2024 IEEE International Instrumentation and Measurement Technology Conference (I2MTC),
20-23 May 2024, Glasgow, Scotland, UK
DOI:10.1109/I2MTC60896.2024.10561189

Selection and Aggregation of Low-cost Particle Sensors for Outdoor Particulate Matter Measurement

1st Jie Li
College of General Aviation and Flight
Nanjing University of Aeronautics and Astronautics
Nanjing, China
email lijienuaa@nuaa.edu.cn

2nd Zhengjia Xu
School of Aerospace, Transport, and Manufacturing
Cranfield University
Cranfield MK43 0AL, UK
email Billy.Xu@cranfield.ac.uk

3rd Zaheer Nasar
School of Water, Energy and Env.
Cranfield University
Cranfield MK43 0AL, UK
email z.a.nasar@cranfield.ac.uk

4th Valerio Ferracci
School of Water, Energy and Env.
Cranfield University
Cranfield MK43 0AL, UK
email v.ferracci@cranfield.ac.uk

5th Neil Harris
School of Water, Energy and Env.
Cranfield University
Cranfield MK43 0AL, UK
email neil.harris@cranfield.ac.uk

***Abstract*—A growing number of low-cost sensors (LCS) have been used to monitor air pollution in outdoor air. The benefit of utilizing LCS lies in its ability to offer increased spatial coverage, which provides real-time measurements at a reduced cost. The selection and combination of low-cost sensors represent the primary challenge in conducting observations using such sensors. This paper employs a sensor quality ranking strategy, utilizing random forest (RF) for aggregating the selected LCS combination, followed by evaluating the correction results using various model evaluation metrics. The LCS used in this study, regardless of their quality grades, achieves a coefficient of determination of 0.93 or higher after model calibration, indicating the effectiveness of employing RF for aggregation. It is found that using a pair of top and averaged LCS can significantly enhance the measurement quality by 25% in RMSE. Using RF to calibrate a single LCS increases the measurement performance at least two times in terms of MSE, RMSE, and MAE. Using paired LCS with RF aggregation for measuring PM2.5, the aggregated observation significantly approximates the reference measurement with $R^2 = 0.986$.**

***Index Terms*—low-cost sensor, air quality monitoring, particulate matter measurement, random forest, sensor ranking**

I. Introduction

As one dominant environmental factor in human health [1], approximately 6.7 million deaths associated with myocardial infarction, strokes, heart failure, asthma, chronic obstructive pulmonary disease, and lung cancer are yielded by ambient air pollution and poor air quality [2]. There has been a persistent policy commitment in the UK to tackling air quality monitoring within the UK Clean Air Strategy [3] [4]. Moreover, the strategy of enhancing transparency by sharing local and national monitoring data with an accessible portal facilitates assessing air quality with diverse graded particle sensors.

This work was supported by the China Scholarship Fund (Nos. 202006835001).

The major metric of representing outdoor particulate matter (PM) is particle mass concentration characterized by particle size. The primary particles are combinations of fine particles and ultrafine particles with diameters less than 2.5 μm (PM2.5) that are mostly generated from combustion, while the secondary particles are coarse particles generated by mechanical or chemical reactions with diameters smaller than 10 μm (PM10) [5].

Low-cost sensor (LCS) is a promising optical measurement method, allowing light-scattering large city-scale deployment to obtain temporal-spatial high-quality measurements as well as mitigation of high expenses associated with regulatory-graded sensors [6]. Nevertheless, a dominant challenge of interpreting measurements from LCS is time-dependent diversities and data variations in terms of sensitivity affected by environmental factors like temperature and humidity [7], which drives the demand for understanding sensor performance and error sources beforehand [8]. Specifically, LCS performance without prompt calibrations [9] presents wide-range bias, drift and offset distributions with strong non-linearity in error models. Consequently, the pre-estimation of LCS quality and cross-referencing with reference sensors are significant in interpreting LCS outputs to prevent overall system degradation with adoption of a small number of poor LCS. Furthermore, given the deployed multiple sensors in certain locations, identifying the most suitable combination of sensors to maximize the measurement performance with minimum computational expenses is prominent in practice.

To diminish data variety and enhance precision from LCS outputs, fusion methods of multiple LCS for PM observations are promising to fulfil anticipated performance requirements over traditional instrumentation. For instance, this study [6] found that deploying those affordable LCS presents substantial potential to expand spatial coverage for precise measurements that expect to mitigate economic limitations with the deployment of high costs associated with regulatory-grade monitoring. Particularly, the utilization of the machine learning approach with the aggregation method shows significant potential in discovering buried patterns and leveraging weights among multiple sources.

This study aims to demonstrate a solution with practical field-test samples for maximizing efficiency in using multiple LCS, i.e. to select, identify and utilize LCS combinations that will maximize the observation performance. Practical LCS performance is first cross-compared with reference sensors to perform individual LCS assessments so as to determine possible LCS combinations. A machine learning aided fusion approach is adopted to aggregate LCS measurement. The optimal LCS combination is hereby provided by identifying the biggest significant performance achievements over LCS under the fusion scheme.

II. Data and Methodology

A. Data Sources

Outdoor PM measurements are collected by 6 PurpleAir LCS units located at Cranfield's authority network (AURN) monitoring station $(0.689°W, 52.074°N)$. The LCS is installed on top of the roof of a two-floor building with approx. 10 m a.g.l. altitude. The figure of showing on-site installation scheme is depicted in Figure 1. The PM measurements, i.e. AURN datasets are captured from 10th August to 21st September 2023 with a time resolution averaged of 15 min.

Fig. 1. Spatial distribution of the LCS network and the reference station (main panel) and an overview of the location of the network (upper right). Maps obtained from © Google Earth contributors 2023.

B. Characterization and Grading of LCS Observations

Given diversity in LCS performance, after preprocessing of LCS data by removing outliers, and interpolating blank samples, characteristics of measurements are extracted to assess LCS observation qualities quantitatively. These LCS performance indicators comprise fundamental statistics of averaging, medium, and maximization functions, as well as correlation coefficient and similarity analysis to assess accuracy by cross-comparison with a reference sensor.

TABLE I
Statistic analysis result and grading of LCS with correlation and similarity measurements.

Sensor ID	PM	Avg.	Med.	Max.	r	s	Gr. Score
LCS1	PM1	6.75	3.22	44.69	0.95	0.97	Top 0.914
	PM2.5	10.65	5.27	67.41	0.96	0.93	
	PM10	13.20	6.73	78.76	0.774	0.90	
LCS2	PM1	7.32	3.74	50.63	0.93	0.97	Good 0.902
	PM2.5	11.17	5.76	71.60	0.93	0.93	
	PM10	13.96	7.20	83.89	0.752	0.90	
LCS3	PM1	6.90	3.24	47.58	0.95	0.97	Top 0.914
	PM2.5	10.62	5.04	70.3	0.96	0.93	
	PM10	12.25	5.88	77.77	0.772	0.90	
LCS4	PM1	6.07	2.51	44.36	0.94	0.97	High 0.905
	PM2.5	9.96	4.34	68.36	0.94	0.93	
	PM10	11.36	4.99	76.31	0.752	0.90	
LCS5	PM1	7.85	4.24	51.59	0.94	0.96	Good 0.903
	PM2.5	11.91	6.33	75.46	0.94	0.92	
	PM10	13.61	6.99	82.36	0.756	0.90	
LCS6	PM1	6.11	2.73	42.67	0.94	0.97	High 0.906
	PM2.5	9.89	4.55	65.86	0.94	0.93	
	PM10	11.30	5.18	74.89	0.754	0.90	
Ref	PM1	5.74	3.17	55.96	-	-	-
	PM2.5	7.32	4.67	59.84	-	-	
	PM10	12.10	8.69	77.83	-	-	

- Correlation analysis

The Pearson correlation coefficient r is applied to measure linearity between two measurements by calculating the normalized correlation function in between. The r value ranges in $[-1, 1]$, where r presents a positive proportion to relationships [10].

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \tag{1}$$

where x_i, y_i are two sets of sample observations indexed with i; $\bar{x}$, $\bar{y}$ are the averaged values of x and y.

- Similarity analysis

Another performance indicator taken into account is a similarity metric s aiming to quantitatively evaluate the similarity between two LCS observations, with the equation given below. The s ranges in $(0, 1]$ and presents a positive proportion to the similarity.

$$s = e^{-\frac{d^2}{2\delta^2}} \tag{2}$$

where d represents Euclidean distance between two observations; and δ is a weight defined in the Gaussian kernel function, controlling the decay rate.

- Grading criteria

Since the statistic measurements merely reflect shapes of data distribution affected by noises rather than approximation towards ground truth, the grading function G_i based on s and r calculations is formulated using a weighted accumulative format as factors are proportional to the ultimate performance:

$$G_i = \sum_{j=1}^{M} \gamma_j (\alpha s_i + \beta r_i) \tag{3}$$

where j stands for a measurement from one LCS; M is the number of measurements from one LCS, i.e. 3 in this case corresponding to PM1, PM2.5 and PM10; α, β, and γ are weights to leverage significance between correlation and similarity or sensitivity to different measurements, where $\alpha + \beta = 1$, and $\gamma_1 + \gamma_2 + \gamma_3 = 1$.

When choosing averaged weight values of $\alpha = \beta = 1/2$ and $\gamma_1 = \gamma_2 = \gamma_3 = 1/3$, the grade function for each LCS along with performance indicator results are displayed in Table I. This study selects 3 grades, i.e. 'Top', 'High', and 'Good' to differentiate LCS quality, thus the LCS are graded accordingly based on G_i outcomes.

C. Determination of LCS Combinations

From Table I, it is noticeable that the selected LCS are mostly more sensitive to PM1 and PM2.5 particles, given relatively high values in r and s than PM10. Specifically, the maximum similarity between LCS and the reference reaches up to 97% indicating promising applications of LCS. The sensitivity to PM10 is relatively low with averaged similarity of 90%, suggesting a degraded performance in measuring coarse particles.

When analysing performance differences among LCS, LCS present diverse performance given distinctions in sample distribution and similarity. The calculation of G allows ranking of LCS based on r and s. Accordingly, the top 2 LCS, i.e. LCS3 and LCS1 are classified as top quality, followed by LCS4 and LCS5 graded as high quality, and good quality grades of LCS5 and LCS2.

Consequently, the following LCS combinations are hereby identified manually with multiple combination criteria detailed in Table II. The combination criteria consider the LCS number impact on the performance improvement and different combinations of heterogenous graded LCS mixed with multiple graded LCS solutions.

D. Random Forest for LCS Calibration and Aggregation

As a classical boosting-based ensemble learning technique, Random Forest (RF) is efficient in handling aggregation tasks from multi-source objectives and succeeds in calibrating single LCS or fusing multiple LCS [7]. The principle of the boosting scheme applies an averaging aggregator to integrate multiple decision boundaries from multiple individual regressors that allow accuracy enhancement in predictions from distinguished expert knowledge domains.

E. Performance Metrics

The following typical metrics are selected for measuring the RF regression performance: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error(MAE), Explained Variance Score (EV) and Coefficient of Determination R^2 [10]:

TABLE II
IDENTIFIED LCS COMBINATION SET.

ID	Devices	Combination criteria
S1	LCS1	Top quality to valuate influence of sensor quality on calibration.
S2	LCS3	ditto
S3	LCS4	High quality to valuate influence of sensor quality on calibration.
S4	LCS6	ditto
S5	LCS5	Good quality to valuate influence of sensor quality on calibration.
S6	LCS2	ditto
S7	LCS3+LCS1	Two different quality sensors to valuate the influence of quality sensor quantity on calibration.
S8	LCS3+LCS6	ditto
S9	LCS3+LCS2	ditto
S10	LCS6+LCS2	ditto
S11	LCS4+LCS6	ditto
S12	LCS5+LCS2	ditto
S13	LCS1+LCS3 LCS4+LCS6	Four different quality sensors to the influence of quality and quantity on calibration
S14	LCS1+LCS3 LCS5+LCS2	ditto
S15	LCS4+LCS6 LCS5+LCS2	ditto
S16	LCS1+LCS3 LCS4+LCS6 LCS2+LCS5	Six different quality sensors to the influence of quality and quantity on calibration.

- Mean Squared Error (MSE):

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i \right)^2 \tag{4}$$

- Root Mean Squared Error (RMSE):

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} \|y(i) - \hat{y}(i)\|^2}{N}}, \tag{5}$$

- Mean Absolute Error(MAE):

$$\text{MAE} = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n} \tag{6}$$

- Explained Variance Score (EV):

$$v(y, \hat{y}) = 1 - \frac{\text{Var}\{y - \hat{y}\}}{\text{Var}\{y\}} \tag{7}$$

- Coefficient of determination R^2:

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}} \tag{8}$$

where SS_{res} and SS_{tot} stand for the sum of squared residuals, and total sum of squares, respectively.

III. Experimentation and Result Analysis

A. LCS Performance Comparison

This experiment aims to evaluate performance differences among different graded LCS with or without RF calibrations. The evaluation demonstrates LCS1 (top quality), LCS6 (high quality), and LCS2 (good quality), over PM1 measurements, and the performance metrics results are presented in Figure 2.

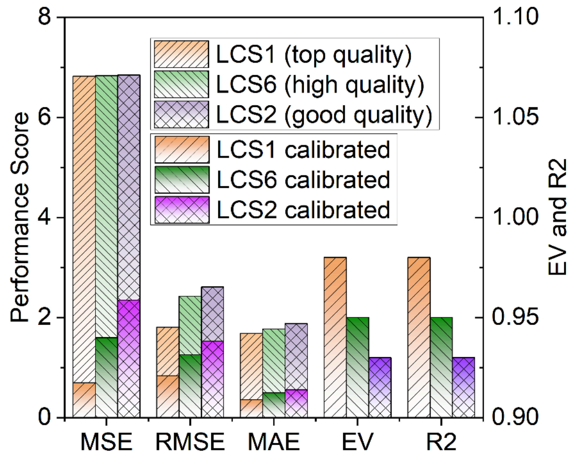

Fig. 2. Performance indices based on different quality sensors, LCS3 (top quality), LCS6 (high quality), and LCS2 (good quality) using the RF calibration model.

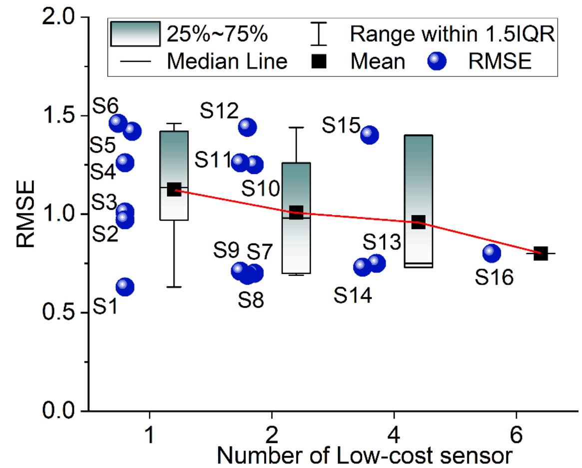

Fig. 3. RMSE of PM2.5 calibration using a single low-cost sensor, two, four, and six low-cost sensors.

It is observable from Figure 2 that the top-quality LCS has the minimum values among all the performance metrics before and after RF calibrations. Similarly, the LCS2 performs the worst with or without calibration in terms of MSE, RMSE, MAE, EV, and R^2. It is concluded that RF calibration preserves equal consequences over different graded LCS, whilst the enhancement degree may be worth further investigation.

With the adoption of RF calibration, the metric values of MSE, RMSE, and MAE are significantly reduced by more than 50% on average, suggesting the effectiveness of RF in compensating LCS observation errors.

The diversity of LCS still remains after RF calibration, especially in EV and R^2 metrics. Specifically, the EV and R^2 of LCS1 are two times than LCS3 after the calibration.

B. Identification of Optimal LCS Combination

This experiment aims to identify the optimal LCS combination from the pre-selected combination set from Table II. Specifically, the most efficient number when using LCS for PM measurement shall be identified, as well as analysis of how different combinations with RF aggregation may enhance the measurement performance most.

• Identifying Combination Number

RMSE is demonstrated as the primary performance indicator for identifying the combination number, and the performance result with RF aggregation over PM2.5 measurements is presented in Figure 3.

It is concluded from Figure 3 that the aggregation scheme with combinations of LCS facilitates measurement performance. Using 6 LCS for aggregation with RF reduces RMSE by approximately 30% than using a single LCS on average. With an increment of LCS number during aggregation, the diversity in combining with different LCS reduces as the length of the RMSE bar shrinks gradually. It is also indicated that choosing the number of 2 in the LCS combination is recommended given the fastest decline rate in the RMSE curve whilst preserving the minimum LCS number.

By analysing RMSE results for single LCS performance, LCS order after RF calibration anticipates the order ranked from G function, which shows a similar result of equality in RF calibration.

When observing the RMSE result aggregated with 2 LCS, it is found that the top combination set of S7, S8 and S9 outperforms the worst combination of S12 by 4 times, while the performance differences between S7, S8 and S9 are negligible. At least one top-graded LCS is included in the top combination set, which verifies the effectiveness of the grading function. It is found that the selection of degraded LCS is more likely to provide degraded combination outcomes. For instance, the ranking order of the combination with two LCS anticipates being $S8 > S7 > S9 > S10 > S11 > S12$, while this order is still satisfied especially when analysing the worst 3 combinations.

When analysing the aggregated outcome with 4 LCS, the combinations of S13 and S14 outperform S15 by 4 times, where S13 and S14 both consist of the top-graded LCS. This phenomenon may suggest the necessity of adopting a top or at least a high-graded LCS during aggregation that will significantly improve the overall performance, whilst the aggregation over low-graded LCS even with more numbers still offers limited performance improvement.

• Detailed Performance Metrics for Two LCS

To identify the optimal combination with all the performance metrics taken into account, the performance metrics results of PM2.5 observations are depicted in Figure 4.

The combination set selects S1 as the up-boundary, and S6 as the low-boundary. Similar results are obtained by analysing the overall performance metrics. For instance, the combination set of S7, S8 and S9 comprising at least one top-graded LCS fosters aggregating the measurement performance significantly, particularly for enhancing MSE and RMSE. The combination set without incorporation of top LCS can only offer a limited improvement compared with the performance using the worst single LCS. The adoption of the two worst

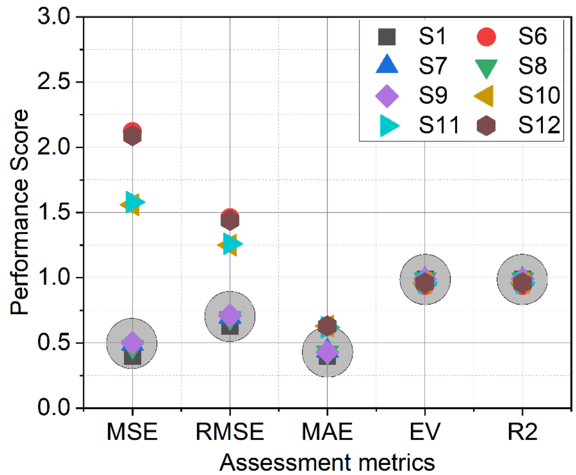

Fig. 4. The evaluation metrics for the networking schemes involving S1, S6, S7, S8, S9, S10, S11, and S12 are depicted. The shaded circle area in the figure represents the region centered around S9.

LCS facilitates the performance to a negligible degree, hereby such a combination is anticipated to be avoided in practice.

Consequently, the comparative analysis reveals that using a pair of sensors with a top and a high-graded LCS tends to yield comparatively stable, precise observational measurements with smaller sensors to achieve cost-effective measurement solutions.

C. Performance Analysis with Selected Pair

This experiment aims to evaluate the relationship with aggregation of the selected combination between references, where the evaluation demonstrates the S9 combination corresponding to the pair of LCS2 and LCS3. The measurement samples over PM2.5 with and without aggregation are cross-referenced with a ground truth station with results displayed in Figure 5.

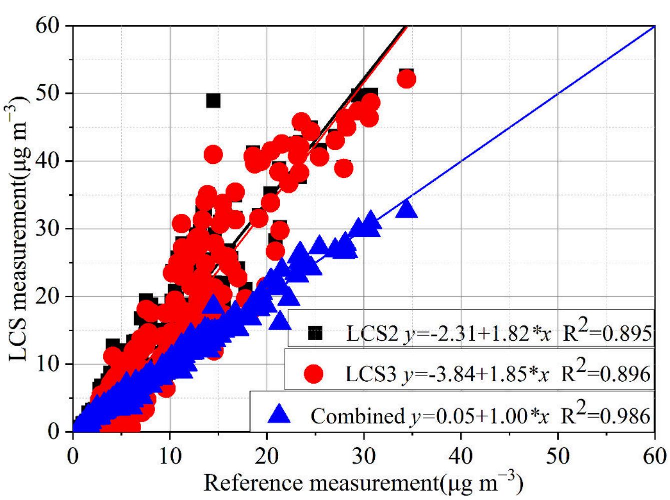

Fig. 5. Comparison of calibrated data using LCS3 and LCS2, as well as their raw data, with reference stations measurement (using PM2.5 as an example).

From Figure 5, the pair of a top and low-graded LCS approximate the true reference evidently. The linearity slope of the linear regression outcomes shows a high value of 1.8 without aggregation, whilst the slope drops down to 1.0 after aggregation, indicating an improved linearity capability. Moreover, the initial drift errors are also reduced to 0.05 after aggregation, suggesting an improved sensitivity against small measurements. The concentration of measurements is improved, suggesting a reduced noise bias and outline numbers. Accordingly, this result consolidates the solution of incorporating LCS pairs for aggregating measurement performance.

Conclusion

Given the fact of utilising more LCS for monitoring outdoor air pollution, this paper aims to analyse the performance distinctions among LCS and identify the optimal LCS combination to enhance measurement accuracy. A grading function based machine learning aggregation method is attempted to fuse multiple LCS. The grading function grades LCS qualities, and different combinations of LCS are aggregated using RF. Through analysing practical LCS measurements of PM1, PM2.5 and PM10, the LCS performance is ranked and graded by the grade function scheme. It is found that using a pair of top and averaged LCS can significantly enhance the measurement quality by 25% in RMSE. Using RF to calibrate a single LCS increases the measurement performance at least two times in terms of MSE, RMSE, and MAE. Using paired LCS with RF aggregation for measuring PM2.5, the aggregated observation significantly approximates the reference measurement with $R^2 = 0.986$.

References

[1] Public Health England, Health matters: air pollution, https://www.gov.uk/government/publications/health-matters-air-pollution/health-matters-air-pollution (last access: 20 November 2023), 2018.

[2] Health Effects Institute, State of Global Air 2020, https://www.stateofglobalair.org/ (last access: 20 November 2023), 2023.

[3] Defra, UK Air Information Resource – Defra, UK [data set], https://uk-air.defra.gov.uk/data (last access: 24 May 2023), 2023.

[4] Defra and DfT, UK plan for tackling roadside nitrogen dioxide concentrations: An overview, https://www.gov.uk/government/publications/air-quality-plan-for-nitrogen-dioxide-no2-in-uk-2017 (last access: 20 November 2023), 2017.

[5] Simões Amaral and João Andrade De Carvalho and Maria Angélica Martins Costa and Cleverson Pinheiro, An overview of particulate matter measurement instruments, Atmosphere, 6(9), pp. 1327-1345, 2015.

[6] Nuria Castell and Franck R. Dauge and Philipp Schneider and Matthias Vogt and Uri Lerner and Barak Fishbain and David Broday and Alena Bartonova, Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates?, Environ. Int., 99, pp. 293–302, 2017.

[7] Tony Bush and Nick Papaioannou and Felix Leach and Francis D. Pope and Ajit Singh and G. Neil Thomas. et al. , Machine learning techniques to improve the field performance of low-cost air quality sensors, Atmos. Meas. Tech., 15, pp. 3261–3278, 2021.

[8] Michael R. Giordano and Carl Malings and Spyros N. Pandis and Albert A. Presto and V. F. McNeill and Daniel M. Westervelt, et al., From low-cost sensors to high-quality data: A summary of challenges and best practices for effectively calibrating low-cost particulate matter mass sensors, J. Aerosol Sci., 158, 105833, 2021.

[9] Naomi Zimmerman and Albert A. Presto and Sriniwasa P. N. Kumar and Jason Gu and Aliaksei Hauryliuk and Ellis S. Robinson et al. , A machine learning calibration model using random forests to improve sensor performance for lower-cost air quality monitoring, Atmos. Meas. Tech.,11, pp. 291–313, 2018.
[10] Fabian Pedregosa and Gael Varoquaux and Alexandre Gramfort and Vincent Michel and Bertrand Thirion and Olivier Grisel, Scikit-learn: Machine Learning in Python, J. Mach. Learn Res., 12, pp. 2825-2830, 2011.

Selection and aggregation of low-cost particle sensors for outdoor particulate matter measurement

Li, Jie

2024-06-28

Li J, Nasar Z, Ferracci V, et al., (2024) Selection and aggregation of low-cost particle sensors for outdoor particulate matter measurement. In: 2024 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), 20-23 May 2024, Glasgow, Scotland, UK

https://doi.org/10.1109/I2MTC60896.2024.10561189