SUPPLEMENTAL MATERIAL

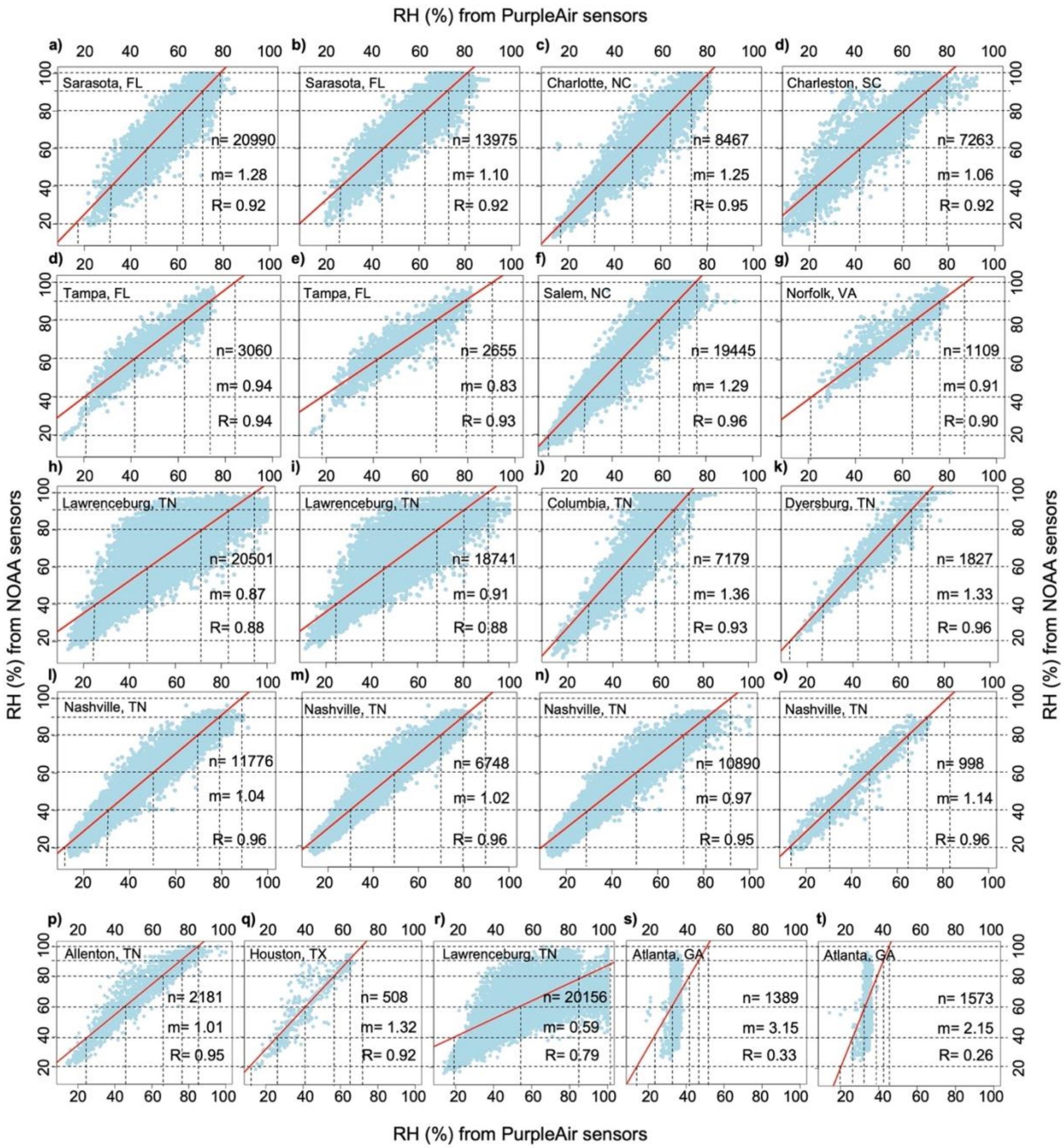

Figure S1: Correlation graphs between RH from each PA sensor and RH from the nearest NOAA sensor (average distance of ~10 miles (min: ~1.65, max: ~25.50 miles)). The 21 graphs correspond to the 21 PA sensors for the 0.5-km radius. R, n and m correspond to the estimated Pearson correlation, the number of data points and the slope of the linear regression.

Except for Figures S1-s and S1-t that present a very low Pearson correlation R, every individual PA displays a correlation R varying between 79% and 96% with 16/21 PA sensors presenting an

R equal or greater than 90%. As reported by recent studies (Tryner et al., 2020, Magi et al., 2020, Giordano et al., 2021; Barkjohn et al., 2022), PA sensors tend to report dryer humidity measurements than ambient conditions with a general difference of 10% to 20% for our sensors (Figure S2). However, Figure S1-r shows that one of our PA sensors reported more humid measurements for RH values of 80% or greater.

Moreover, the slope of the linear regression estimated for each PA sensor (Figure S1) shows that RH from Figures S1-r, S1-s and S1-t exhibit the larger bias metrics. Figure S1-r shows that RH from the PA sensor tend to underestimate ambient RH while PA sensors represented in Figures S1-s and S1-t tend to more than doubled or tripled RH values from NOAA.

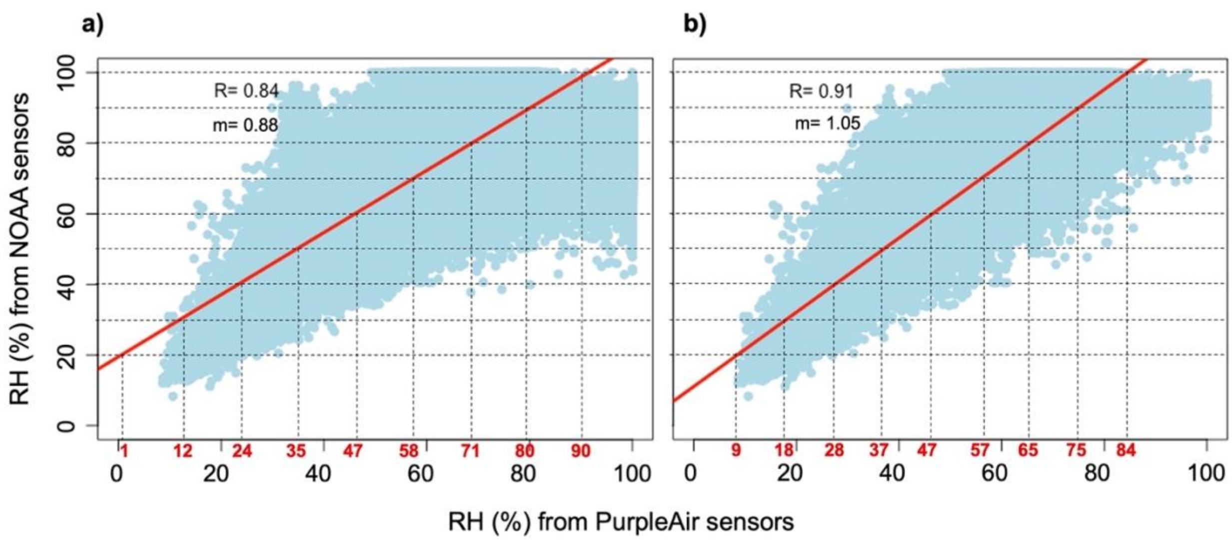

Figure S2: Correlation graphs between RH from PA sensors and RH from the nearest NOAA sensor to each PA sensor. a) All the 21 PA sensors for the 0.5-km radius, b) All PA sensors except r), s) and t) from Figure S1.

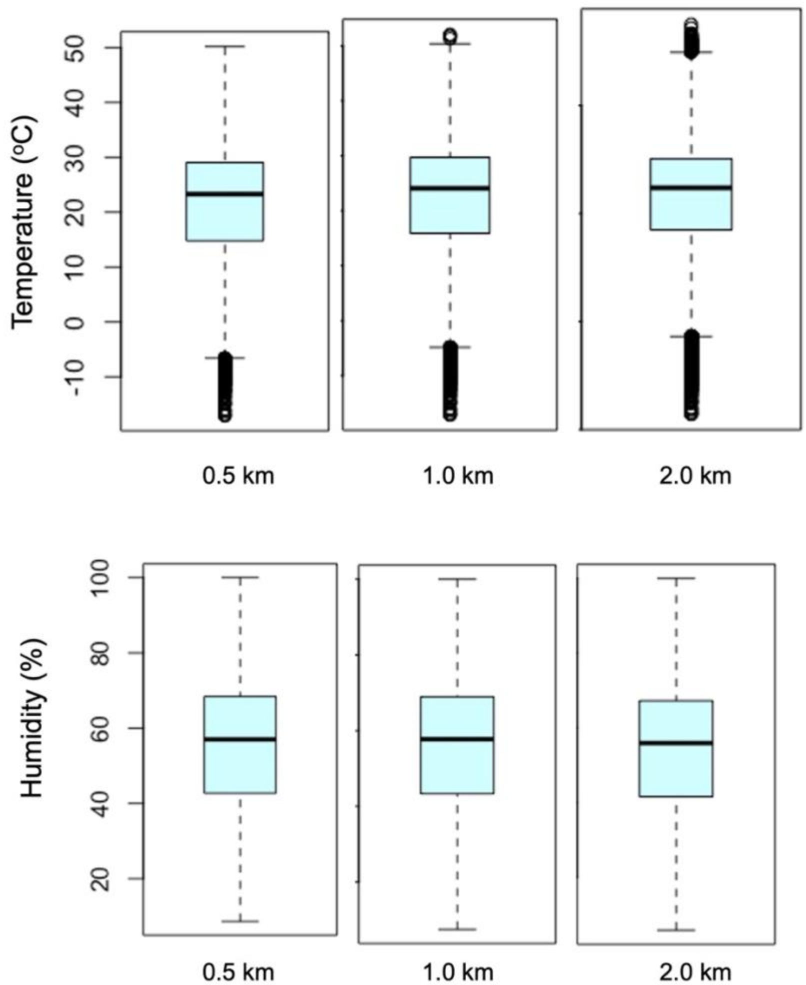

Figure S3: Distribution of RH and T from PurpleAir data for our three buffers (0.5 km, 1.0 km and 2.0 km)

Model fit using the 1.0-km radius dataset:

Table S1: MLR model development (model fit using hourly data) for the 1.0-km radius

Parameters		Model fit with hourly data			
Models		R^2 (%)	RMSE ($\mu g/m^3$)	MAE ($\mu g/m^3$)	R (%)
Model 1	$4.3410721+0.3796856\ PA_i$	58	3.85	2.47	76
Model 2	$6.9941051+0.3872666\ PA_i\ -0.0489237\ RH_i$	60	3.76	2.40	77
Model 3	$1.6915454+0.\ 3849136\ PA_i\ +0.\ 1149728\ T_i$	62	3.68	2.39	78

Model 4	4.1204142+0. 3907494 PA_i -0. 0405732 RH_i + 0. 1050501 T_i	63	3.61	2.32	79
Model Bj	5.72+0.524 PA_i -0.0852 RH_i	60	4.40	2.94	77

Table S2: SSC model development for the 1.0-km radius

Parameters		**Model fit with hourly data**			
Clusters (Number of observations)	**Models**	**R^2 (%)**	**RMSE (μg/m³)**	**MAE (μg/m³)**	**R (%)**
RH ≤ 50 (85616)	2.782329 + 0.368994 PA_i - 0.010616 RH_i + 0.122888 T_i	57	3.94	2.43	75
RH >50 (152431)	5.1538241 + 0.3980145 PA_i - 0.0539108 RH_i + 0.0943790 T_i	67	3.40	2.26	82

Model fit using the 2.0-km radius dataset:

Table S3: MLR model development (model fit using hourly data) for the 2.0-km radius

Parameters		**Model fit with hourly data**			
Models		**R^2 (%)**	**RMSE (μg/m³)**	**MAE (μg/m³)**	**R (%)**
Model 1	4. 7265899 +0. 3763097 PA_i	55	4.30	2.72	74
Model 2	7. 5916138 +0. 3844184 PA_i -0. 0545752 RH_i	58	4.19	2.63	76
Model 3	1. 7548043 +0. 3803865 PA_i +0. 1250425 T_i	59	4.13	2.62	77
Model 4	4. 3418026 +0. 3862249 PA_i -0. 0425548 RH_i + 0. 1101893 T_i	60	4.06	2.55	78
Model Bj	5.72+0.524 PA_i -0.0852 RH_i	58	4.84	3.10	76

Table S4: SSC model development for the 2.0-km radius

Parameters		Model fit with hourly data			
Clusters (Number of observations)	Models	R^2 (%)	RMSE (µg/m³)	MAE (µg/m³)	R (%)
RH ≤ 50 (154276)	2.6452739 + 0.3676529 PA_i - 0.0057266 RH_i + 0.1303605 T_i	54	4.45	2.74	74
RH >50 (239734)	6.0381100 + 0.3926179 PA_i - 0.0646265 RH_i + 0.0961319 T_i	65	3.77	2.42	80

Cross-validation using LGOCV for the 0.5-km radius

Table S5: Cross-validation results – MLR models

Parameters		Model fit with hourly data			
Models		R^2 (%)	RMSE (µg/m³)	MAE (µg/m³)	R (%)
Model 1	3.6667550+0.4053418 PA_i	69	3.19	2.13	83
Model 2	6.3384228+0.4143437 PA_i -0.0506037 RH_i	71	3.06	2.05	84
Model 3	1.7642336+0.4109897 PA_i +0.0847196 T_i	71	3.05	2.06	84
Model 4	4.3295358+0.4182906 PA_i -0.0445768 RH_i + 0.0752867 T_i	73	2.95	1.98	85

Table S6: Cross-validation – SSC models

Parameters		Model fit with hourly data			
Clusters (Number of observations)	Models	R^2 (%)	RMSE (µg/m³)	MAE (µg/m³)	R (%)
RH ≤ 50 (59405)	2.738732 + 0.425834 PA_i - 0.008944 RH_i + 0.079210 T_i	71	2.93	1.86	84

RH >50 (100243)	7.230374 + 0.412683 PA_i - 0.085278 RH_i + 0.070655 T_i	74	2.92	2.02	86

Table S7: Methods evaluated to determine the optimal number of clusters using NbClust

#	**Methods**	**Number of clusters**	**Value Index**
1	KL	6	12.525
2	CH	5	2455.144
3	Hartigan	5	1479.653
4	CCC	2	86.151
5	Scott	5	2750.553
6	Marriot	5	349227082927.000
7	TrCovW	3	107827250739.000
8	TraceW	5	135580.700
9	Friedman	14	52.457
10	Rubin	5	11.169
11	Cindex	6	0.253
12	DB	2	0.882
13	Silhouette	2	0.384
14	Duda	2	0.723
15	PseudoT2	2	1043.876
16	Beale	2	0.381
17	Ratkowsky	4	0.379
18	Ball	3	195959.500
19	PtBiserial	5	0.520
20	Frey	2	2.381
21	McClain	2	0.186
22	Dunn	13	0.010
23	Hubert	0	0.000
24	SDindex	4	0.203
25	Dindex	0	0.000
26	SDbw	14	0.280

Among all indices:
Eight (8) proposed 2 as the best number of clusters.
Two (2) proposed 3 as the best number of clusters.

Two (2) proposed 4 as the best number of clusters.
Seven (7) proposed 5 as the best number of clusters.
Two (2) proposed 6 as the best number of clusters.
One (1) proposed 13 as the best number of clusters.
Two (2) proposed 14 as the best number of clusters.

Conclusion: According to the majority rule, the best number of clusters is 2.

Model fit using NOAA RH and T for the 0.5-km radius dataset:

Table S8: MLR and SSC model development (model fit using hourly data) for the 0.5-km radius

Parameters		**Model development**				**Sensitivity analysis**			
Models		**R^2 (%)**	**RMSE (μg/m^3)**	**MAE (μg/m^3)**	**R (%)**	**R^2 (%)**	**RMSE (μg/m^3)**	**MAE (μg/m^3)**	**R (%)**
MLR	4.4968840+0.4184462 PA_i - 0.0353587 RH_i + 0.0779764 T_i	72	2.99	2.01	85	78	2.28	1.62	88
SSC (RH ≤ 50)	2.874778+0.461934 PA_i - 0.009394 RH_i + 0.077146 T_i	76	2.72	1.74	87	85	1.94	1.32	92
SSC (RH > 50)	5.6571930+0.4101217 PA_i - 0.0472842 RH_i + 0.0724931 T_i	72	3.03	2.05	85	77	2.31	1.64	88