Air Emory Initial Data Validation Report (September – November)

Saikawa Lab at Emory University

February 2022

This report contains a table of data quality indicators and their calculated values for $PM_{2.5}$, PM_{10}, and CO. Visualizations of the regression models and time series data are presented following the data tables.

Eight Airly sensors, including five PM sensors (ID# 7194, 7276, 7284, 7304) and three PM+gas sensors (ID# 7160, 7205, 7250) were collocated with the federal regulatory monitor (FRM) at the South Dekalb site in Georgia from 09/06/2021 00:00:00 EST to 11/08/2021 17:00:00 EST.

Three Dylos sensors that measure PM concentrations were collocated with federal regulatory monitor (FRM) at the South Dekalb site in Georgia from 10/05/2021 00:00:00 EST to 11/09/2021 23:59:59 EST.

Table 1. Data Quality Indicators (DQI)

DQI	Quality control activities and checks	Equation for DQI
Precision	Variation around the mean of a set of measurements reported concurrently by three or more sensors of the same type collocated under the same sampling conditions. Precision is estimated using the standard deviation (SD) and coefficient of variation (CV).	$\mathrm{SD} = \sqrt{\frac{1}{(N \times M) - 1}\sum_{j=1}^{M}\left[\sum_{d=1}^{N}(x_{dj} - \overline{x_d})^2\right]}$ $\mathrm{CV} = \frac{\mathrm{SD}}{\overline{x}} \times 100$ N = number of 1-hour periods during which all identical instruments are operating and returning valid averages over the duration of the field test M = number of identical sensors operated simultaneously during a field test x_{dj} = 24-hour averaged sensor concentration for day d and sensor j $\overline{xd}$ = 24-hour averaged sensor concentration for day d $\overline{x}$ = deployment averaged sensor concentration for a field test
Bias	The systematic (non-random) or persistent disagreement between the concentrations reported by the sensor and reference instruments. Bias is determined using the linear regression slope and intercept.	Using a simple linear regression model (y = mx + b) with the sensor measurements as a dependent variable (y) and the FRM/FEM measurements as an independent variable (x), calculate the slope (m), intercept (b), and the coefficient of determination (R^2).
Linearity	A measure of the extent to which the measurements reported by a sensor are able to explain the concentrations reported by the reference instrument. Linearity is determined using the coefficient of determination (R^2).	

Error	A measure of the disagreement between the pollutant concentrations reported by the sensor and the reference instrument. The error is etimated using the root mean square error (RMSE) and normalized root mean square error (NRMSE).	$\mathrm{RMSE} = \sqrt{\frac{1}{N \times M}\sum_{j=1}^{M}\left[\sum_{d=1}^{N}(x_{dj} - R_d)^2\right]}$ $\mathrm{NRMSE} = \frac{\mathrm{RMSE}}{\overline{R_d}} \times 100$ *N* = number of 1-hour periods during which all identical instruments are operating and returning valid averages over the duration of the field test *M* = number of identical sensors operated simultaneously during a field test x_{dj} = valid 24-hour averaged sensor concentration for day *d* and instrument *j* R_d = valid 24-hour averaged FRM/FEM concentration for day *d* $\overline{Rd}$ = valid 24-hour averaged FRM/FEM concentration over the entire testing period
Accuracy	Accuracy is the degree of closeness between the sensors' measured values and the reference value. This is optional and used as a complement to the performance metric of bias.	$A\ \% = 100 - \frac{\lvert X - R \rvert}{R} \times 100$ *X* is the average concentration measured by the sensors throughout testing period and *R* is the average concentration measured by the official air quality monitoring station during the testing period
Inter-sensor variability (%)	Evaluate how close the measurements from the units of the same sensor type are to each other. This is used in the data validation for the co-located sensors.	$Inter - sensor\ variability = \frac{Mean.highest - Mean.lowest}{Mean.average} \times 100$

Table 2. a) Precision; b) Bias and Linearity and c) Error and Accuracy for $PM_{2.5}$

a)	Precision (between collocated sensors)		Data Quality		
	CV (%)	SD ($\mu g/m^3$)	Uptime (%)	Number of concurrently reported sensor concentration values	
Metric Target Range	≤ 30	≤ 5	90%*		
Airly Deployment Value	11.8	1.70	100	64	
Dylos Deployment Value	23.0	2.97	100	36	
b)	Bias and Linearity			Data Quality	
	R^2	Slope	Intercept (b) ($\mu g/m^3$)	Uptime (%)	Number of paired sensor and FRM/FEM
Metric Target Range	≥ 0.70	1.0 ± 0.35	-5 ≤ b ≤ 5	90%*	
Airly 7160	0.947	1.92	-5.84	100	64
Airly 7194	0.954	1.98	-6.01	100	64
Airly 7205	0.957	1.98	-4.90	100	64
Airly 7250	0.951	2.07	-4.74	100	64
Airly 7276	0.954	2.21	-5.94	100	64
Airly 7284	0.957	2.20	-5.33	100	64
Airly 7304	0.960	2.21	-5.85	100	64
Airly 7396	0.944	2.42	-5.24	100	64

Airly Mean	0.953	2.12	-5.48	100	64
Number of Airly Meeting Target Value	**8**	**0**	**2**	**8**	
Dylos 1	0.392	0.949	5.68	100	36
Dylos 2	0.470	0.936	2.51	100	36
Dylos 3	0.355	1.08	6.10	100	36
Dylos mean	0.492	0.988	4.76	100	36
Number of Sensors Meeting Target Value	**0**	**3**	**1**	**3**	

c)	RMSE ($\mu g/m^3$)	NRMSE (%)	Accuracy (%)	Inter-sensor Variability (%)
Metric Target Range	<= 7	<= 30	The closer to 100% the better, but a value of 80% and above will be considered acceptable.	
Airly	6.95	74.2	46.6	36.8
Dylos	6.33	75.0	44.5	36.8

Table 3. a) Precision; b) Bias and Linearity and c) Error and Accuracy for PM_{10}

a)	Precision (between collocated sensors)		Data Quality	
	CV (%)	SD ($\mu g/m^3$)	Uptime (%)	Number of concurrently reported sensor concentration values
Metric Target Range	≤ 30	≤ 21	90%*	
Airly Deployment Value	11.7	1.80	100	64

b)	Bias and Linearity			Data Quality	
	R^2	Slope	Intercept (b) ($\mu g/m^3$)	Uptime (%)	Number of paired sensor and FRM/FEM
Metric Target Range	≥ 0.70	1.0 ± 0.3	$-21 \leq b \leq 21$	90%*	
Airly 7160	0.641	1.23	-9.40	100	64
Airly 7194	0.649	1.26	-10.1	100	64
Airly 7205	0.644	1.28	-9.04	100	64
Airly 7250	0.638	1.34	-9.24	100	64
Airly 7276	0.649	1.49	-11.3	100	64
Airly 7284	0.645	1.45	-10.4	100	64
Airly 7304	0.638	1.42	-10.6	100	64
Airly 7396	0.636	1.60	-11.3	100	64
Airly Mean	0.643	1.38	-10.2	100	64
Number of Airly Meeting Target Value	**8**	**3**	**8**	**8**	

c)	RMSE ($\mu g/m^3$)	NRMSE (%)	Accuracy (%)	Inter-sensor Variability (%)
Metric Target Range	<= 7	<= 30	The closer to 100% the better, but a value of 80% and above will be considered acceptable.	
Airly	6.72	36.2	83.1	33.4

Table 4. a) Precision; b) Bias and Linearity and c) Error and Accuracy for CO

a)	Precision (between collocated sensors)		Data Quality	
	CV (%)	SD ($\mu g/m^3$)	Uptime (%)	Number of concurrently reported sensor concentration values
Metric Target Range	≤ 30		90%*	
Airly Deployment Value	15.1	0.0423	100	50

b)	Bias and Linearity			Data Quality	
	R^2	Slope	Intercept (b) (ppm)	Uptime (%)	Number of paired sensor and FRM/FEM
Metric Target Range	≥ 0.80	1.0 ± 0.2	-0.643 ≤ b ≤ 0.643	90%*	
Airly 7160	0.587	0.373	0.166	100	64
Airly 7205	0.342	0.386	0.107	78	50
Airly 7250	0.695	0.528	0.140	100	64
Airly Mean	0.541	0.429	0.138	92.7	59.3
Number of Airly Meeting Target Value	**0**	**0**	**3**	**3**	

c)	RMSE (ppm)	Accuracy (%)	Inter-sensor Variability (%)
Metric Target Range	<= 0.640	The closer to 100% the better, but a value of 80% and above will be considered acceptable.	
Airly	0.129	86.7	28.9

Note: The GAS-type sensor 7205 was not available until 09/20/2021 due to an issue with software version. It only has 50 data points in the report. The precision, RMSE, accuracy and inter-sensor variability are therefore calculated only for the time between 9/20/2021 and 11/08/2021 during when all the GAS-type sensors had data.

The following linear regression model was also used to correct CO data, considering the effect of ambient temperature (Temp) and relative humidity (RH). The following multiple linear regression model was established: $\boldsymbol{CO_{Airly} = a_1 CO_{RFM} + a_2 T + a_3 RH + b}$

Table 5. Bias and Linearity and for CO

		Bias and Linearity				Data Quality	
	R^2	Slope a1	Slope a2	Slope a3	Intercept (b) (ppm)	Uptime (%)	Number of paired sensor & FRM/FEM
Metric Target Range	≥ 0.80	1.0 ± 0.2			-0.643 ≤ b ≤ 0.643	90%*	
Airly 7160	0.674	0.379	-0.00454	-0.000569	0.297	100	64
Airly 7205	0.631	0.436	-0.0143	0.00108	0.237	78	50
Airly 7250	0.706	0.522	-0.00203	-0.000822	0.252	100	64
Airly Mean	0.670	0.446	-0.00696	-0.000104	0.262	92.7	59.3
Number of Sensors Meeting Target Value	0	0			3	3	

Figures

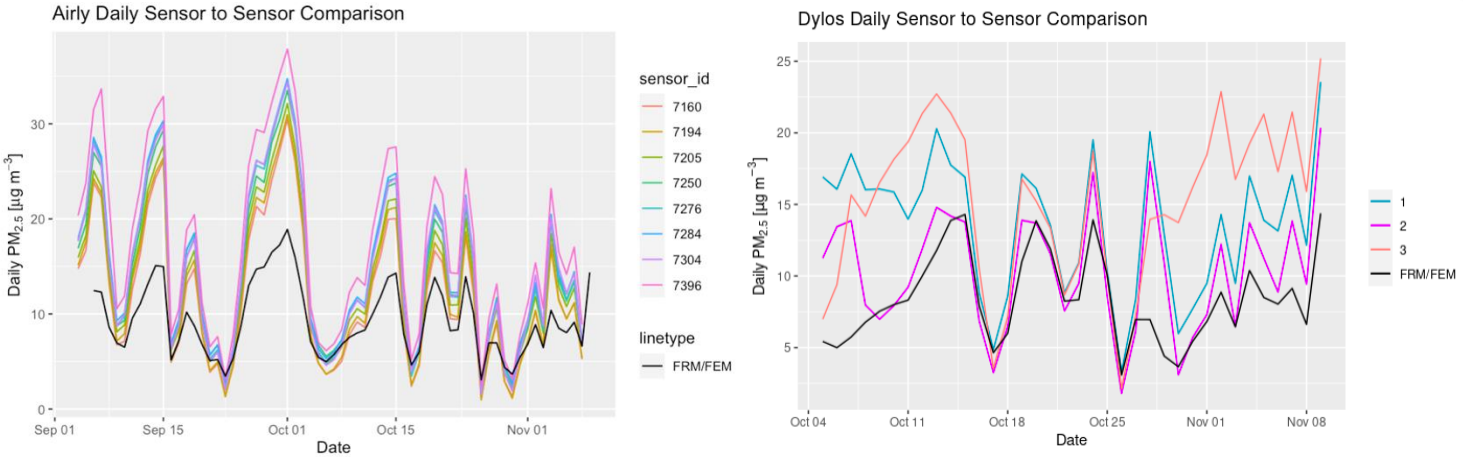

Figure 1. Daily $PM_{2.5}$ concentrations comparison of a) Airly and b) Dylos to FRM/FEM

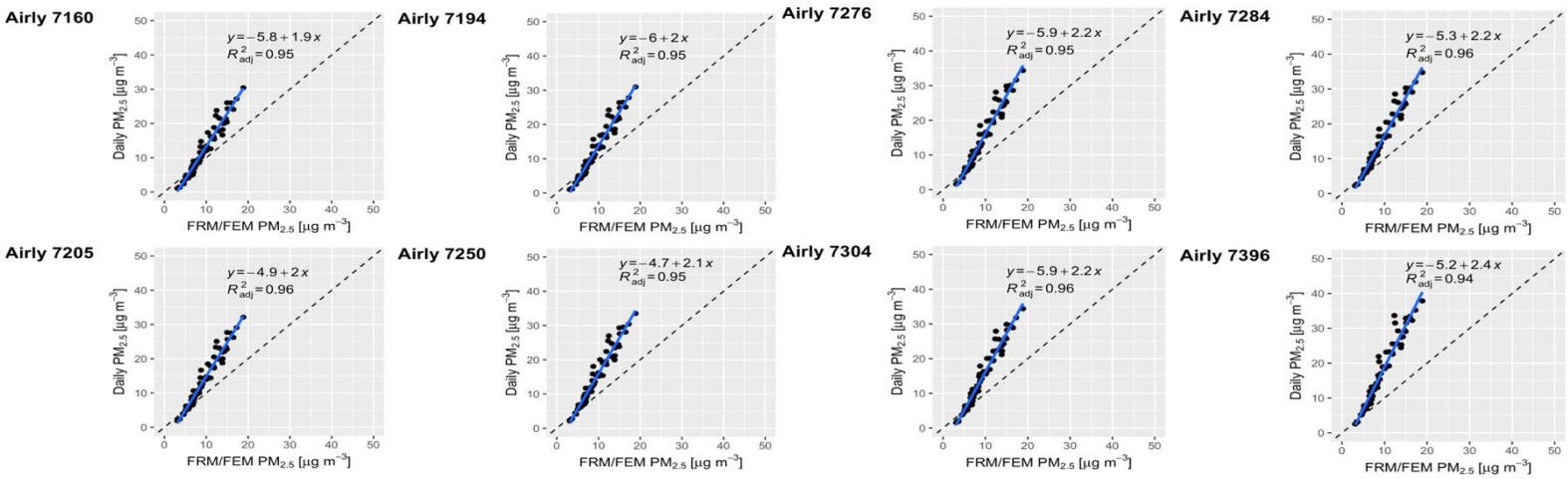

Figure 2. Daily $PM_{2.5}$ concentrations comparison of each Airly to FRM/FEM

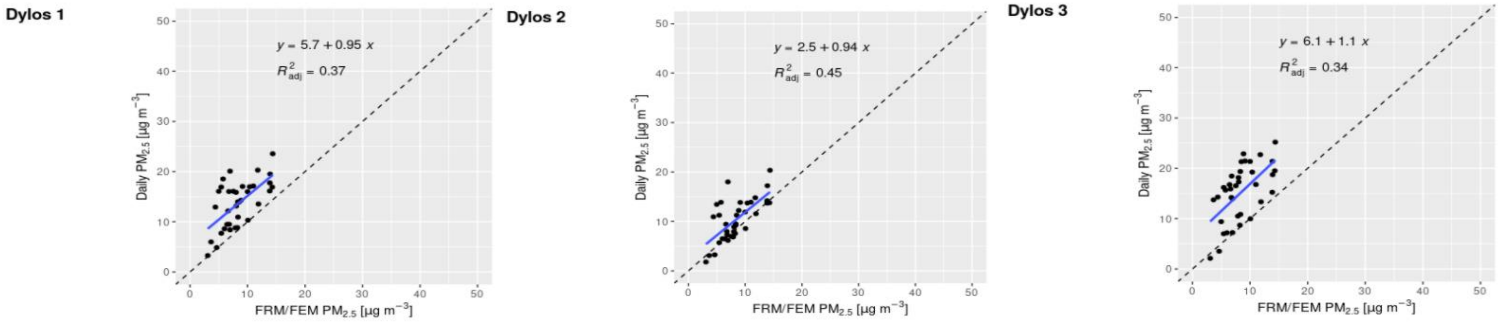

Figure 3. Daily $PM_{2.5}$ concentrations comparison of each Dylos to FRM/FEM

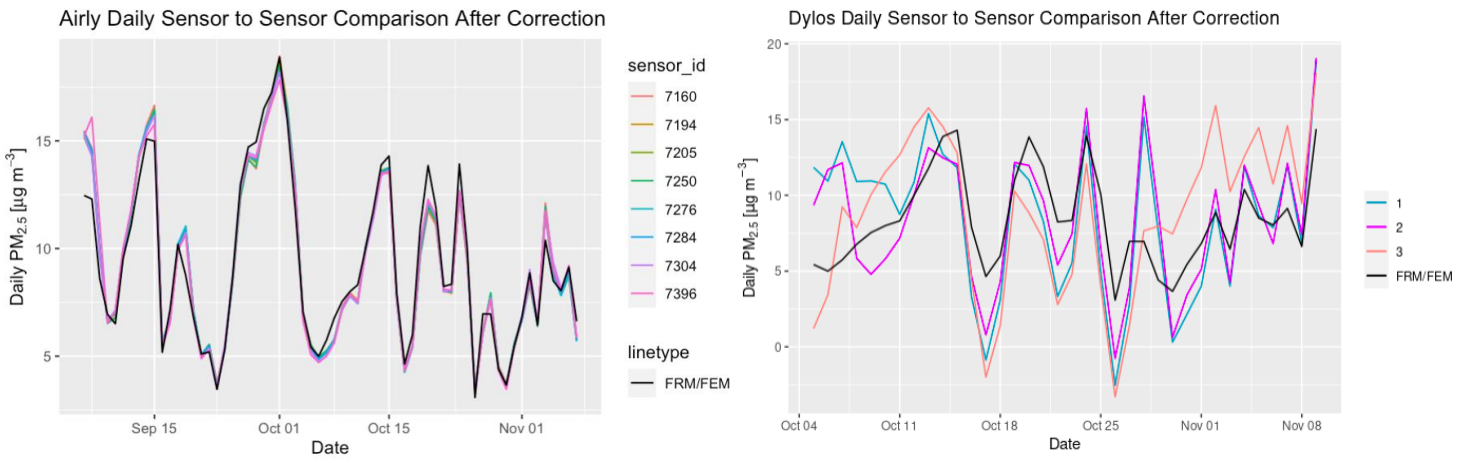

Figure 4. Daily $PM_{2.5}$ concentrations comparison of Airly (left) and Dylos (right) to FRM/FEM after bias correction

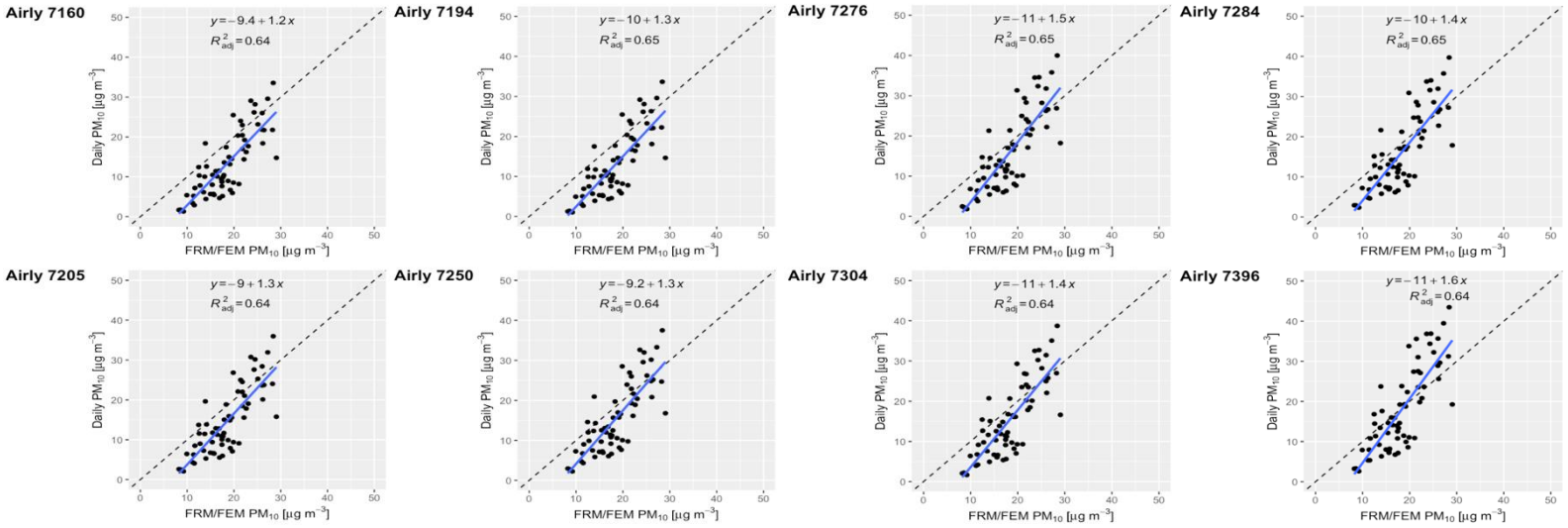

Figure 5. Daily PM_{10} concentrations comparison of each Airly to FRM/FEM

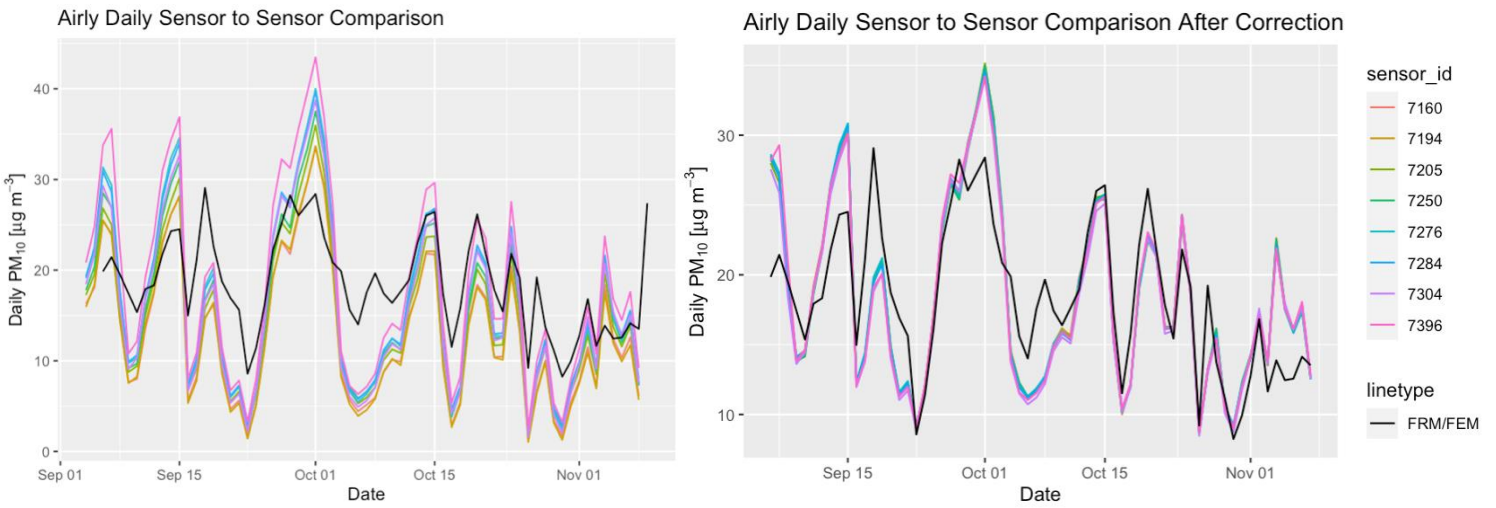

Figure 6. Daily PM_{10} concentrations comparison of Airly to FRM/FEM before (left) and after (right) bias correction

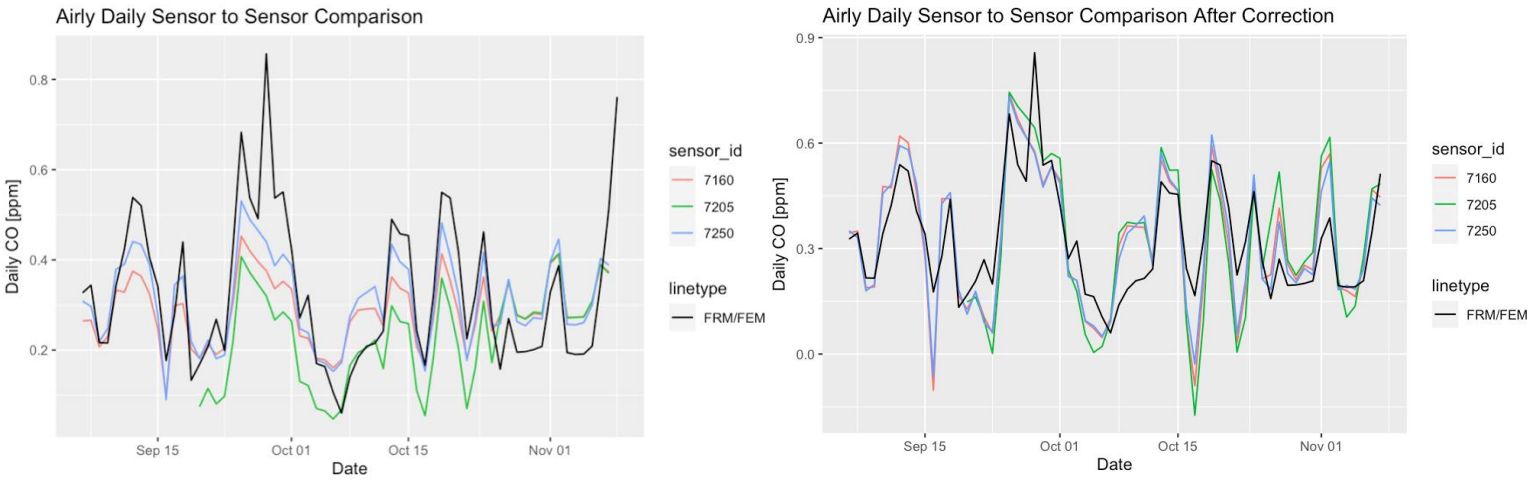

Figure 7. Daily CO concentrations comparison of Airly to FRM/FEM before (left) and after bias correction

Meteorological Effects

Understanding Meteorological Influence on Sensors

- **Relative frequency** is the percentage of a dataset corresponding to a particular measurement. For example, if 10°C has a relative frequency of 25%, it means that a temperature of 10°C was recorded 25% of the time throughout the entire monitoring period. Relative frequency is calculated using this formula, where M_i is a particular measurement of temperature or relative humidity:

$$Relative\ Frequency\ = \frac{\#\ of\ occurrences\ of\ M_i}{Total\ \#\ of\ measurements}$$

- **Normalized $PM_{2.5}$** is the ratio of $PM_{2.5}$ concentrations between the low-cost sensors and federal monitor. A value greater than one indicates that the low-cost sensor has overestimated $PM_{2.5}$ concentrations. A value less than one indicates an underestimate. A value of exactly one indicates that the low-cost sensor's observational value is equal to that of the federal monitor. A horizontal line of y = 1 (shown in gray on the $PM_{2.5}$ graphs) represents an ideal correlation, where temperature and humidity have no effect on low-cost sensor measurements. The formula for normalized $PM_{2.5}$ is the following:

$$PM_{2.5,\,norm}\ =\ \frac{PM_{2.5\ (Airly\ or\ Dylos)}}{PM_{2.5\ (FRM)}}$$

- The monitoring period for Airly sensors is 09/06/2021 (00:00 EST) to 11/08/2021 (17:00 EST), a total of 64 days. The monitoring period for Dylos sensors is 10/05/2021 (00:00 EST) to 11/09/2021 (23:59 EST), a total of 36 days.

- Federal monitors and Dylos sensors take measurements every minute, while Airly sensors take measurements every five minutes. For all sensors, measurements are averaged over 24 hours.

- The functional temperature and humidity ranges for Airly sensors are -40°C — +80°C and 0 — 100%, respectively. The ranges for Dylos are not specified by the manufacturer.

Number of 24-hour periods outside manufacturer-listed temperature target criteria	0
Number of 24-hour periods outside manufacturer-listed relative humidity target criteria	0

Airly

Number of paired, normalized concentration and temperature values	64 per sensor, 512 total
Number of paired, normalized concentration and humidity values	64 per sensor, 512 total

Dylos

Number of paired, normalized concentration and temperature values	36 per sensor, 108 total
Number of paired, normalized concentration and humidity values	36 per sensor, 108 total

Meteorological Conditions During Deployment

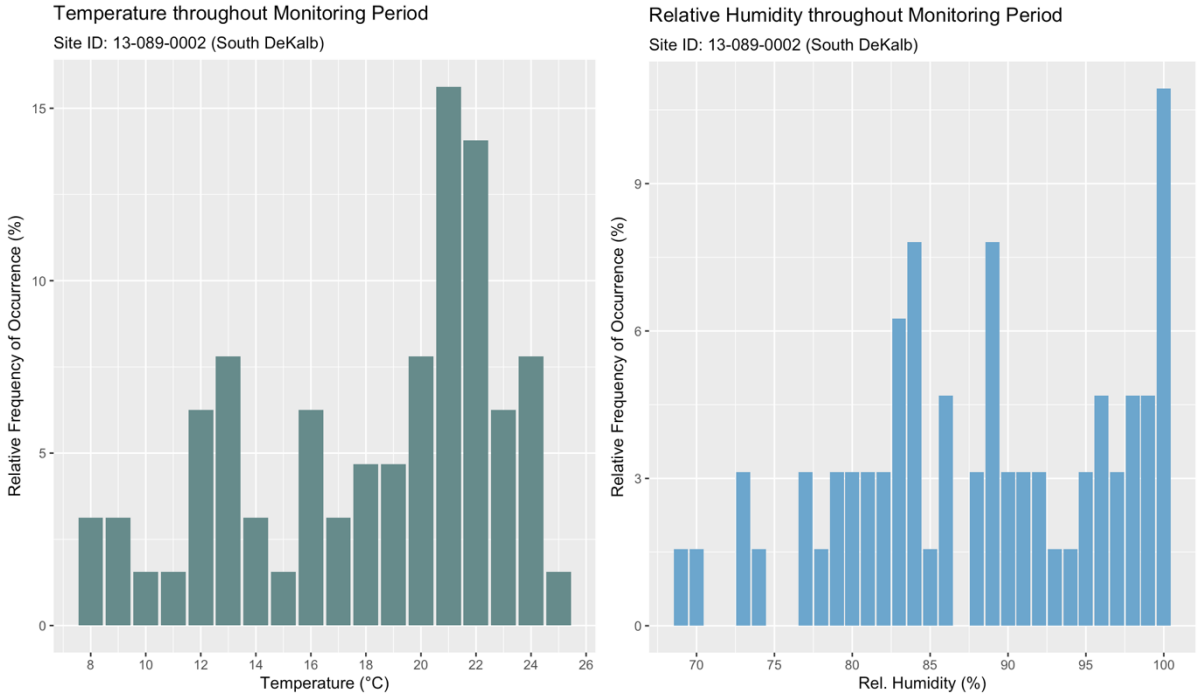

Figure 8. Relative frequency of temperature and relative humidity observed in FRM during the monitoring period

Meteorological Influence: Airly

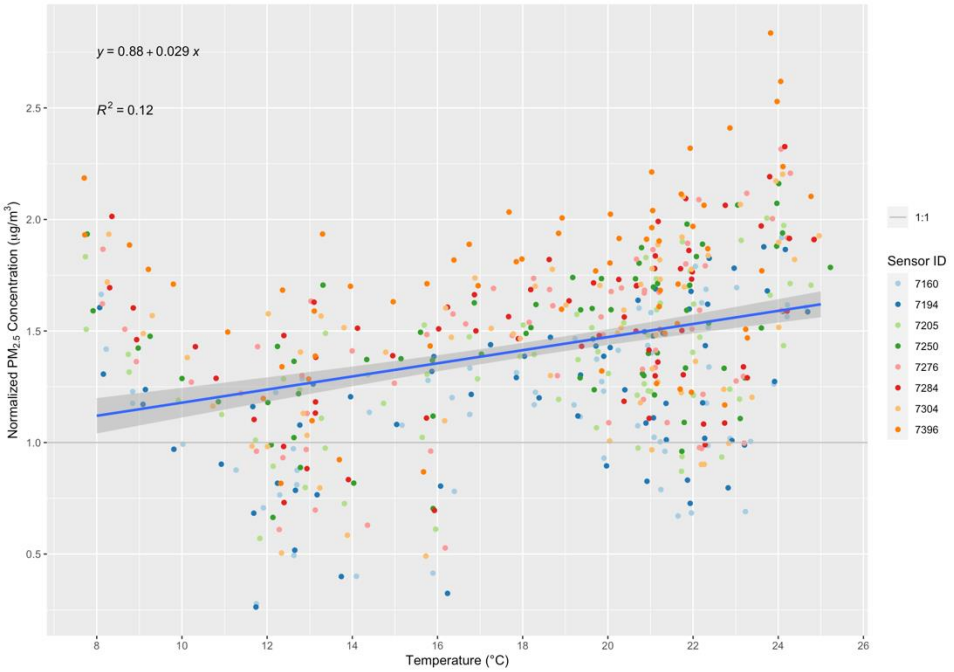

Figure 9. Temperature and normalized $PM_{2.5}$ concentrations for Airly sensors

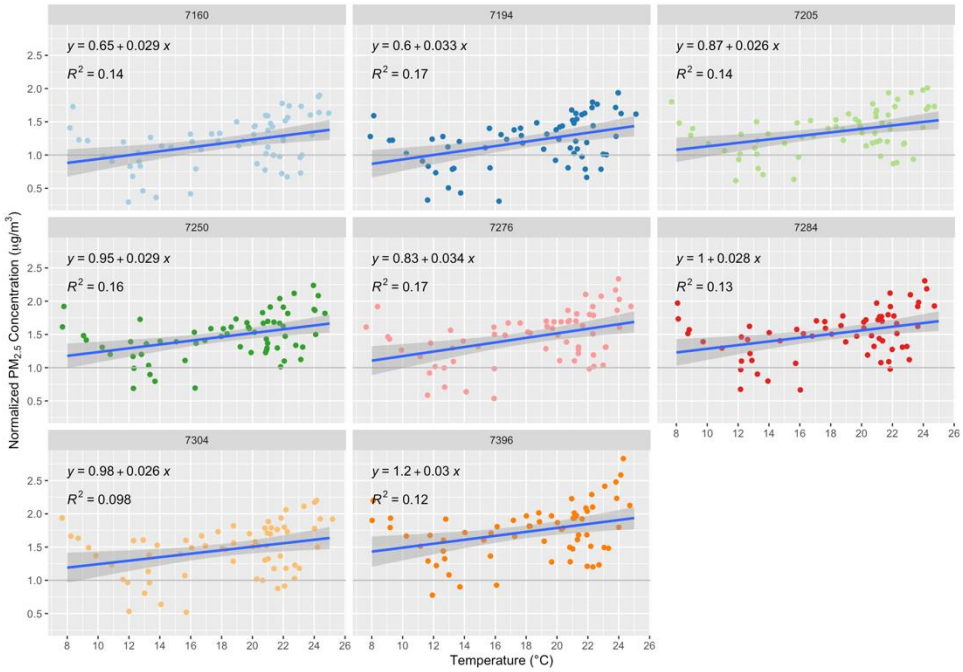

Figure 10. Temperature and normalized $PM_{2.5}$ concentrations for each Airly sensor

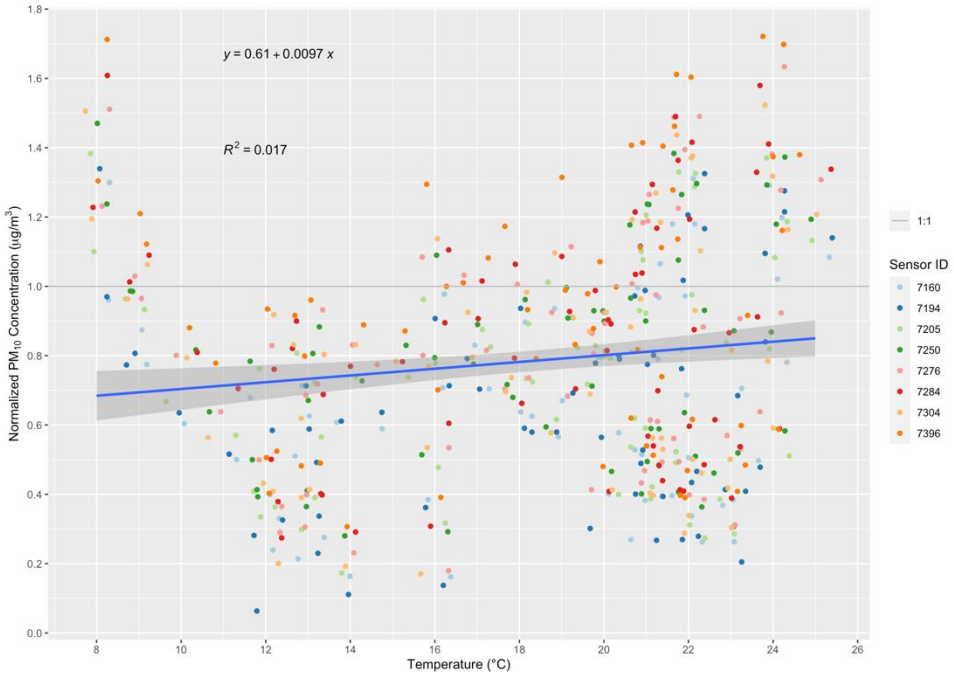

Figure 11. Temperature and normalized PM_{10} concentrations for Airly sensors

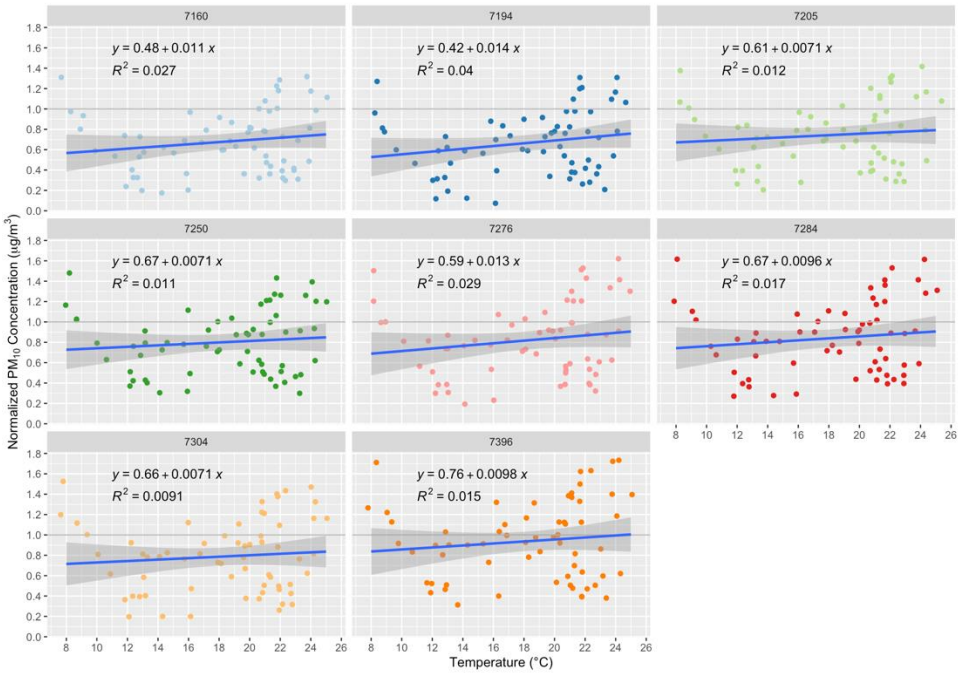

Figure 12. Temperature and normalized PM_{10} concentrations for each Airly sensor

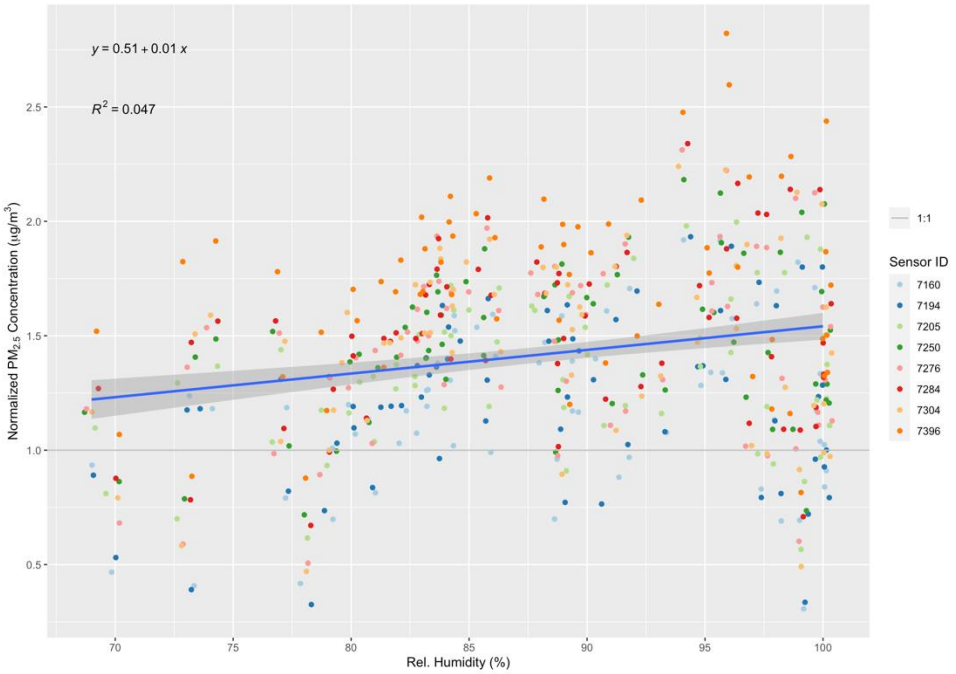

Figure 13. Relative humidity and normalized $PM_{2.5}$ concentrations for Airly sensors

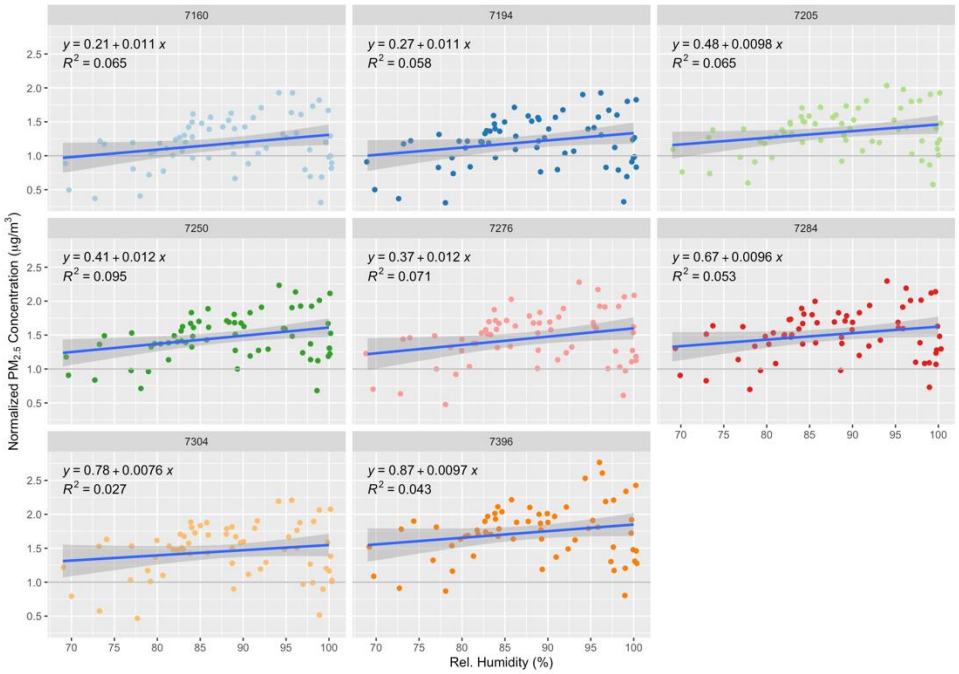

Figure 14. Relative humidity and normalized $PM_{2.5}$ concentrations for each Airly sensor

Figure 15. Relative humidity and normalized PM_{10} concentrations for Airly sensors

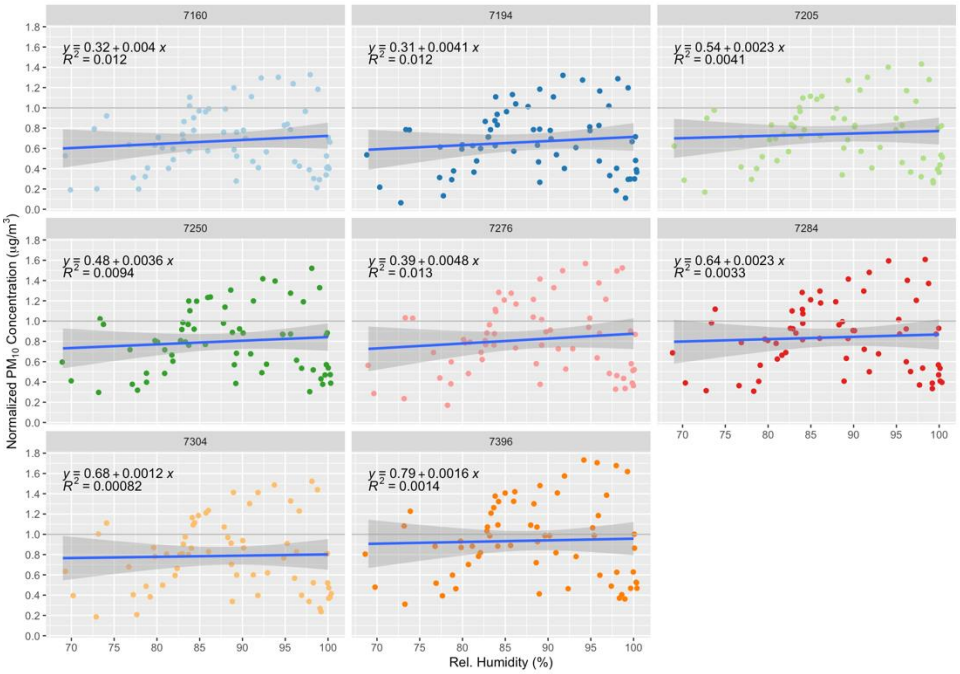

Figure 16. Relative humidity and normalized PM_{10} concentrations for each Airly sensor

Meteorological Influence: Dylos

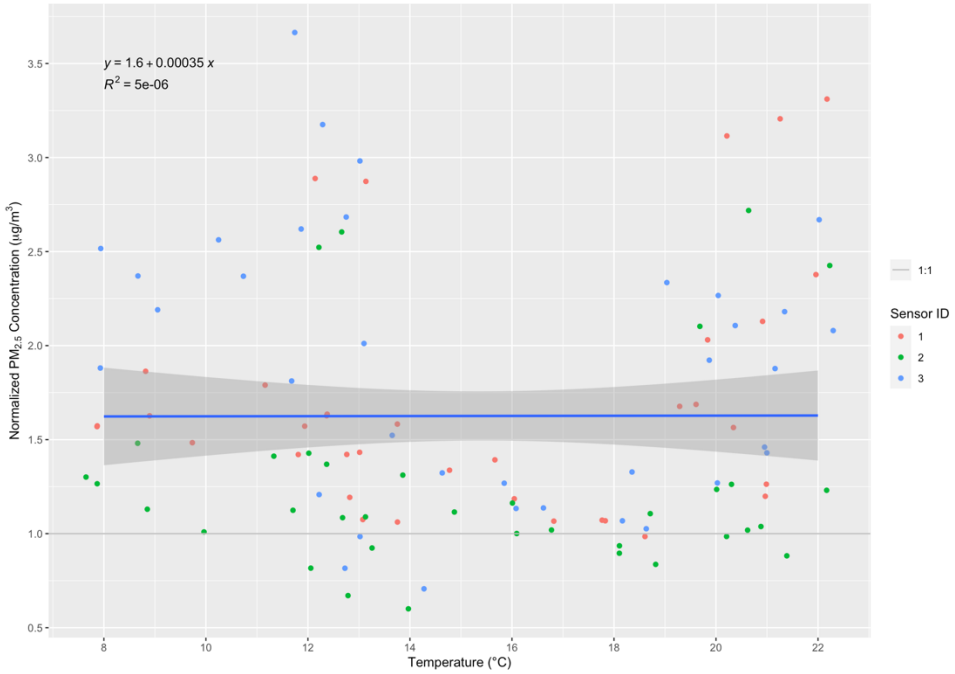

Figure 17. Temperature and normalized $PM_{2.5}$ concentrations for Dylos sensors

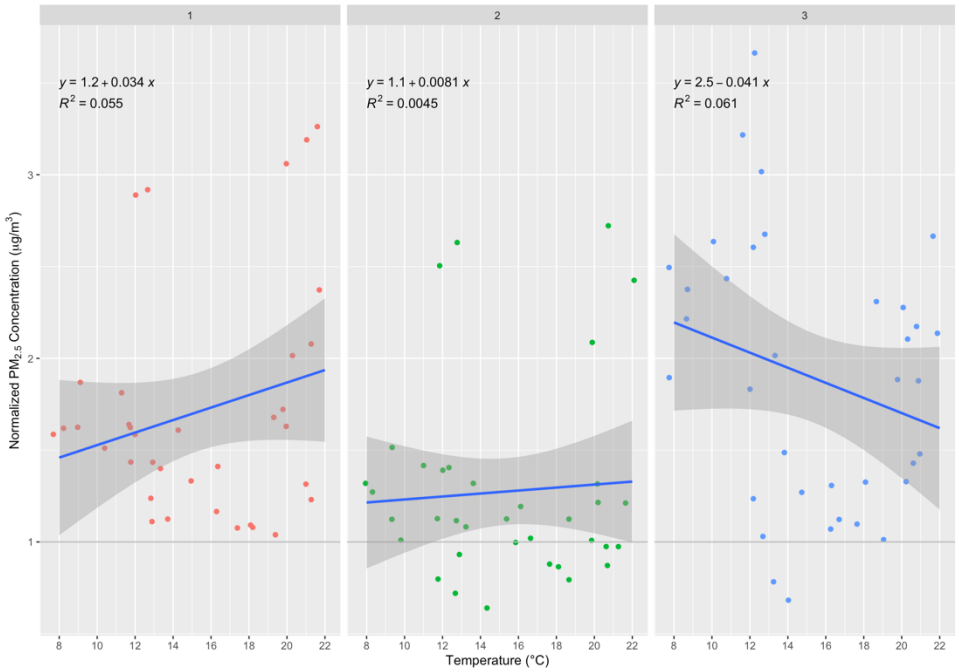

Figure 18. Temperature and normalized $PM_{2.5}$ concentrations for each Dylos sensor

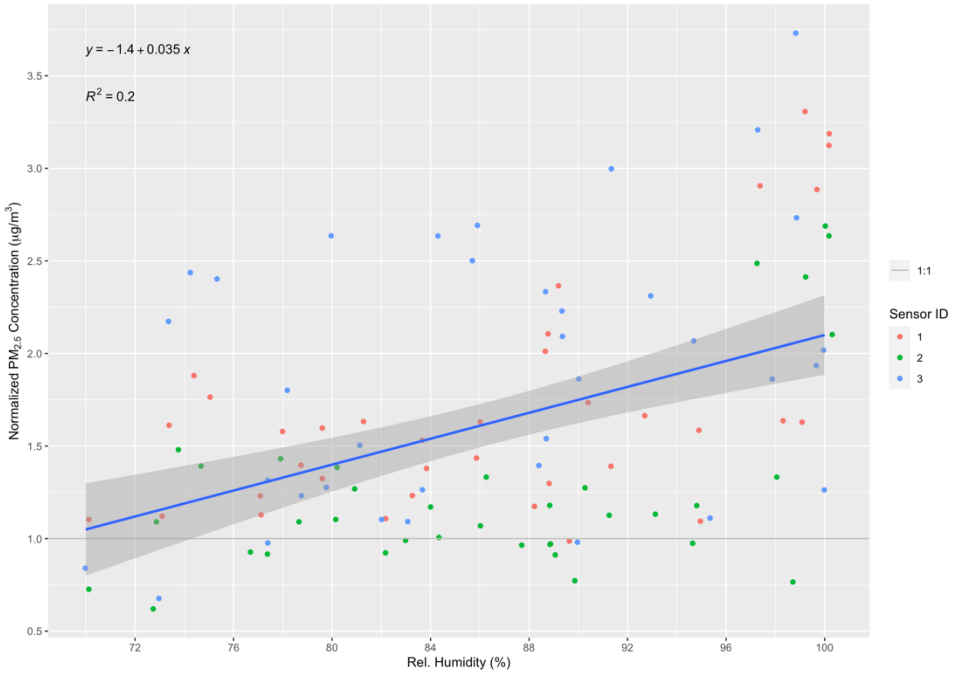

Figure 19. Relative humidity and normalized $PM_{2.5}$ concentrations for Dylos sensors

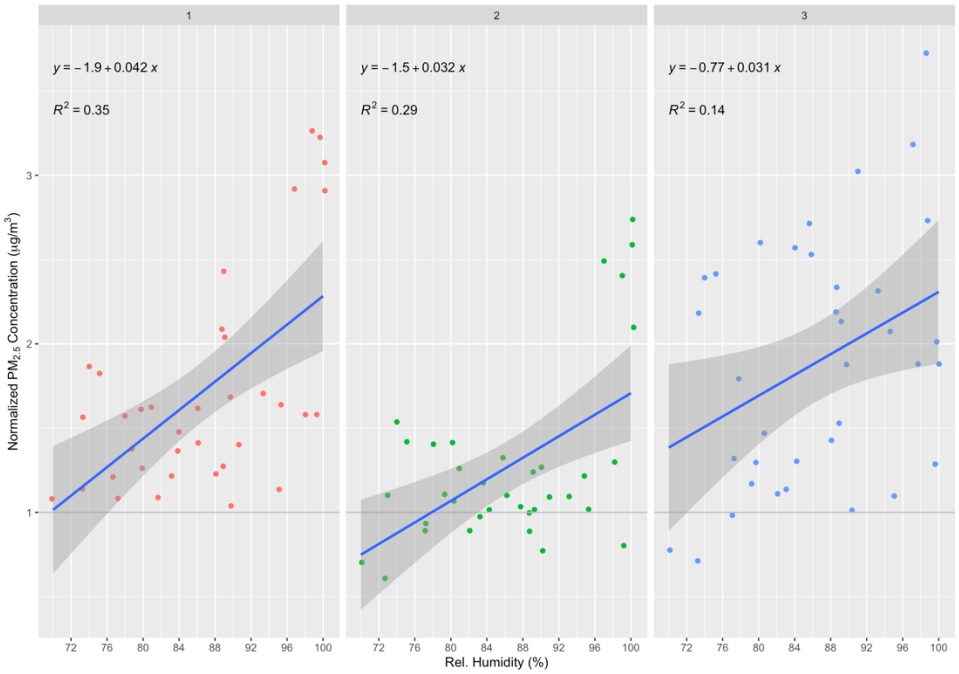

Figure 20. Relative humidity and normalized $PM_{2.5}$ concentrations for each Dylos sensor