	Min/Max:1/24,700 Mean: 12,533 Median:12,378
South Carolina	19	249,528	Min/Max:0/33,611 Mean:11,501 Median:9,881
South Dakota	5	123,915	Min/Max:0/34,687 Mean:16,455 Median:16,431
Tennessee	38	378,017	Min/Max:0/32,475 Mean:10,606 Median:8,541
Texas	215	2,216,316	Min/Max:0/34,786 Mean:11,973 Median:10,260
Utah	492	7,640,288	Min/Max:0/39,873 Mean:13,670 Median:12,735
Vermont	10	121,301	Min/Max:0/34,687 Mean:12,310 Median:9,612
Virginia	80	910,758	Min/Max:0/35,097 Mean:11,954 Median:10,989
Washington	677	5,787,583	Min/Max:0/35,244 Mean:10,324 Median: 7,676
West Virgina	16	183,751	Min/Max:0/34,689 Mean:12,177 Median:10,178
Wisconsin	56	604,028	Min/Max:0/33,275 Mean:11,427 Median: 10,189
Wyoming	33	156,177	Min/Max:0/32,212 Mean: 6,101 Median: 4,678

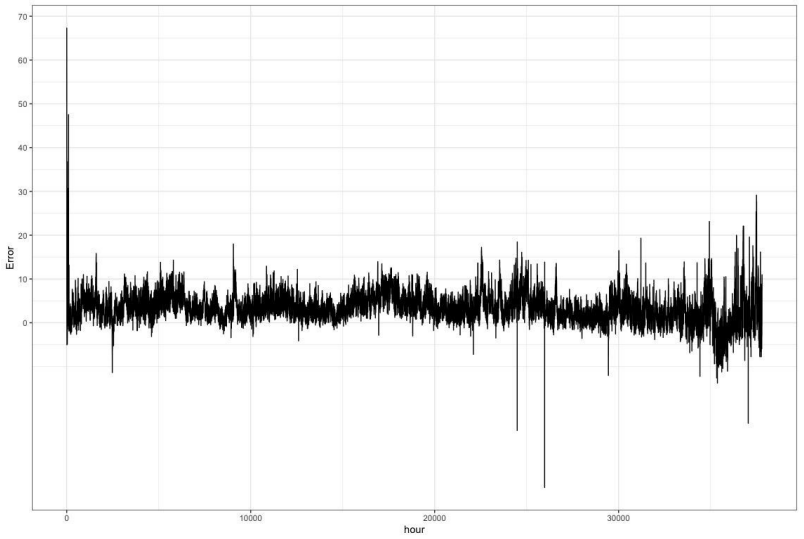
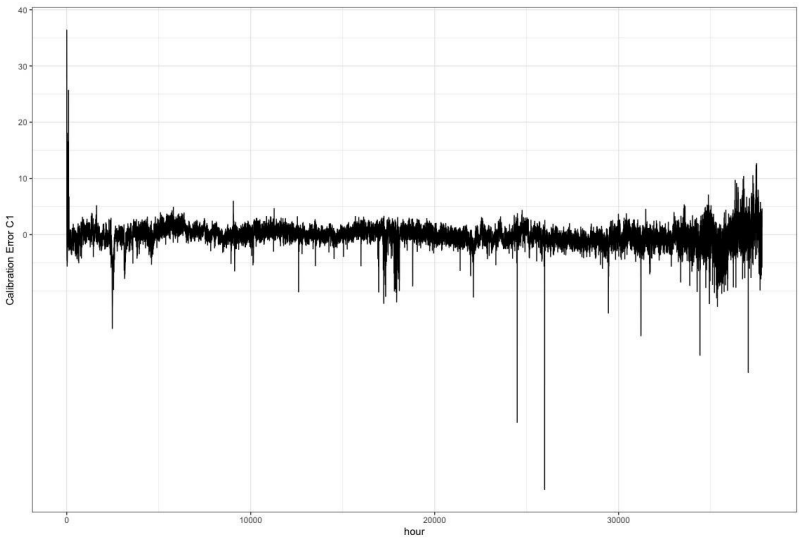
Table S2*: Summary statistics of measurements and period of operation of PurpleAir monitors considered in this study by climate*

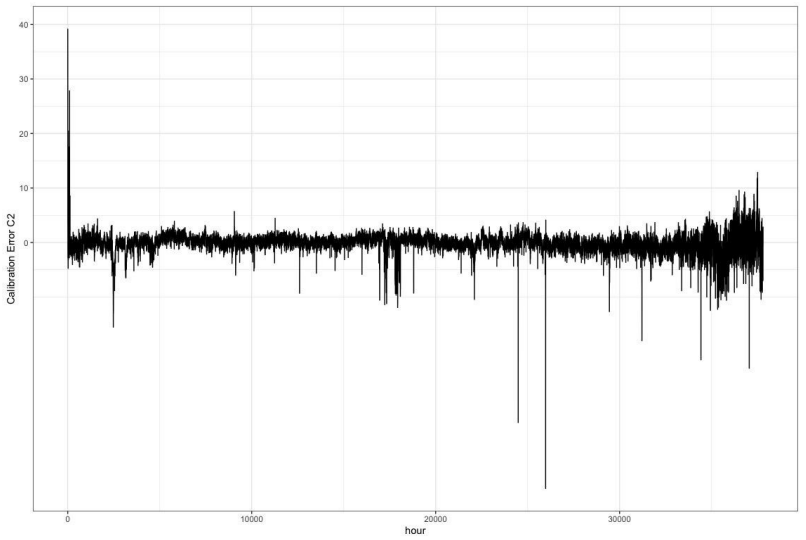
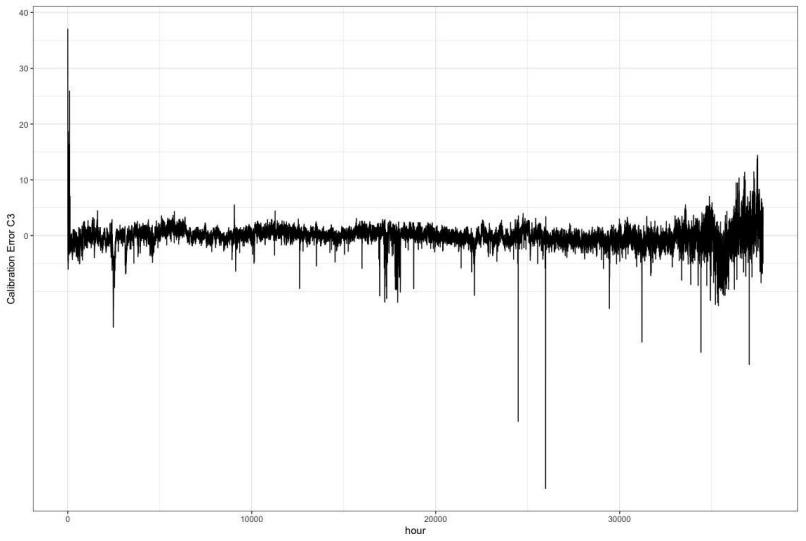
	All (Including flagged measurements)	**Co-located (After removing flagged measurements)**

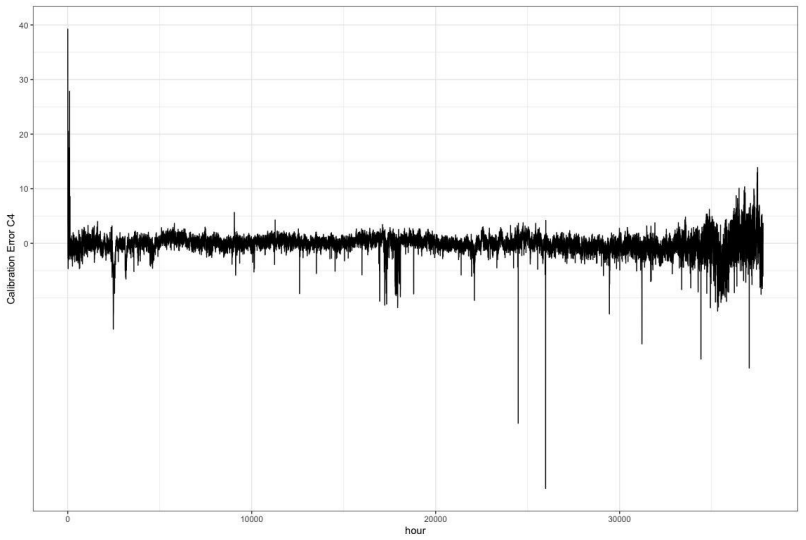
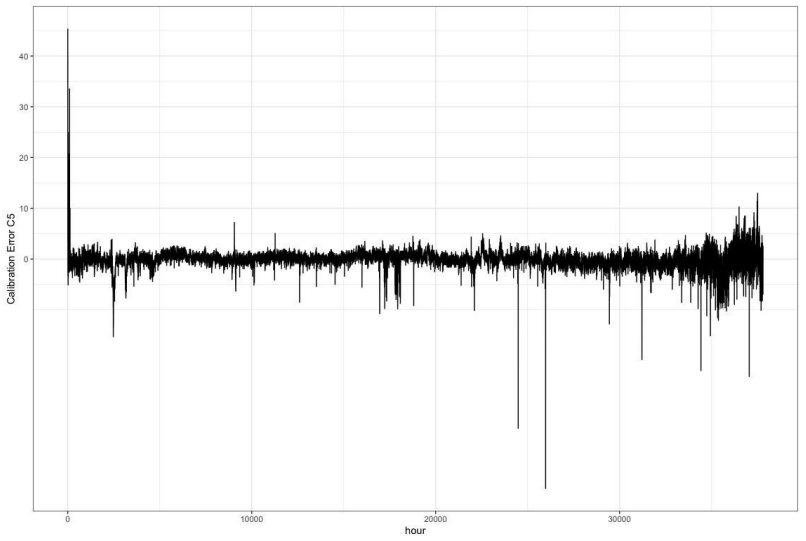
	Number of monitors	Number of measurements	Hours of operation per monitor (Min/Max, Mean, Median)	Min/Mean/Median/Max $PM_{2.5}$ concentrations reported by PurpleAir (% of measurements > 100 µg/m³)	Number of monitors	Number of measurements	Hours of operation per monitor (Min/Max, Mean, Median)	Min/Mean/Median/Max $PM_{2.5}$ concentrations reported by PurpleAir (% of measurements > 100 µg/m³)
Cold	2,458	26,704,292	Min/Max:0/39,873 Mean: 12,163 Median: 10,795	0/9/4/1454 (0.5%)	39	345,425	Min/Max:0/37,803 Mean: 14,066 Median: 13,307	0/13/8/1295 (0.5%)
Hot-Dry	2,680	27,210,216	Min/Max:0/40,077 Mean: 11,487 Median:9,404	0/13/6/1459 (1.1%)	72	603,505	Min/Max:0/35,289 Mean:15,444 Median: 14,414	0/15/8/773 (0.7%)
Hot-Humid	281	3,243,029	Min/Max:0/37,281 Mean: 12,273 Median: 11,027	0/11/8/1105 (0.1%)	1	15,432	Min/Max:0/21,661 Mean: 12,893 Median: 12,572	0/8/5/216 (0.1%)
Marine	4,842	41,448,103	Min/Max: 0/39,892 Mean: 8,644 Median: 6,178	0/10/4/1457 (1.0%)	21	318,255	Min/Max: 0/34,594 Mean: 14,641 Median: 13,648	0/12/6/1050 (0.7%)
Mixed-Dry	361	3,527,237	Min/Max: 0/37,805 Mean: 12,197 Median: 10,786	0/11/4/1445 (1.7%)	2	35,041	Min/Max:1/26,359 Mean: 15,424 Median: 16,109	0/8/1/586 (1.7%)
Mixed-Humid	750	6,959,785	Min/Max: 0/39,123 Mean: 11,389 Median: 9,434	0/12/9/1113 (0.3%)	16	145,498	Min/Max: 0/30,827 Mean: 9,932 Median: 8,386	0/13/10/247 (0.1%)
Subarctic	58	839,563	Min/Max:0/30,029 Mean:11,324 Median: 11,054	0/7/2/1292 (0.7%)	-	-	-	-
Very Cold	108	991,326	Min/Max:0/34,946 Mean: 11,242 Median: 10,326	0/5/2/859 (0.3%)	-	-	-	-
NA	394	3,336,389	Min/Max:0/36,294 Mean: 9,684 Median: 6,709	0/9/4/1206 (0.7%)	-	-	-	-
Outside								
	Number of monitors	**Number of measurements**	**Hours of operation per monitor (Min/Max, Mean, Median)**	**Min/Mean/Median/Max $PM_{2.5}$ concentrations**	-	-	-	-

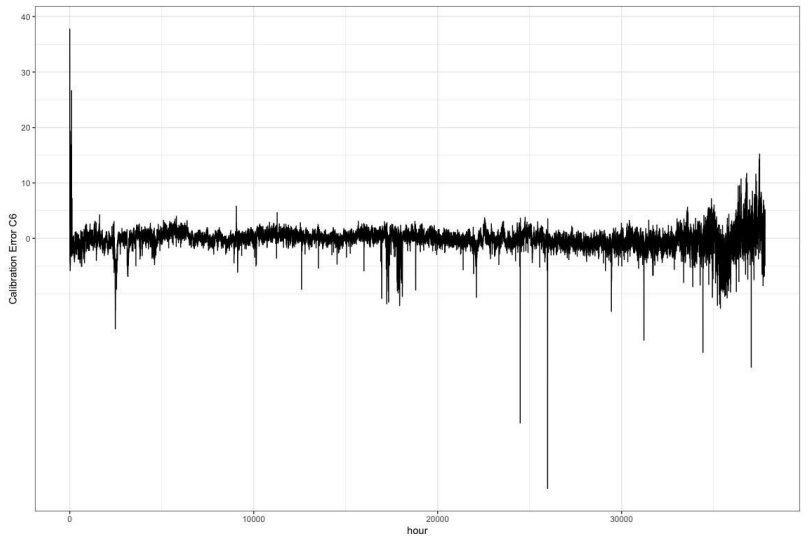
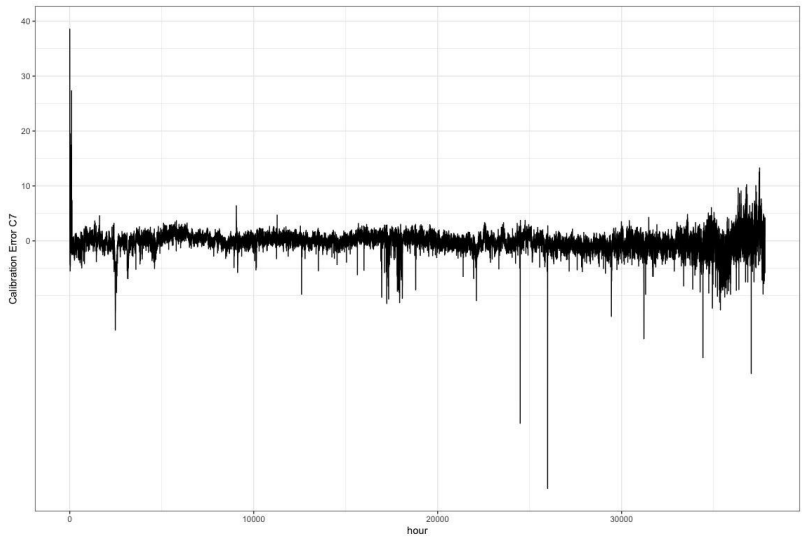
				reported by PurpleAir (% of measurements > 100 µg/m^3)				
Cold	2,333	25,584,998	Min/Max:0/39,873 Mean: 12,165 Median: 10,790	0/10/4/1455 (0.5%)	-	-	-	-
Hot-Dry	2,464	25,154,043	Min/Max: 0/40,077 Mean: 11,547 Median: 9,473	0/14/7/1459 (1.1%)	-	-	-	-
Hot-Humid	268	3,134,869	Min/Max: 0/37,281 Mean: 12,407 Median: 11,182	0/11/8/1105 (0.1%)	-	-	-	-
Marine	4,372	37,247,271	Min/Max:0/39,892 Mean: 8,627 Median: 6,155	0/10/5/1457 (1.1%)	-	-	-	-
Mixed-Dry	337	3,325,742	Min/Max: 0/37,805 Mean: 12,064 Median: 10,584	0/12/4/1445 (1.8%)	-	-	-	-
Mixed-Humid	697	6,590,197	Min/Max: 0/39,123 Mean: 11,539 Median: 9,630	0/12/9/991 (0.2%)	-	-	-	-
Subarctic	57	835,664	Min/Max: 0/30,029 Mean: 11,347 Median: 11,096	0/7/2/1292 (0.7%)	-	-	-	-
Very Cold	106	968,988	Min/Max: 0/34,946 Mean: 11,295 Median: 10,383	0/5/2/859 (0.3%)	-	-	-	-
NA	363	3,106,013	Min/Max:0/36,294 Mean: 9,766 Median: 6,810	0/9/4/1206 (0.7%)	-	-	-	-

Table S3*: Performance of the correction models as captured using root mean square error (RMSE), and Pearson correlation (R). LOSO CV was used to prevent overfitting in the random forest models.* A detailed description of the correction models proposed can be found in (deSouza et al., 2022). Model 2 corresponds to the correction model described using Equation (1)

Model Number	Name	Equation	Evaluating Correction Model		Summary of error: Mean (Median) ($\mu g/m^3$)	
			R	RMSE ($\mu g/m^3$)		
	Raw PurpleAir measurements					
0	Raw		0.88	12.5	3.89 (0.95)	
	Multivariate Regression (LOSO CV)					
1	Linear	$PM_{2.5,\ corrected} = PM_{2.5} \times s1 + b + \varepsilon$	0.88	6.8	3.4 (2.2)	

2	+RH	$PM_{2.5,\ corrected} = PM_{2.5}$ x s_1 + RH x s_2 + b + ε	0.89	6.6	3.3 (2.2)	
3	+T	$PM_{2.5,\ corrected} = PM_{2.5}$ x s_1 + T x s_2 + b + ε	0.88	6.7	3.3 (2.2)	

4	+RH x T	$PM_{2.5,\ corrected} = PM_{2.5} \times s_1 + RH \times s_2 + T \times s_3 + RH \times T \times s_4 + b + \varepsilon$	0.89	6.6	3.3 (2.3)	
5	PM x RH	$PM_{2.5,\ corrected} = PM_{2.5} \times s_1 + RH \times s_2 + RH \times PM_{2.5} \times s_3 + b + \varepsilon$	0.89	6.4	3.1 (2.0)	

6	PM x T	$PM_{2.5,\ corrected} = PM_{2.5} \times s_1 + T \times s_2 + T \times PM_{2.5} \times s_3 + b + \varepsilon$	0.88	6.7	3.3 (2.2)	
7	PM x nonlinear RH	$PM_{2.5,\ corrected} = PM_{2.5} \times s_1 + RH^2/(1\text{-}RH) \times s_2 + RH^2/(1\text{-}RH) \times PM_{2.5} \times s_3 + b + \varepsilon$	0.88	6.7	3.3 (2.2)	

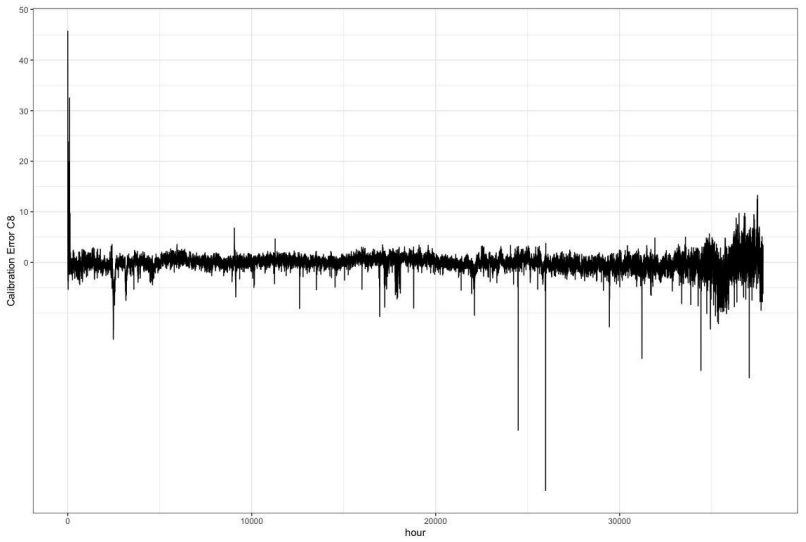
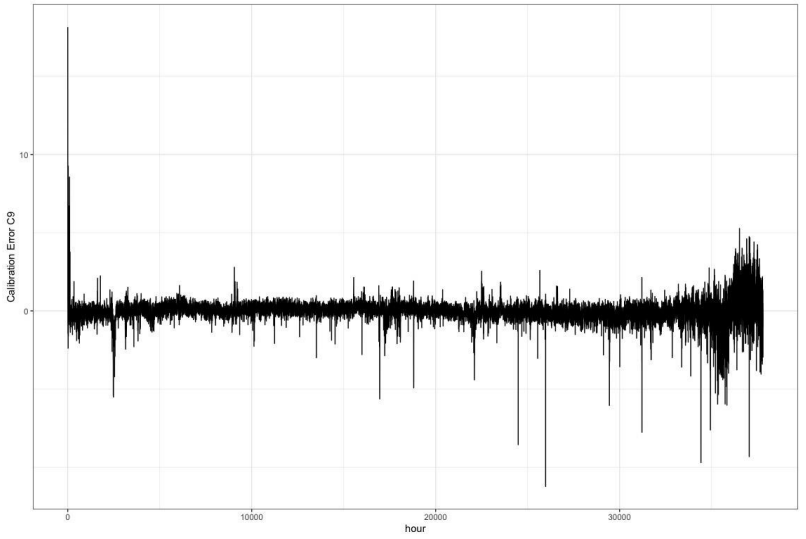
8	PM x RH x T	$PM_{2.5,\ corrected} = PM_{2.5} \times s_1 + RH \times s_2 + T \times s_3 + PM_{2.5} \times RH \times s_4 + PM_{2.5} \times T \times s_5 + RH \times T \times s_6 + PM_{2.5} \times RH \times T \times s_7 + b + \varepsilon$	0.90	6.2	3.1 (2.1)	
	Machine Learning (LOSO CV)					
9	Random Forest	$PM_{2.5,\ corrected} = f(PM_{2.5}, T, RH)$	0.99	2.4	1.3 (0.9)	

***Table S4**: Associations between the correction error and year of operation*

	Associations (95% Confidence Interval)								
Dataset	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**	**Model 8**	**Model 9**
All	-0.28* (-0.29, -0.27)	-0.12* (-0.13, -0.10)	-0.17* (-0.18, -0.16)	-0.08* (-0.10, -0.07)	-0.07* (-0.08, -0.06)	-0.16* (-0.17, -0.15)	-0.23* (-0.24, -0.22)	-0.04* (-0.05, -0.03)	-0.05* (-0.05, -0.05)
Climate Zone (Outside Devices Only)									
Cold	-0.26* (-0.29, -0.24)	-0.27* (-0.29, -0.25)	-0.19* (-0.21, -0.17)	-0.21* (-0.23, -0.19)	-0.20* (-0.22, -0.18)	-0.20* (-0.22, -0.17)	-0.22* (-0.24, -0.20)	-0.14* (-0.16, -0.12)	-0.08* (-0.08, -0.07)

Hot-Dry	0.03* (0.02, 0.05)	0.08* (0.06, 0.09)	0.09* (0.08, 0.11)	0.10* (0.09, 0.12)	0.13* (0.11, 0.14)	0.11* (0.09, 0.13)	0.04* (0.03, 0.06)	0.13* (0.11, 0.14)	0.01 (-0.00, 0.01)
Hot-Humid	-0.86* (-1.04, -0.68)	-0.92* (-1.10, -0.75)	-0.89* (-1.06, -0.71)	-0.93* (-1.10, -0.76)	-0.88* (-1.05, -0.71)	-0.89* (-1.06, -0.71)	-0.86* (-1.04, -0.69)	-0.90* (-1.06, -0.73)	-0.35* (-0.41, -0.28)
Marine	-0.40* (-0.43, -0.38)	-0.13* (-0.15, -0.10)	-0.27* (-0.29, -0.24)	-0.07* (-0.09, -0.04)	-0.07* (-0.10, -0.05)	-0.24* (-0.27, -0.22)	-0.29* (-0.32, -0.27)	-0.01 (-0.04, 0.01)	-0.02* (-0.03, -0.01)
Mixed-Dry	-0.93* (-1.03, -0.84)	-0.31* (-0.40, -0.21)	-0.53* (-0.63, -0.44)	-0.28* (-0.37, -0.18)	-0.27* (-0.36, -0.18)	-0.56* (-0.66, -0.47)	-0.83* (-0.92, -0.74)	-0.29* (-0.37, -0.20)	-0.34* (-0.37, -0.30)
Mixed-Humid	-0.46* (-0.51, -0.41)	-0.28* (-0.33, -0.23)	-0.32* (-0.37, -0.27)	-0.20* (-0.25, -0.16)	-0.20* (-0.25, -0.16)	-0.29* (-0.34, -0.24)	-0.39* (-0.44, -0.35)	-0.13* (-0.17, -0.08)	-0.01 (-0.03, 0.00)

(**p< 0.05*)

Table S5*: Results from regressing hour and interaction of hour and the cumulative number of high $PM_{2.5}$ measurements recorded on correction error*

	Coefficients (95% CI)		
Intercept	-1,566* (-1,774, -1,357)	867.2* (674.8, 1059.5)	801.9* (618.5, 985.3)
Hour	-0.18* (-0.19, -0.17)	-0.15* (-0.17, -0.14)	-0.05* (-0.06, -0.04)
Cumulative $PM_{2.5}$ > 50	29.5* (28.8, 30.1)		-
Cumulative $PM_{2.5}$ > 100	-	57.7* (54.8, 60.5)	-
Cumulative $PM_{2.5}$ > 500	-	-	864.4* (832.0, 896.7)
Hour: Cumulative $PM_{2.5}$ > 50	-9.0×10^{-4}* (-9.2×10^{-4}, -8.7×10^{-4})	-	-
Hour: Cumulative $PM_{2.5}$ > 100	-	-1.5×10^{-3}* (-1.7×10^{-3}, -1.4×10^{-5})	-
Hour: Cumulative $PM_{2.5}$ > 500	-	-	-0.06* (-0.07, -0.06)

(**p< 0.05*)

Table S6*: Results from regressing hour and interaction of hour and the cumulative number of high $PM_{2.5}$ measurements recorded on normalized correction error*

	Coefficients (95% CI)		
Intercept	7,306* (2,125, 12,488)	6,601* (1,842, 11,359)	8,263* (3,727, 12,799)
Hour	0.6*	0.5*	0.5*

	(0.3, 0.9)	(0.2, 0.8)	(0.2, 0.8)
Cumulative $PM_{2.5}$ > 50	2.9 (-14.1, 19.8)	-	-
Cumulative $PM_{2.5}$ > 100	-	90 (19, 160)	-
Cumulative $PM_{2.5}$ > 500	-	-	-238.9 (-1039.4, 561.6)
Hour: Cumulative $PM_{2.5}$ > 50	-1.6×10^{-4} $(-8.3 \times 10^{-4}, 5.1 \times 10^{-4})$	-	-
Hour: Cumulative $PM_{2.5}$ > 100	-	-2.6×10^{-3} $(-5.5 \times 10^{-3}, 0.2 \times 10^{-3})$	-
Hour: Cumulative $PM_{2.5}$ > 500	-	-	8.2×10^{-3} $(-5.0 \times 10^{-2}, 6.7 \times 10^{-2})$

(*$*p< 0.05$*)