



Supplement of

Spectral analysis approach for assessing the accuracy of low-cost air quality sensor network data

Vijay Kumar et al.

Correspondence to: Suresh Dhaniyala (sdhaniya@clarkson.edu)

The copyright of individual parts of the supplement might differ from the article licence.

S1 Performance metric equations

The metrics of RMSE (Root Mean Square Error) and MAE (Mean Absolute Error) are frequently utilized in model comparison due to their quantitative evaluation capabilities, error magnitude measurement, and ability to aid in model selection.

The RMSE and MAE are defined as follows;

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - R_i)^2} \quad (1)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |x_i - R_i| \quad (2)$$

where,

RMSE = root mean square error $\mu g/m^3$

MAE = mean absolute error $\mu g/m^3$

10 N = number of 1-hour periods during which all identical instruments are operating and returning valid averages over the duration

x_i = valid 1-hour averaged PA sensor PM2.5 concentration for hour i $\mu g/m^3$

R_i = valid 1-hour averaged EPA PM2.5 concentration for hour i $\mu g/m^3$

15 The locations of sampling sites and the EPA/PA reference numbers for the sites are listed in Table S1. In Table S2, the descriptive statistics related to Figure 3 is shown.

Table S1. Locations of sampling sites of PM_{2.5}

ID	Location	Name	Latitude	Longitude
E1	4500 W. 123rd St.	170310001	41.67099	-87.73246
E2	7801 Lawndale	170310076	41.75140	-87.71349
E3	Kingery Expressway & Torrence Avenue	170310119	41.57862	-87.55741
E4	9511 W. Harrison St	170314007	42.06029	-87.86322
E5	13th St. & 50th Ave.	170316005	41.86443	-87.74890
P1	The Heart Of Riverside	10416	41.83212	-87.81295
P2	Sasa_Pa2_SI_W	8476	41.86120	-87.62989
P3	Sheffield Neighbors	12981	41.91974	-87.65742
P4	Ukrainian Village	3499	41.89270	-87.68569
P5	Wheaton Highlands	36395	41.86352	-88.08800
P6	West Lawn	39173	41.76776	-87.73013
P7	West Rogers Park	4395	41.99600	-87.69200
P8	Mulford Manor	4404	42.02362	-87.76341
P9	1138 Plymouth	5588	41.86889	-87.62944

S2 Baseline removal

To create the average PSD curve, we obtained the PSD curves for each of the 5 EPA sites and the 9 PA sites. To remove the baseline of the PSD curve, we first smoothed the PSD data with a moving average filter and then subtracted the smoothed PSD curve from the original PSD curve. For the moving average smoothing, we used a window size of 100 hours.

Table S2. Descriptive Statistics of PM_{2.5} measurements from EPA and PA sampling locations

ID	E1	E2	E3	E4	E5	P1	P2	P3	P4	P5	P6	P7	P8	P9
Q1	5.3	5.1	6.3	4.5	4.5	4.5	5.8	4.8	4.2	5.5	5.6	3.9	4.4	4.5
Median	7.9	8.3	9.9	8.5	7.7	9.8	11.6	10.7	8.8	11.1	11.1	8.5	9.7	9.6
Mean	8.8	9.1	11.0	9.7	8.8	12.6	14.3	13.5	12.0	14.1	14.6	11.0	12.2	12.4
Q3	11.2	12.2	14.6	3.4	11.8	17.8	20.6	19.3	16.9	20.1	21.1	15.6	17.3	17.5

Table S3. Local correction model performance

Location	Model	R^2	RMSE	MAE	Time
US	St correction	0.84	2.5	1.7	daily
Chicago	Local correction	0.86	1.5	0.3	daily
Chicago	Local correction	0.73	2.6	1.9	hourly

In Table S3, the performance of the standard US-wide correction model in comparison to a local correction model for the prediction of PA data is shown. Two local correction models were built - one based on daily data and the other on hourly data. The standard correction model was built on daily data and was used accordingly in this study.