Science of the Total Environment 818 (2022) 151769

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Review

Performance evaluation of low-cost air quality sensors: A review

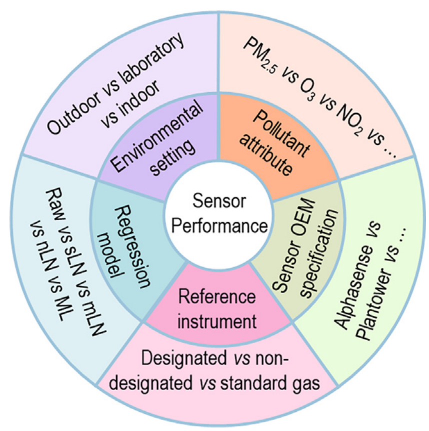
Ye Kang [a], Lu Aye [b], Tuan Duc Ngo [b], Jin Zhou [a,*]

[a] Department of Civil Engineering, Monash University, Clayton, Victoria 3800, Australia
[b] Department of Infrastructure Engineering, Faculty of Engineering and Information Technology, The University of Melbourne, Parkville, Victoria 3010, Australia

HIGHLIGHTS

- This work reviewed 112 studies on the performance of low-cost air quality sensors.
- r^2, *RMSE*, *MNB* and *CV* were applied to evaluate sensor performance.
- Environmental settings, reference instruments, and regression models affect sensors.
- The impact of pollutant attributes and sensor OEM specifications is inconclusive.
- Limited studies followed US EPA guideline to examine low-cost air quality sensors.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 8 September 2021
Received in revised form 27 October 2021
Accepted 14 November 2021
Available online 19 November 2021

Editor: Hai Guo

Keywords:
Statistical measure
Environmental setting
Pollutant attribute
Sensor original equipment manufacturer (OEM) specification
Reference instrument
Regression model

ABSTRACT

The monitoring of air quality compliance requires the use of Federal Reference Methods (FRM)/Federal Equivalent Methods (FEM); nevertheless, the validity and reliability of low-cost sensors deserve attention due to their affordability and accessibility. This review examines the methodologies of previous studies to characterise the performance of low-cost air quality sensors and to identify the influential factors in sensor evaluation experiments. The data on four statistical measures (Correlation of Determination, r^2; Root Mean Square Error, *RMSE*; Mean Normalised Bias, *MNB*; and Coefficient of Variation, *CV*) and details about five methodological factors in experimental design (environmental setting, reference instrument, regression model, pollutant attribute, and sensor original equipment manufacturer (OEM) specification) were extracted from a total of 112 primary articles for a detailed analysis. The results of the analysis suggested that low-cost air quality sensors exhibited improved r^2 and *RMSE* in the experiments with stable environmental settings, in the comparison against non-designated reference instruments, or in the analysis where advanced regression models were used to adjust the sensor readings. However, the pollutant attribute and sensor OEM specification had inconclusive effects on r^2 and *RMSE* due to contradictory results and lack of sufficient data. *MNB* and *CV*, two measures that US EPA recommends to determine the suitable application tier of air quality sensors, varied significantly among published experiments due to the discrepancy in experimental design. The outcomes of this work could provide direction to researchers regarding sensor evaluation experiments and guide practitioners to effectively select and deploy low-cost sensors for air quality monitoring.

* Corresponding author.
E-mail address: Jenny.zhou@monash.edu (J. Zhou).

https://doi.org/10.1016/j.scitotenv.2021.151769

Contents

1. Introduction

The epidemiologic link between exposure to air pollutants and adverse health outcomes has been identified in previous studies (Antonini et al., 2004; Makri and Stilianakis, 2008). Traditionally, air pollutants are monitored using complex and expensive stationary equipment at fixed locations (Kularatna and Sudantha, 2008). This paradigm is changing with the availability of low-cost, easy-to-use air pollutant sensors that can provide real-time measurements at an affordable price, allowing higher spatial coverage than current expensive instruments (Karagulian et al., 2019). These attributes offer new opportunities to enhance existing air monitoring capabilities and possibly present new air pollutant monitoring applications (Snyder et al., 2013).

A number of review articles have already addressed the emerging area of low-cost sensors in air quality monitoring. Many reviews either focused on sensor technologies for monitoring particulate and gaseous air pollutants (McKercher et al., 2017; Morawska et al., 2018) or provided a general overview of the potential application areas for low-cost sensors and their networks (Borghi et al., 2017; Clements et al., 2017; Kumar et al., 2015; Kumar et al., 2016). Although these studies all raised concerns regarding the data quality of low-cost sensors, they did not offer a universal solution to this problem. Three reviews focused on quantitative data that measured the validity and reliability of low-cost air quality sensors. Jovasevic-Stojanovic et al. (2015) appraised approximately 20 published articles that evaluated the performance of low-cost sensors for particulate matter (PM) detection. This review article mainly discussed how environmental settings and pollutant attributes affected the Coefficient of Determination (r^2) between measurements of low-cost sensors and that of reference instruments. Rai et al. (2017) reviewed nearly 40 existing studies, and in addition to PM sensors, their performance assessment included sensors for measuring gaseous pollutants, such as carbon monoxide (CO), ozone (O_3), and nitrogen dioxide (NO_2). In their study, not only r^2 but also Root Mean Square Error (*RMSE*) and Coefficient of Variation (*CV*) were applied to evaluate the sensor performance of different environmental settings and sensor original equipment manufacturer (OEM) specifications. Karagulian et al. (2019) reviewed a total of 64 independent articles and utilised r^2 to evaluate the correlation of low-cost sensors with the reference for PM, CO, O_3, and NO_2 measurements. In addition to environmental settings, they analysed the influence of the regression model on sensor performance.

The review articles described in the previous paragraph presented three main knowledge gaps that motivated the current work. The first and the most important knowledge gap was that existing reviews of low-cost sensor studies dominantly focused on r^2 between low-cost sensors and reference instruments. Limited insights were presented on other statistical measures, such as *RMSE* and *CV*. In addition, Mean Normalised Bias (*MNB*), a statistical measure recommended by the United States Environmental Protection Agency (US EPA) guideline (Williams et al., 2014), was not covered in existing review studies. Secondly, a few experimental studies tested the impact of relative humidity (Malings et al., 2019b; Hong et al., 2021), pollutant concentration (K.K. Johnson et al., 2018; Hong et al., 2021; Sayahi et al., 2019a), sensor type (Kelly et al., 2017; Hong et al., 2021; Spinelle et al., 2015a,b), reference instrument (Kelly et al., 2017), and regression model (K.K. Johnson et al., 2018; Kelly et al., 2017; Malings et al., 2019b; Hong et al., 2021) on sensor performance, but their results have not yet been quantitatively synthesised and reported in previous review articles. Reviews that comprehensively diagnose factors affecting the data quality of low-cost sensors are currently lacking. Finally, published review articles only included low-cost sensors for monitoring PM, O_3, NO_2, and CO, neglecting many other gaseous pollutants (e.g. NO and SO_2).

To address these research gaps, this review aims: i) to examine how previous studies were conducted to characterise the performance of low-cost air quality sensors and ii) to identify the influential factors in sensor evaluation experiments. The outcome of this research could benefit future low-cost sensor evaluation experiments and provide useful information for individuals interested in the use of low-cost sensors.

2. Methods

2.1. Definition of low-cost sensors

There is no agreed definition of a "sensor". One should distinguish between the sensor module and the whole monitoring system, which typically includes one or multiple sensor modules together with a protective box, power system, electronic hardware, and components for data transmission, storage, and retrieval. In this article, the authors applied the term "sensor" to the sensor module either directly provided by the original equipment manufacturer (OEM) or from the whole monitoring system.

The term "low-cost" refers to the purchase price of sensors, compared with the purchase price and operating cost of the reference instrument. The term is relative, depending on the user and the specific purposes. The low-cost air quality sensor module or monitors in many published works (e.g. Hall et al., 2014; Rai et al., 2017) cost from several hundred to a few thousand dollars. In this work, the low-cost sensor was defined as a sensor module with a price of less than US $1000.

2.2. Literature search and selection criteria

The literature search was carried out using abstract and citation databases, including Scopus, Web of Science, and IEEE Xplore Digital Library for peer-reviewed articles published in English until July 2021, and comprised the fields "article title"; "abstract"; and "keywords". The authors used the following search terms for query: ("low cost" or "inexpensive" or "affordable") and ("sens*" or "monitor*" or "device") and ("evaluat*" or "assess*" or "calibrat*") and ("air pollut*" or "gaseous pollut*" or "particulate matter" or "PM"). Initial database searches generated 358 publications, and 21 additional studies were sought from reference lists of retrieved studies and review articles (Fig. S1). Following the removal of duplicates, articles using secondary data sources, studies without full articles, articles not written in English, and irrelevant citations based on fast screening, 157 publications were further assessed for selection. Given the study scope and data acquisition requirements, original studies were eligible for inclusion in this review when the work 1) focused on the air quality sensors that were categorised as low-cost (≤US $ 1000); 2) specified the manufacturer of the sensor being tested; 3) specified the reference instrument used in the experiment; and 4) applied one or more statistical measures to evaluate the sensor performance. A total of 112 studies met all the four selection criteria and were included in the final data synthesis of this review. Relevant documents were collected and organised in Endnote X9.

2.3. Data extraction and classification

Data on two domains, including i) metrics and measures that were used to describe the sensor performance and ii) methodological factors in experimental design that might influence the sensor evaluation results, were extracted from all identified references. The data are summarised in file *source_data.xlsx* in supporting information.

2.3.1. Performance metrics and statistical measures

Considering the various end-use requirements for different applications, the research community adopts a variety of performance metrics and statistical measures based largely on convenience with respect to their respective demands, and consequently, the sensor performance evaluation suffers from ambiguous and inconsistent interpretations. This review focused on statistical measures across three key performance metrics: (1) correlation, (2) measurement error, and (3) precision. Correlation was a measure of the extent to which the estimated sensor readings and reference values were related. Measurement error compared the difference between the sensor estimates and the corresponding reference measurements. Precision measured the agreement among repeated measurements of the same property under identical or substantially similar conditions.

2.3.2. Methodological factors in experimental design

The features on the following five facets of experimental design were extracted from published studies.

(1) *Environmental setting.* The authors categorised the testing environments into i) laboratory, ii) indoor, and iii) outdoor settings due to the profound variations in meteorological parameters and background concentrations. Laboratory circumstances referred to those investigations that were carried out in controlled chambers, while studies conducted in indoor environments, such as, residential and office spaces, were regarded as indoor conditions. In terms of outdoor tests, sensors and instruments were deployed at open-air sites to detect pollutant concentrations in outdoor ambient air.

(2) *Pollutant attribute.* Air pollutants consist of a composite mixture of particulate matter (PM) and various gaseous pollutants. Besides the pollutant types and concentrations, the authors also documented the size and origin properties of particles. Particles were divided into PM_1, $PM_{2.5,}$ and PM_{10} to align with the conventional definition of size groups. Particle emission sources found in published laboratory experiments were classified into four categories: i) Arizona Road Dust (ARD, standard test dust composed of different minerals); ii) particles generated from combustion (e.g. incense, cigarette, and wood smoke); iii) organic particles (e.g. dust mite, oleic acid, and polystyrene latex spheres); and iv) salt (e.g. sodium chloride and ammonium sulphate) and oxide particles (e.g. aluminium oxide). Regarding the gaseous pollutants, the main species in both outdoor and indoor air pollution were recorded.

(3) *Sensor OEM specification.* The sensors usually perform differently due to the variations in product characteristics and quality assurance. Therefore, the manufacturer, measurement technique, and if available, the upper detection limit of sensors were identified from published works. For the commercial monitor with built-in sensors, OEM of the built-in sensor, rather than the monitor brand, was documented for further discussion. For instance, both AirAssure and Foobot air quality monitors utilised the same sensor modules manufactured by Sharp; and consequently, Sharp was documented.

(4) *Reference instrument.* The evaluation of sensor performance requires comparison against a reference instrument. Reference instruments adopted in published articles were classified into three groups: i) instrument designated by national accreditation bodies (US Environmental Protection Agency, 2017); ii) non-designated equipment; and iii) gas cylinders with known analyte concentrations (standard gas) for the evaluation of sensors for gaseous pollutants. Please note that keeping the non-designated group in this review is not to promote its use as the reference but due to the fact that it has been widely applied in published studies.

(5) *Regression model.* The robustness of regression models is dependent on the complexity of the algorithms. In this article, regression models were categorised into four groups: i) simple linear (sLN), ii) multiple linear (mLN), iii) non-linear (nLN), iv) machine learning (ML), and v) raw. sLN and mLN models described the output as an affine function of single input and multiple inputs, respectively. nLN models involved more complex polynomial equations, such as the quadratic equation, of the input variables, while ML models (e.g. Random Forest and Neural Network) had the ability to capture sophisticated non-parametric relationships between various inputs and target outputs. It should be noted that some studies did not apply any regression model and used raw outputs to reflect the "out of the box" performance of low-cost sensors. This scenario was denoted as "raw" in this work.

2.4. Data analysis

Fig. 1 displays a schematic representation of data classification and statistical analysis approaches. Delineating all combinations of the levels in every methodological factor for a full factorial comparison was not feasible because of insufficient data records. Instead, this

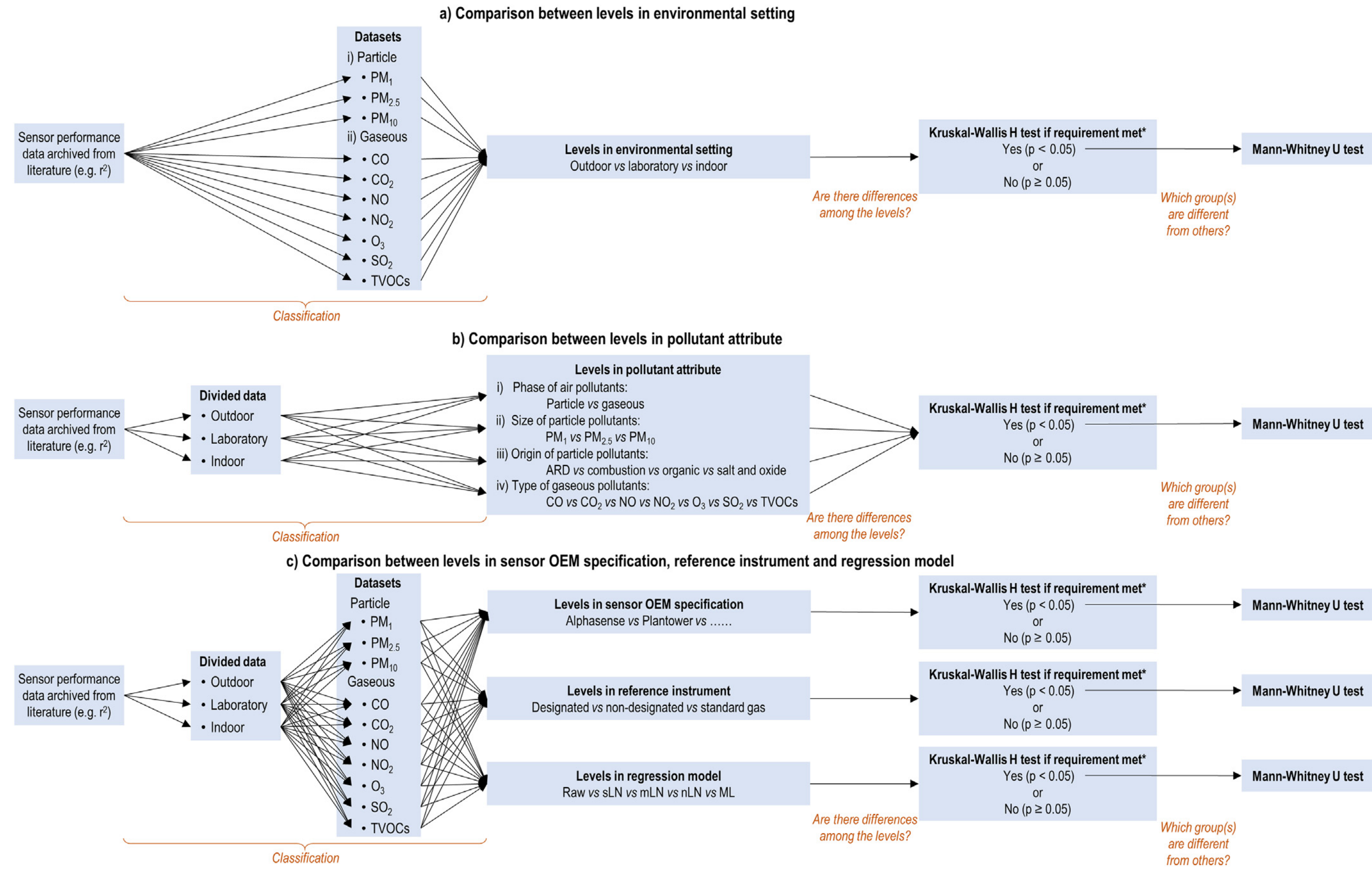


Fig. 1. Schematic representation of data classification and statistical analysis approaches for the comparison at different levels: a) environmental setting, b) pollutant attribute, and c) sensor original equipment manufacturer (OEM) specification, reference instrument, and regression model. *The comparison was applied to the levels with a sample size >7 to ensure the adequate power of the Kruskal-Wallis H and Mann-Whitney *U* tests.

study employed a partial and mixed method to classify the sensor performance data. Three different rules in data classification were selected because they could maximise the number of datasets that met the sample size requirement for the statistical analysis method.

(1) The sensor performance data were classified into 10 datasets according to the types of measured air pollutants (particles: PM_1, $PM_{2.5}$, and PM_{10}; gaseous pollutants: CO, CO_2, NO, NO_2, O_3, SO_2, and TVOCs). The comparison was performed in every dataset to distinguish if any variations existed between the three levels of environmental settings.
(2) To reveal the influence of pollutant attributes, comparisons were made between two broad types of air pollutants, among particles of three different sizes, among particles of four different origins, and among seven species of gaseous pollutants. The data were disaggregated by environmental settings.
(3) The comparisons among multiple OEMs, among three levels in the reference instrument, and among five levels in the regression model were also performed in the 10 datasets mentioned in Section 2.4 (1), but in combination with all environmental settings.

Kruskal-Wallis H and Mann-Whitney U tests were adopted for the assessment of the difference between levels in each dataset. If a significant difference was found in the H test ($p < 0.05$), then the *U* test was used as a post hoc analysis to determine where the difference lay. To ensure the adequate power of the H test and U test, the comparison was applied to the levels with a sample size >7 (Gore and Altman, 1982).

The matched-pair comparison, as in Pearce (2016), was used as a supplement approach if the sample size could not meet the requirement. To ensure a fair comparison, the matched-pair referred to the datasets extracted from the same study with variation in only one methodological factor.

MNB and *CV* are the two statistical measures recommended by the US EPA air sensor guideline (Williams et al., 2014). Therefore, the comparison was made against the performance goals for five application areas (Tier I–Tier V) defined in the guideline. The *Tier V*: *Regulatory Monitoring* required the highest quality data ($-0.07 < MNB < 0.07$ and $CV < 0.07$ for O_3, $-0.10 < MNB < 0.10$ and $CV < 0.10$ for CO, SO_2, $PM_{2.5}$ and PM_{10}, and $-0.15 < MNB < 0.15$ and $CV < 0.15$ for NO_2), followed by *Tier III*: *Supplemental Monitoring* ($-0.2 < MNB < 0.2$ and $CV < 0.2$ for O_3, PM, CO, NO_2, SO_2 and TVOCs). Both *Tier II: Hotspot Identification and Characterisation* and *Tier IV: Personal Exposure* shared the same performance goal ($-0.3 < MNB < 0.3$ and $CV < 0.3$ for all pollutants). *Tier I: Education and Information* allowed a relatively large uncertainty ($-0.5 < MNB < 0.5$ and $CV < 0.5$ for all pollutants).

3. Characteristics of the reviewed literature

The data synthesised in file *source_data.xlsx* was visualised as a Sankey diagram (Fig. 2). The steps in the Sankey diagram correspond to the sensor performance measures and methodological factors described in Section 2.3. The total number in each step may exceed the total number of identified studies because the categories in each step are not mutually exclusive.

Most of the experiments were conducted in outdoor settings ($n = 88$), followed by laboratory ($n = 28$) and indoor ($n = 13$) circumstances. Previous work evaluated sensors for various particle sizes and gaseous pollutant monitoring. Sensors for the measurement of many criteria pollutants, such as $PM_{2.5}$ ($n = 68$), PM_{10} ($n = 20$), O_3 ($n = 38$), NO_2 ($n = 33$), CO ($n = 23$), and SO_2 ($n = 5$), were also among

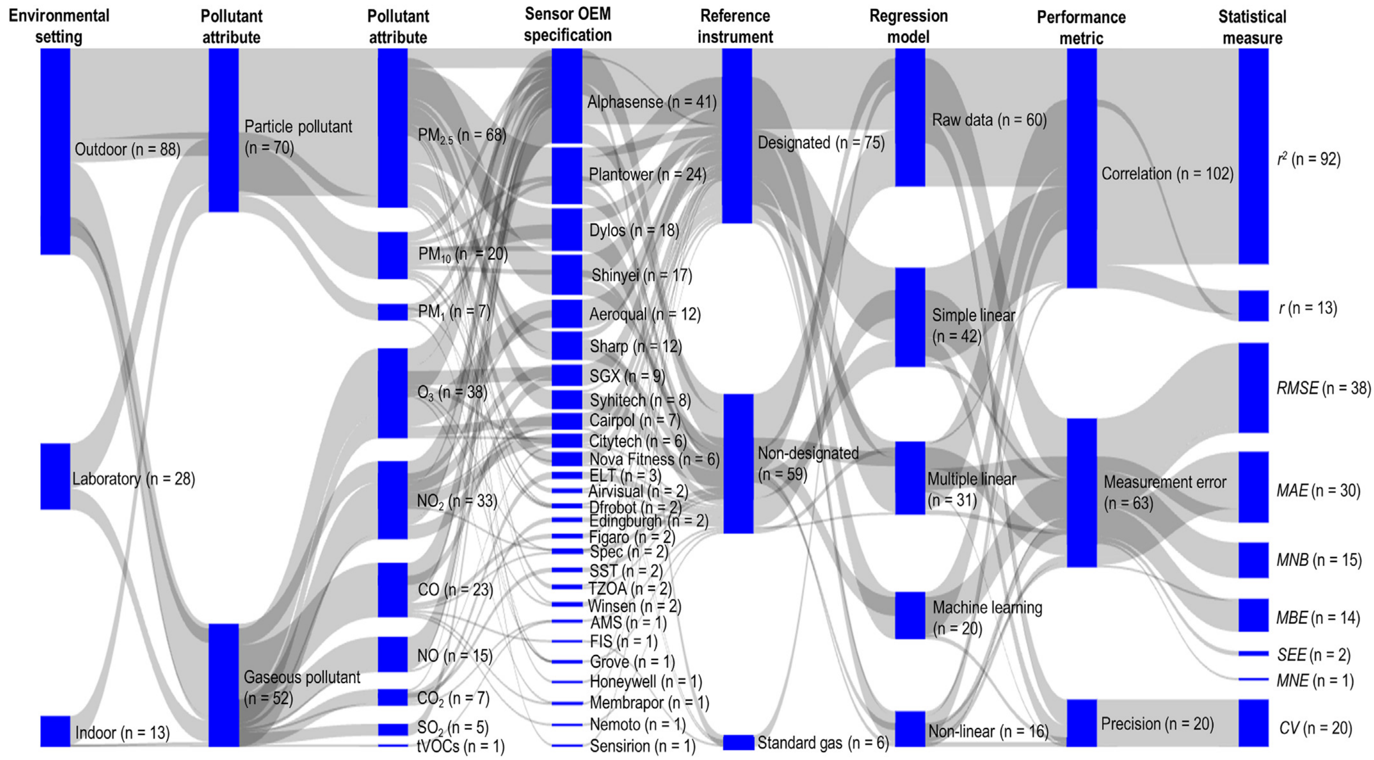


Fig. 2. Sankey diagram showing the main characteristics of 112 reviewed articles. The total number in each step may exceed the total number of identified studies because the categories in each step are not mutually exclusive.

the low-cost sensors tested in previous studies. Out of 27 OEMs, Alphasense sensors (n = 41) were used most often in previous studies, as they can monitor both PM and gaseous pollutants. PM sensors provided by Plantower (n = 24), Dylos (n = 18), Shinyei (n = 17), Sharp (n = 12), Syhitech (n = 8), and Nova Fitness (n = 6), and the sensors for gas produced by Aeroqual (n = 12), SGX (n = 9), Cairpol (n = 7), and Citytech (n = 6) were also found in published studies. It is worth noting that the measurement technique of sensors is dependent on the target pollutant. As indicated in file *source_data.xlsx*, the light scattering technique is applied for particle monitoring. In addition, the electrochemical (EC) and metal-oxide-semiconductor (MOS) methods are the measurement techniques for gaseous pollutants, except for sensors used in CO_2 and TVOCs detection, which relies on the non-dispersive infrared (NDIR) and photoionisation detector (PID) technique, respectively.

The majority of investigations compared the sensor readings with the values from designated instruments (n = 75) or non-designated instruments (n = 59). Standard gas was also applied in six published articles. More than half of the research articles (n = 60) used raw outputs to assess sensor performance. Various regression models, including simple linear (n = 42), multiple linear (n = 31), machine learning (n = 20), and non-linear models (n = 16), were also adopted in previous studies.

Pearson correlation coefficient (r) (n = 13) or its square value (r^2) (n = 92) were commonly adopted in previous studies to describe how well the responses of low-cost sensors correlated with those of reference instruments. These studies applied five different statistical measures to depict the measurement error of low-cost sensors. They were *RMSE* (n = 38), Mean Absolute Error (*MAE*) (n = 30), *MNB* (n = 15), Mean Bias Error (*MBE*) (n = 14), Standard Error of Estimate (*SEE*) (n = 2), and Mean Normalised Error (*MNE*) (n = 1). *RMSE*, *MAE*, and *SEE* measured the magnitude of the measurement error, but the error weighting varied. *RMSE* and *SEE* provided more weighting to the most significant errors, while *MAE* treated each error equally (Pal, 2017). *MNE*, the normalisation of MAE, was expressed as a percentage to facilitate the comparison between datasets with different scales. Both *MBE* and *MNB* described the overall direction of the error. *MBE* measured the deviation (either positive or negative) from the reference instrument, but the *MNB* reported the deviation as a percentage of the reference instrument. A total of 20 studies utilised *CV* to assess the precision of low-cost sensors, a statistical measure of the intra-model variability of low-cost sensors. The equations of all statistical measures are summarised in the supporting information (Table S1).

4. Influences of methodological factors on sensor performance measurement

This section is structured according to the basis of the characteristics of previous studies. Section 4.1 synthesises r^2 values (including r that has been converted to r^2) to evaluate the variation of sensor correlation with the reference instrument. Section 4.2 analyses the *RMSE* data, as it is the most common statistical measure used in the literature for measurement error (Fig. 2). The US EPA air sensor guideline (Williams et al., 2014) suggests *MNB* (another measure for measurement error) and *CV* (a precision measure) as performance indicators. Therefore, Section 4.3 focuses on these two measures and discusses the compliance of low-cost sensors with the performance goals suggested by the US EPA guideline.

4.1. Correlation coefficient (r) and its square value (r^2)

The r^2 (including r that were converted to r^2) measured the degree to which low-cost sensors and reference instruments were in agreement with each other. In this work, r^2 values <0.5 are considered as a weak correlation, 0.5–0.7 as an acceptable agreement, and >0.7 as a strong agreement (Moore et al., 2013).

4.1.1. Sensor readings show better correlations with the reference in well-controlled environmental settings

Both $PM_{2.5}$ and PM_{10} sensors exhibited better correlations with the reference instruments in laboratory experiments than those obtained from non-laboratory tests (Fig. 3a). Similar trends were found in the sensors for the measurement of gaseous pollutants (Fig. 3b), although the significance was only observed in NO_2 sensors. In addition, laboratory data spanned over a much narrower interquartile range (IQR) than those of indoor and outdoor experiments. The discrepancy in medians ($\widetilde{r^2}$) and IQR could be attributed to environmental interference and cross-sensitivity artefacts. Compared with the laboratory environment in which the temperature and relative humidity (RH) are well controlled and a single source of air pollutants is dosed for sensor evaluation, the indoor and outdoor settings provide a more dynamic and diverse environment, which becomes a confounding factor in sensor performance evaluation. The sensor response can change with water vapour as it could modify the optical measures for the light scattering of aerosol particles (Wang et al., 2015) or alter the humidity equilibrium between sampled air and gas sensing electrodes (Sun et al., 2016). There is also evidence to suggest the dependence of sensor response on temperature; the temperature can strongly affect the electrical conductivity of oxide materials (Afshar-Mohajer et al., 2018) and the current of the electrodes (Cross et al., 2017). Additionally, cross-sensitivity is a prevailing challenge for sensors that measures gaseous pollutants. A sensor reacts not only to the target gas but also to other accompanying gases. Multiple studies demonstrated biased measurements because of the interference between O_3 and NO_2 (Cross et al., 2017; Duvall et al., 2016; Zimmerman et al., 2018) as well as between CO and molecular hydrogen (Mead et al., 2013).

4.1.2. No huge difference is induced by pollutant attributes, except cross-sensitivity issue for NO_2 sensors

The $\widetilde{r^2}$ of sensors for gaseous pollutants was slightly higher than that of PM sensors in both outdoor (Fig. 4a(i)) and laboratory settings (Fig. 4b(i)), but statistical significance was only observed in laboratory experiments. The size (Fig. 4a(ii) and b(ii)) and origin (Fig. 4b(iii)) of particles had no distinguishable effect on the correlation between PM sensors and reference instruments. Regarding outdoor gaseous pollutant monitoring (Fig. 4a(iv)), the sensors for O_3, CO, and NO detection ($\widetilde{r^2} = 0.75–0.78$) performed better than NO_2 sensor ($\widetilde{r^2} = 0.57$) because NO_2 sensors (both MOS sensors and EC sensors) are prone to cross-sensitivities with O_3 (Zimmerman et al., 2018). The comparison of gaseous sensors in laboratory experiments was carried out using matched-pair analysis, and the results are presented in Table S2. As expected, this difference was not observed in the laboratory experiments, which usually have a single type of air pollutant dosed for sensor evaluation.

4.1.3. Correlations differ among manufacturers due to the discrepancy in the sensing technique and the upper detection limit

Shinyei sensors ($\widetilde{r^2} < 0.5$) did not perform as well as other sensors ($\widetilde{r^2} > 0.7$) in the measurement of outdoor $PM_{2.5}$ (Fig. 5a(i)). The differences in the signal processing and the airflow mechanism lead to its decreased accuracy.

	$\widetilde{r^2}$	(Q1, Q3)	IQR	n
a) Particle pollutants				
PM_1				
Outdoor	0.90	(0.82, 0.91)	0.09	5
$PM_{2.5}$				
Laboratory	0.97[a]	(0.92, 0.98)	0.06	45
Indoor	0.86[b]	(0.71, 0.90)	0.19	11
Outdoor	0.72[b]	(0.53, 0.85)	0.32	80
PM_{10}				
Laboratory	0.95[a]	(0.78, 0.98)	0.20	11
Outdoor	0.71[b]	(0.48, 0.84)	0.36	20
b) Gaseous pollutants				
CO				
Laboratory	0.99	(0.99, 1.00)	0.01	5
Outdoor	0.78	(0.64, 0.87)	0.23	32
NO				
Outdoor	0.75	(0.27, 0.82)	0.55	23
NO_2				
Laboratory	1.00[a]	(0.99, 1.00)	0.01	8
Outdoor	0.57[b]	(0.24, 0.84)	0.60	59
O_3				
Laboratory	1.00	(0.99, 1.00)	0.01	7
Outdoor	0.78	(0.51, 0.85)	0.34	49
SO_2				
Outdoor	0.63	(0.18, 0.95)	0.77	6

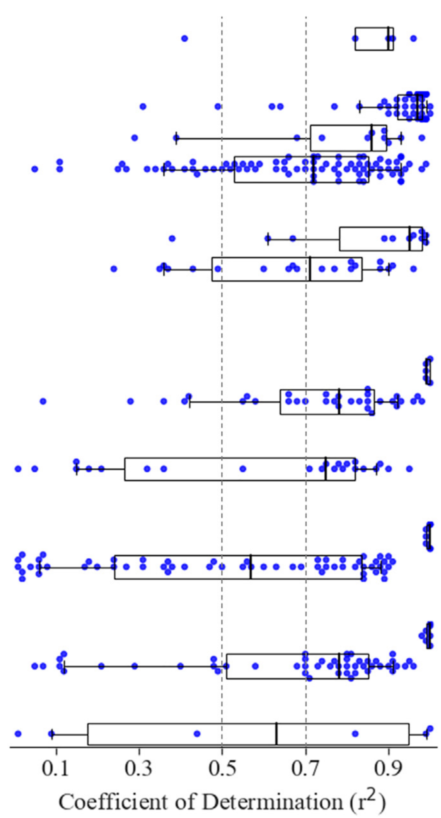


Fig. 3. Influence of environmental settings on the correlation of sensors with the reference instrument. The data were classified by the phase of air contaminants, occurring as a) particle or b) gaseous phase. Each group was further disaggregated by pollutant types. In the same classification, medians ($\widetilde{r^2}$) with different superscript letters are statistically different from each other ($p < 0.05$), while those that share the same superscript letter are not. The statistical test was not performed for the dataset with the sample size ≤7, and the corresponding median has no superscript. In the box plot, whiskers correspond to the 10th and 90th percentiles; upper and lower bounds of the box represent 25th and 75th percentiles; solid lines within the box are the medians. The statistics for levels with the sample size $n < 5$ are not shown in this figure.

	$\widetilde{r^2}$	(Q1, Q3)	IQR	n
a) Outdoor				
i) Phase of air pollutants				
Particle	0.73[a]	(0.52, 0.86)	0.34	105
Gaseous	0.74[a]	(0.37, 0.85)	0.48	169
ii) Size of particle pollutants				
$PM_{2.5}$	0.72[a]	(0.53, 0.85)	0.32	80
PM_{10}	0.71[a]	(0.48, 0.84)	0.36	20
PM_1	0.90	(0.82, 0.91)	0.09	5
iv) Type of gaseous pollutants				
O_3	0.78[a]	(0.51, 0.85)	0.34	49
CO	0.78[a]	(0.64, 0.87)	0.23	32
NO	0.75[a]	(0.27, 0.82)	0.55	23
NO_2	0.57[b]	(0.24, 0.84)	0.60	59
SO_2	0.63	(0.18, 0.95)	0.77	6
b) Laboratory				
i) Phase of air pollutants				
Particle	0.96[a]	(0.91, 0.98)	0.07	56
Gaseous	1.00[b]	(0.99, 1.00)	0.01	24
ii) Size of particle pollutants				
$PM_{2.5}$	0.97[a]	(0.92, 0.98)	0.06	45
PM_{10}	0.95[a]	(0.78, 0.98)	0.20	11
iii) Origin of particle pollutants				
Salt and oxide	0.98[a]	(0.71, 0.99)	0.28	14
ARD	0.96[a]	(0.92, 0.98)	0.06	13
Combustion	0.95[a]	(0.91, 0.98)	0.07	23
Organic	0.98	(0.95, 0.99)	0.04	6
iv) Type of gaseous pollutants*				
O_3	1.00	(0.99, 1.00)	0.01	7
CO	0.99	(0.99, 1.00)	0.01	5
NO_2	1.00	(0.99, 1.00)	0.01	8
c) Indoor				
i) Phase of air pollutants				
Particle	0.86	(0.72, 0.89)	0.17	12
ii) Size of particle pollutants				
$PM_{2.5}$	0.86	(0.71, 0.90)	0.19	11

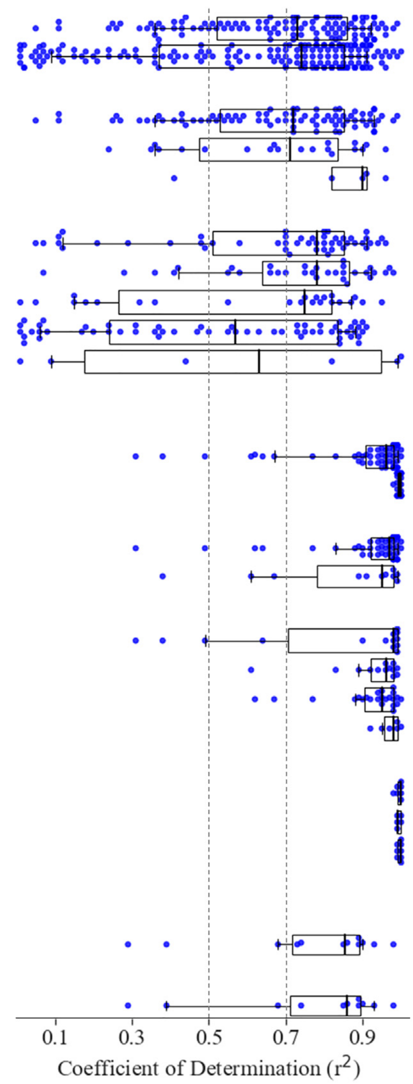


Fig. 4. Influence of pollutant attribute on the correlation of sensor with the reference instrument in a) outdoor, b) laboratory, and c) indoor environmental settings. In each setting, the data were classified by i) phases of air pollutants, ii) sizes and iii) sources of particles, and iv) types of gaseous contaminants. Please refer to Fig. 3 for detailed notations. The statistics for levels with the sample size $n < 5$ are not shown in this figure. *Comparison was made by matched-pair approach because the sample size does not meet the requirement for Kruskal-Wallis H/Mann-Whitney *U* test.

- Shinyei sensor does not count the number of particles, but instead, the fraction of time that particles are detected by the photodiode (K.K. Johnson et al., 2018). It sends pulses when particles are detected in the beam, and the fraction of time when pulses occur over the total time is calculated as Low Pulse Occupancy (LPO). Shinyei sensor then applies LPO to estimate particle mass concentration (N.E. Johnson et al., 2018). It has been suggested that the LPO-based technique has larger measurement errors in particle detection, especially under non-laboratory circumstances, in which particles are usually in relatively low concentrations, as compared with the laboratory experiments involving particle dosing (Gao et al., 2015; N.E. Johnson et al., 2018).
- The PM monitoring sensor unit usually comes with a fan inside the detection chamber to draw airflow past the laser beam. However, the air movement inside the Shinyei sensor is driven by resistive heating (Wang et al., 2015). This temperature-gradient-induced flow is strongly affected by ambient thermal conditions and could be less effective in carrying aerosols, especially large particles.

The analysis of $PM_{2.5}$ data collected in laboratory experiments (Fig. 5b(i)) showed that the $\widetilde{r^2}$ of Dylos was slightly lower than that of the Sharp sensors ($p < 0.05$). The deterioration was mainly caused by the unsatisfactory correlation of Dylos sensors ($r^2 = 0.31$) as reported by Vercellino et al. (2018). They conducted the experiments with particle concentrations that exceeded the upper level of qualification (106 particles/cm^3), and thus, a high level of coincidence loss occurred.

In terms of outdoor O_3 monitoring (Fig. 5a(ii)), Aeroqual ($\widetilde{r^2} = 0.87$) performed significantly better than Alphasense ($\widetilde{r^2} = 0.70$). Compared with the EC sensors used by Alphasense, the MOS sensors of Aeroqual are fabricated with a relatively coarse microstructure to mitigate the cross-sensitivity problem (Williams et al., 2009).

4.1.4. Particle sensors correlate better with non-designated reference

Particle monitoring sensors exhibited a better correlation with non-designated reference (Fig. 6a(i)) than with designated reference because low-cost PM sensors (except Shinyei) employ a similar

	$\widetilde{r^2}$	(Q1, Q3)	IQR	n
a) Outdoor				
i) Particle pollutants				
$PM_{2.5}$				
Nova Fitness	0.84^a	(0.78, 0.85)	0.07	9
Plantower	0.82^a	(0.65, 0.90)	0.25	26
Dylos	0.74^a	(0.59, 0.81)	0.22	10
Alphasense	0.72^a	(0.48, 0.85)	0.37	9
Shinyei	0.44^b	(0.35, 0.73)	0.38	19
PM_{10}				
Alphasense	0.68	(0.67, 0.85)	0.18	7
ii) Gaseous pollutants				
CO				
Alphasense	0.82	(0.70, 0.89)	0.19	28
NO				
Alphasense	0.78	(0.63, 0.83)	0.20	19
NO_2				
Alphasense	0.68^a	(0.32, 0.84)	0.52	34
Cairpol	0.65^a	(0.28, 0.76)	0.48	8
Aeroqual	0.24	(0.08, 0.38)	0.30	5
Citytech	0.75	(0.21, 0.88)	0.67	6
SGX	0.33	(0.11, 0.40)	0.29	6
O_3				
Aeroqual	0.87^a	(0.78, 0.91)	0.13	10
Alphasense	0.70^b	(0.48, 0.80)	0.32	25
SO_2				
Alphasense	0.63	(0.18, 0.95)	0.77	6
b) Laboratory				
i) Particle pollutants				
$PM_{2.5}$				
Sharp	0.98^a	(0.98, 0.99)	0.01	11
Dylos	0.94^b	(0.90, 0.98)	0.08	9
Plantower	0.96	(0.95, 0.98)	0.03	7
Syhitech	0.94	(0.91, 0.97)	0.06	7
Shinyei	0.92	(0.71, 0.97)	0.26	7
PM_{10}				
Dylos	0.89	(0.64, 0.97)	0.33	7
ii) Gaseous pollutants				
CO				
Alphasense	0.99	(0.99, 1.00)	0.01	5
NO_2				
Alphasense	1.00	(0.99, 1.00)	0.01	6
O_3				
Alphasense	0.99	(0.99, 1.00)	0.01	5

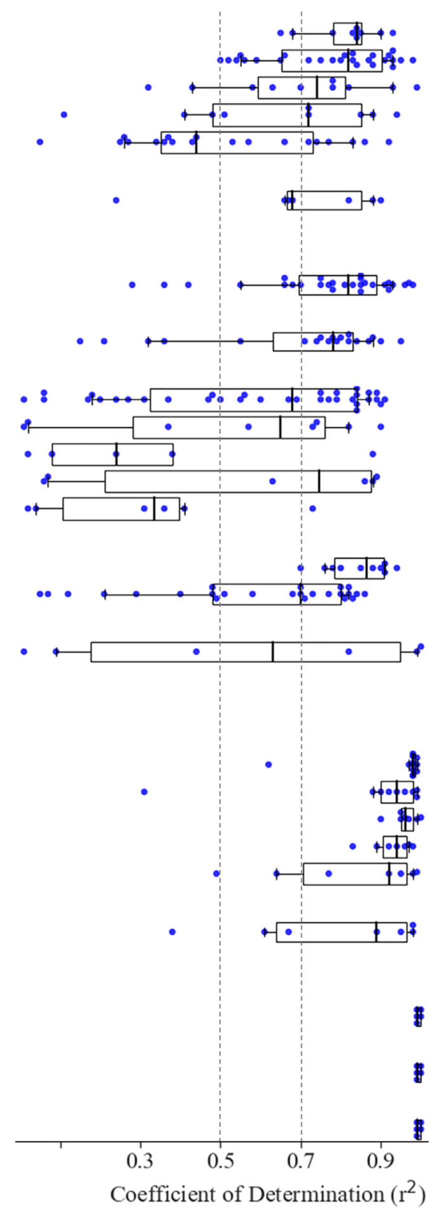


Fig. 5. Influence of original equipment manufacturer on the correlation of sensor with the reference instrument in both a) outdoor and b) laboratory settings. The data classification followed the rules used in Fig. 3, except the comparison was made between sensor manufacturers. Please refer to Fig. 3 for detailed notations. The statistics for levels with the sample size n < 5 are not shown in this figure.

measurement technique as non-designated instruments. They both apply the light scattering technique to count particle numbers in different size ranges (Sousan et al., 2017), which are then converted to mass concentrations with appropriate assumptions on particle density. However, the designated instruments, including beta attenuation (Mukherjee et al., 2017) or tapered element oscillating microbalance (N.E. Johnson et al., 2018), conduct a direct measurement of particle mass concentrations. The authors suspect that stronger correlations with non-designated instruments also exist in non-outdoor testing because the similarity in measurement technique is valid in various environmental settings.

Regarding the reference instrument for gaseous pollutant sensors, designated instrument and standard gas were mostly used for outdoor and laboratory tests, respectively; however, more datasets are required to determine the impact of the reference instrument.

4.1.5. Adjusting sensor readings with advanced regression models improves correlation

As illustrated in Fig. 7a(ii), many sensors for gas exhibited better correlations with the reference when ML models were used to correct the artefact induced by environmental conditions. The use of sLN and mLN models was not so effective because the response of the sensor does not always have a simple linear relationship with environmental parameters. Signals from MOS sensors for gas, for instance, display an exponential dependency on RH (Sohn et al., 2008). For particle sensors, improved correlations were associated with the use of advanced

	$\widetilde{r^2}$	(Q1, Q3)	IQR	n
a) Outdoor				
i) Particle pollutants				
PM_1				
Non-designated	0.90	(0.82, 0.91)	0.09	5
$PM_{2.5}$				
Non-designated	0.85[a]	(0.72, 0.92)	0.20	20
Designated	0.69[b]	(0.50, 0.83)	0.33	60
PM_{10}				
Non-designated	0.88	(0.82, 0.88)	0.06	5
Designated	0.66	(0.40, 0.79)	0.39	15
ii) Gaseous pollutants				
CO				
Designated	0.78	(0.64, 0.87)	0.23	32
NO				
Non-designated	0.75	(0.27, 0.82)	0.55	23
NO_2				
Designated	0.63	(0.24, 0.84)	0.60	53
Non-designated	0.39	(0.28, 0.48)	0.20	6
O_3				
Designated	0.78	(0.51, 0.85)	0.34	48
SO_2				
Designated	0.63	(0.18, 0.95)	0.77	6
b) Laboratory				
i) Particle pollutants				
$PM_{2.5}$				
Non-designated	0.97	(0.92, 0.98)	0.06	45
PM_{10}				
Non-designated	0.95	(0.78, 0.98)	0.20	11
ii) Gaseous pollutants				
NO_2				
Standard gas	1.00	(1.00, 1.00)	0.00	6
O_3				
Standard gas	1.00	(0.99, 1.00)	0.01	6
c) Indoor				
i) Particle pollutants				
$PM_{2.5}$				
Non-designated	0.86	(0.71, 0.90)	0.19	11

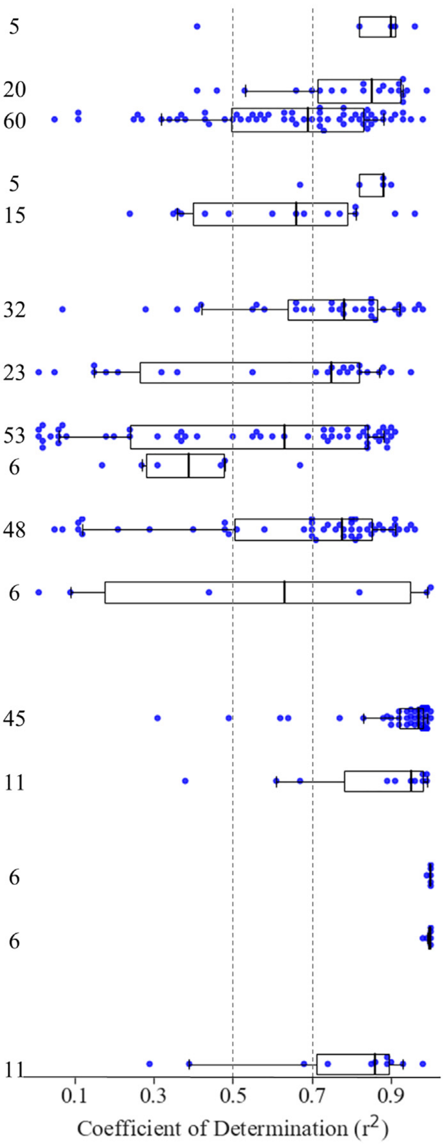


Fig. 6. Influence of reference instruments on the correlation of sensors with the reference instrument in a) outdoor b) laboratory, and c) indoor settings. The data classification followed the rules used in Fig. 3, except the comparison was made between levels in the reference instrument. Please refer to Fig. 3 for detailed notations. The statistics for levels with the sample size n < 5 are not shown in this figure.

regression models, such as ML and nLN models; however, more datasets are required to determine if the difference is statistically significant.

4.2. Root mean square error (RMSE)

Due to limited data on *RMSE*, comparisons were made by means of the matched-pair approach and focused on the impact of two methodological factors (reference instrument and regression model). The results generally followed the trend observed for the r^2 analysis; smaller *RMSE* was found in the experiments with non-designated reference instruments (Table S3) or in the analyses with nLN or ML regression models (Table S4).

The concentrations of pollutants used in sensor evaluation experiments were extracted from published studies and then plotted against *RMSE*. A strong positive relationship existed between these two variables, as evidenced by the plots for Plantower $PM_{2.5}$ sensors (Fig. S2a) and Alphasense NO_2 sensors (Fig. S2b). The authors suspect that this positive relationship may also apply to other low-cost sensors, as *RMSE* is a scale-dependent measure (Hyndman and Koehler, 2006).

4.3. Mean Normalised Bias (MNB) and Coefficient of Variation (CV), recommended by the US EPA guideline

As mentioned in Section 2.4, the US EPA (Williams et al., 2014) proposed to use *MNB* and *CV* to determine the performance of air quality sensors, but out of 112 published studies, only 10 (seven for PM sensors and three for sensors measuring gas) investigated both measures. In terms of PM sensors (Table 1a), Dylos conformed to Tier II: Hotspot Identification and Characterisation and Tier IV: Personal Exposure in all investigations, while the application tier of PM sensors from other OEMs, including Alphasense, Plantower, Sharp, Shinyei, and Syhitech, varied significantly with different particle sources. For example, Sharp

	$\widetilde{r^2}$	(Q1, Q3)	IQR	n
a) Outdoor				
i) Particle pollutants				
$PM_{2.5}$				
Raw	0.72[a]	(0.50, 0.85)	0.35	41
sLN	0.70[a]	(0.55, 0.85)	0.30	17
mLN	0.68[a]	(0.59, 0.83)	0.24	9
nLN	0.93	(0.67, 0.94)	0.27	7
ML	0.75	(0.72, 0.80)	0.08	6
PM_{10}				
Raw	0.67	(0.36, 0.70)	0.34	8
sLN	0.77	(0.46, 0.90)	0.44	7
ii) Gaseous pollutants				
CO				
Raw	0.78	(0.36, 0.85)	0.49	5
sLN	0.74[a]	(0.64, 0.87)	0.23	12
mLN	0.72[a]	(0.54, 0.84)	0.30	8
ML	0.86	(0.79, 0.90)	0.11	6
NO				
sLN	0.74	(0.43, 0.82)	0.39	8
mLN	0.75	(0.15, 0.87)	0.72	5
ML	0.78	(0.32, 0.84)	0.52	5
NO_2				
Raw	0.49[a]	(0.11, 0.84)	0.73	10
sLN	0.39[a]	(0.10, 0.84)	0.74	18
mLN	0.44[a]	(0.25, 0.62)	0.37	18
ML	0.77[b]	(0.70, 0.84)	0.14	10
O_3				
Raw	0.76[a]	(0.21, 0.82)	0.61	11
sLN	0.74[a]	(0.54, 0.85)	0.31	15
mLN	0.64[a]	(0.49, 0.83)	0.34	12
ML	0.82[b]	(0.81, 0.88)	0.07	8
b) Laboratory				
i) Particle pollutants				
$PM_{2.5}$				
Raw	0.97	(0.92, 0.98)	0.06	40
sLN	0.98	(0.95, 0.98)	0.03	5
PM_{10}				
Raw	0.95	(0.78, 0.98)	0.20	11
ii) Gaseous pollutants				
CO				
Raw	0.99	(0.99, 1.00)	0.01	5
NO_2				
Raw	1.00	(0.99, 1.00)	0.01	8
O_3				
Raw	1.00	(0.99, 1.00)	0.01	7
c) Indoor				
i) Particle pollutants				
$PM_{2.5}$				
Raw	0.86	(0.39, 0.89)	0.50	5

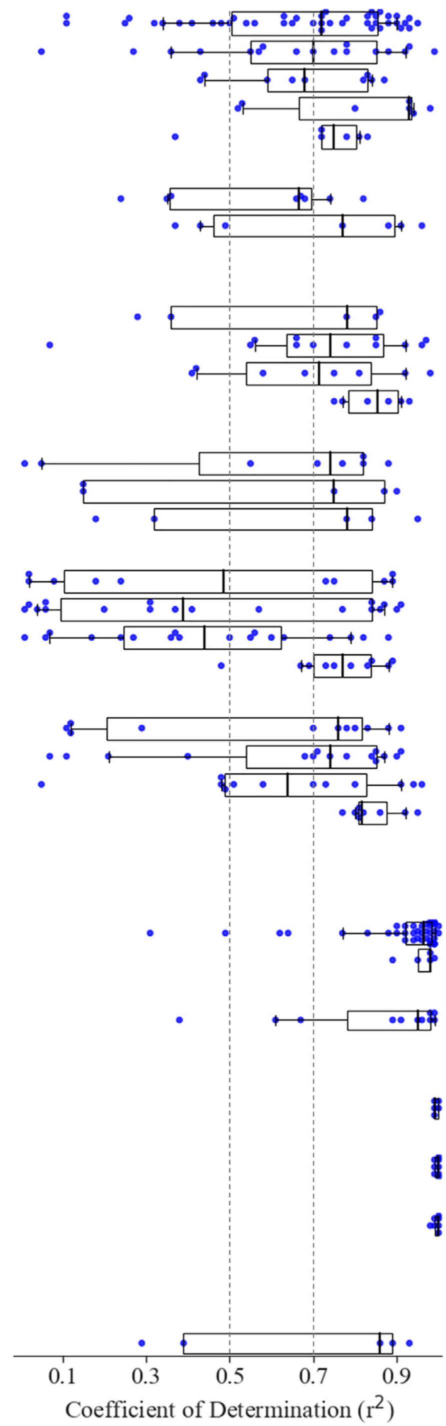


Fig. 7. Influence of regression models on the correlation of sensor with the reference instrument in a) outdoor b) laboratory, and c) indoor settings. The data classification followed the rules used in Fig. 3, except the comparison was made between levels in the regression model. Please refer to Fig. 3 for detailed notations. The statistics for levels with the sample size n < 5 are not shown in this figure.

performed well (met Tier III: Supplemental Monitoring) when exposed to ARD, but displayed large uncertainty when measuring salt (met Tier I: Education and Information) and welding fumes (failed all tiers) (Sousan et al., 2017). The difference in the responses to ARD, salt, and welding fumes was likely due to the prominent variations in the refractive index, which contributed to the variation in sensor detection efficiency (Liu and Daum, 2000).

With regard to sensors for gas pollutants (Table 1b), the existing studies only focused on the compliance of Alphasense sensors with the performance goals for measuring CO, O_3, and NO_2. Even when measuring the same gaseous pollutant, the application tiers of Alphasense sensors varied significantly among different experiments due to the discrepancy in the experimental design. Consistent with the findings of r^2 and *RMSE*, the use of ML models could also improve *MNB* and *CV*. This is supported by the work of Zimmerman et al. (2018): in a similar experimental design, the application tier increased from Tier II/IV to Tier III for O_3 sensors and from Tier V to Tiers II/IV for NO_2 sensors.

5. Conclusions

This investigation reviewed 112 published articles that assessed the validity and reliability of low-cost air quality sensors. A detailed analysis was conducted on statistical measures that described the sensor's correlation with the reference value (r^2), measurement error (*RMSE* and *MNB*), and precision (*CV*). This work also discussed the effects of five methodological factors, including environmental setting, reference instrument, regression model, pollutant attribute, and sensor OEM specification.

The correlation coefficient (r) or its square form (r^2) was the most used measure in previous studies. It was significantly affected by environmental settings, reference instruments, and regression models. However, due to conflicting results and limited data, the effects of pollutant attributes and sensor OEM specifications on sensor correlation could not be conclusively determined. One-third of published studies utilised *RMSE* to analyse the measurement error of low-cost sensors, and the outcomes generally followed the pattern discovered in the correlation analysis. Although *MNB* and *CV* were recommended by US EPA to determine the suitable application areas of low-cost sensors, only a limited number of studies applied these two parameters to characterise the data quality of low-cost sensors.

The findings of this work have a number of important implications for future evaluation experiments on low-cost air quality sensors. First, because sensor performance is sensitive to experimental factors, on-site evaluation is necessary for sensors under conditions as close to the final deployment as possible. Second, although low-cost sensors exhibited improved r^2 and *RMSE* with non-designated reference instruments, the authors suggest that, when conditions permit, future

Table 1
Summary of studies that applied Mean Normalised Bias (*MNB*) and Coefficient of Variation (*CV*), two statistical measures recommended by US EPA (Williams et al., 2014), to evaluate sensor performance.

Article	Sensor OEM specification	Environmental setting	Pollutant attribute	Regression model	Statistical measure		Application tier[a]
					MNB	*CV*	
a) Particle pollutants							
(Sousan et al., 2016a)	Alphasense	Lab	PM_1_ARD	Raw	−0.19	0.15	Tier III
(Sousan et al., 2016a)	Alphasense	Lab	PM_1_combustion	Raw	−0.92	0.10	Failed
(Sousan et al., 2016a)	Alphasense	Lab	PM_1_salt and oxide	Raw	−0.70	0.07	Failed
(Sousan et al., 2016a)	Alphasense	Lab	$PM_{2.5}$_ARD	Raw	−0.02	0.14	Tier III
(Sousan et al., 2016a)	Alphasense	Lab	$PM_{2.5}$_combustion	Raw	−0.91	0.09	Failed
(Sousan et al., 2016a)	Alphasense	Lab	$PM_{2.5}$_ salt and oxide	Raw	−0.57	0.05	Failed
(Sousan et al., 2016a)	Alphasense	Lab	PM_{10}_ARD	Raw	0.63	0.16	Failed
(Sousan et al., 2016a)	Alphasense	Lab	PM_{10}_combustion	Raw	−0.89	0.06	Failed
(Sousan et al., 2016a)	Alphasense	Lab	PM_{10}_ salt and oxide	Raw	−0.55	0.04	Failed
(Sousan et al., 2016b)	Dylos	Lab	PM_{10}_ARD	Raw	−0.18	0.03	Tier III
(Sousan et al., 2016b)	Dylos	Lab	PM_{10}_combustion	Raw	−0.03	0.08	Tier V
(Sousan et al., 2016b)	Dylos	Lab	PM_{10}_ salt and oxide	Raw	0.04	0.01	Tier V
(Jones et al., 2016)	Dylos	Indoor	PM_{10}	sLN	−0.02	0.08	Tier V
(Carvlin et al., 2017)	Dylos	Outdoor	$PM_{2.5}$	sLN	−0.05	0.25	Tier II/IV
(Sayahi et al., 2019b)	Plantower	Lab	$PM_{2.5}$_ salt and oxide	sLN	−0.49	0.04	Tier I
(Zamora et al., 2020)	Plantower	Indoor	$PM_{2.5}$	Raw	0.59	1.60	Failed
(Hong et al., 2021)	Plantower	Outdoor	$PM_{2.5}$	Raw	0.43	0.05	Tier I
(Sousan et al., 2017)	Sharp	Lab	$PM_{2.5}$_ARD	Raw	−0.12	0.08	Tier III
(Sousan et al., 2017)	Sharp	Lab	$PM_{2.5}$_combustion	Raw	−0.82	0.08	Failed
(Sousan et al., 2017)	Sharp	Lab	$PM_{2.5}$_ salt and oxide	Raw	−0.46	0.05	Tier I
(Sousan et al., 2017)	Shinyei	Lab	$PM_{2.5}$_ARD	Raw	−0.53	0.02	Failed
(Sousan et al., 2017)	Shinyei	Lab	$PM_{2.5}$_combustion	Raw	−0.83	0.09	Failed
(Sousan et al., 2017)	Shinyei	Lab	$PM_{2.5}$_ salt and oxide	Raw	−0.36	0.04	Tier I
(Sousan et al., 2017)	Syhitech	Lab	$PM_{2.5}$_ARD	Raw	0.18	0.12	Tier III
(Sousan et al., 2017)	Syhitech	Lab	$PM_{2.5}$_combustion	Raw	−0.86	0.08	Failed
(Sousan et al., 2017)	Syhitech	Lab	$PM_{2.5}$_ salt and oxide	Raw	−0.68	0.25	Failed
(Zamora et al., 2020)	Syhitech	Indoor	$PM_{2.5}$	Raw	3.90	0.52	Failed
b) Gaseous pollutants							
(Afshar-Mohajer et al., 2018)	Alphasense	Lab	CO	Raw	0.03	0.03	Tier V
(Afshar-Mohajer et al., 2018)	Alphasense	Lab	NO_2	Raw	−0.24	0.18	Tier II/IV
(Afshar-Mohajer et al., 2018)	Alphasense	Lab	O_3	Raw	0.10	0.17	Tier III
(Zimmerman et al., 2018)	Alphasense	Outdoor	CO	sLN	0.23	0.22	Tier II/IV
(Zimmerman et al., 2018)	Alphasense	Outdoor	CO	mLN	0.12	0.14	Tier III
(Zimmerman et al., 2018)	Alphasense	Outdoor	CO	ML	0.11	0.12	Tier III
(Malings et al., 2019a)	Alphasense	Outdoor	CO	ML	0.16	0.18	Tier III
(Zimmerman et al., 2018)	Alphasense	Outdoor	NO_2	sLN	1.60	0.75	Failed
(Zimmerman et al., 2018)	Alphasense	Outdoor	NO_2	mLN	0.32	0.28	Tier V
(Zimmerman et al., 2018)	Alphasense	Outdoor	NO_2	ML	0.27	0.25	Tier II/IV
(Malings et al., 2019a)	Alphasense	Outdoor	NO_2	ML	0.25	0.28	Tier II/IV
(Zimmerman et al., 2018)	Alphasense	Outdoor	O_3	mLN	0.22	0.33	Tier II/IV
(Zimmerman et al., 2018)	Alphasense	Outdoor	O_3	ML	0.13	0.16	Tier III
(Malings et al., 2019a)	Alphasense	Outdoor	O_3	ML	0.20	0.26	Tier II/IV

[a] EPA suggested performance goals by application for *MNB* and *CV* (Williams et al., 2014): Tier I: Education and Information ($-0.5 < MNB < 0.5$ and $CV < 0.5$ for all pollutants); Tier II: Hotspot Identification and Characterisation ($-0.3 < MNB < 0.3$ and $CV < 0.3$ for all pollutants); Tier III: Supplemental Monitoring ($-0.2 < MNB < 0.2$ and $CV < 0.2$ for all pollutants); Tier IV: Personal Exposure ($-0.3 < MNB < 0.3$ and $CV < 0.3$ for all pollutants); Tier V: Regulatory Monitoring ($-0.07 < MNB < 0.07$ and $CV < 0.07$ for O_3; $-0.1 < MNB < 0.1$ and $CV < 0.1$ for CO and $PM_{2.5}$; $-0.15 < MNB < 0.15$ and $CV < 0.15$ for NO_2).

studies should assess sensors against designated reference instruments that meet national accreditation standards. Third, sensor calibration that uses advanced regression models (i.e. nLN and ML models) is better than the one that employs raw data, sLN or mLN models because advanced models can correct the non-linear effect of environmental conditions on sensor performance. Finally, considering the lack of uniformity in the statistical measures among existing studies, more efforts are needed to develop a common standardised protocol to examine low-cost sensors at the international level for data comparability in the future.

CRediT authorship contribution statement

Ye Kang: Methodology, Formal analysis, Investigation, Data curation, Writing - original draft, Visualisation. **Lu Aye**: Validation, Writing - review & editing. **Tuan Duc Ngo**: Validation, Writing - review & editing. **Jenny Zhou**: Conceptualisation, Methodology, Validation, Resources, Writing - review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

Acknowledgements

This research was supported by the Building 4.0 CRC program and Commonwealth Scientific and Industrial Research Organisation.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2021.151769.

References

Afshar-Mohajer, N., Zuidema, C., Sousan, S., Hallett, L., Tatum, M., Rule, A.M., Thomas, G., Peters, T.M., Koehler, K., 2018. Evaluation of low-cost electro-chemical sensors for environmental monitoring of ozone, nitrogen dioxide, and carbon monoxide. J. Occup. Environ. Hyg. 15, 87–98.

Antonini, J.M., Taylor, M.D., Zimmer, A.T., Roberts, J.R., 2004. Pulmonary responses to welding fumes: role of metal constituents. J. Toxic. Environ. Health A 67, 233–249.

Borghi, F., Spinazze, A., Rovelli, S., Campagnolo, D., Del Buono, L., Cattaneo, A., Cavallo, D.M., 2017. Miniaturized monitors for assessment of exposure to air pollutants: a review. Int. J. Environ. Res. Public Health 14, 909.

Carvlin, G.N., Lugo, H., Olmedo, L., Bejarano, E., Wilkie, A., Meltzer, D., Wong, M., King, G., Northcross, A., Jerrett, M., English, P.B., Hammond, D., Seto, E., 2017. Development and field validation of a community-engaged particulate matter air quality monitoring network in Imperial, California, USA. J. Air Waste Manage. Assoc. 67, 1342–1352.

Clements, A.L., Griswold, W.G., Rs, A., Johnston, J.E., Herting, M.M., Thorson, J., Collier-Oxandale, A., Hannigan, M., 2017. Low-cost air quality monitoring tools: from research to practice (a workshop summary). Sensors (Basel) 17, 2478.

Cross, E.S., Williams, L.R., Lewis, D.K., Magoon, G.R., Onasch, T.B., Kaminsky, M.L., Worsnop, D.R., Jayne, J.T., 2017. Use of electrochemical sensors for measurement of air pollution: correcting interference response and validating measurements. Atmos. Meas. Tech. 10, 3575–3588.

Duvall, R.M., Long, R.W., Beaver, M.R., Kronmiller, K.G., Wheeler, M.L., Szykman, J.J., 2016. Performance evaluation and community application of low-cost sensors for ozone and nitrogen dioxide. Sensors (Basel) 16, 1698.

Gao, M., Cao, J., Seto, E., 2015. A distributed network of low-cost continuous reading sensors to measure spatiotemporal variations of PM2.5 in Xi'an, China. Environ. Pollut. 199, 56–65.

Gore, S.M., Altman, D.G., 1982. Statistics in Practice. British Medical Association, London.

Hall, E.S., Kaushik, S.M., Vanderpool, R.W., Duvall, R.M., Beaver, M.R., Long, R.W., Solomon, P.A., 2014. Integrating sensor monitoring technology into the current air pollution regulatory support paradigm: practical consideration. Am. J. Environ. Sci. 4, 147–154.

Hong, G.-H., Le, T.-C., Tu, J.-W., Wang, C., Chang, S.-C., Yu, J.-Y., Lin, G.-Y., Aggarwal, S.G., Tsai, C.-J., 2021. Long-term evaluation and calibration of three types of low-cost PM2.5 sensors at different air quality monitoring stations. J. Aerosol Sci. 157 (105829).

Hyndman, R.J., Koehler, A.B., 2006. Another look at measures of forecast accuracy. Int. J. Forecast. 22, 679–688.

Johnson, K.K., Bergin, M.H., Russell, A.G., Hagler, G.S.W., 2018a. Field test of several low-cost particulate matter sensors in high and low concentration urban environments. Aerosol Air Qual. Res. 18, 565–578.

Johnson, N.E., Bonczak, B., Kontokosta, C.E., 2018b. Using a gradient boosting model to improve the performance of low-cost aerosol monitors in a dense, heterogeneous urban environment. Atmos. Environ. 184, 9–16.

Jones, S., Anthony, T.R., Sousan, S., Altmaier, R., Park, J.H., Peters, T.M., 2016. Evaluation of a low-cost aerosol sensor to assess dust concentrations in a swine building. Ann. Occup. Hyg. 60, 597–607.

Jovasevic-Stojanovic, M., Bartonova, A., Topalovic, D., Lazovic, I., Pokric, B., Ristovski, Z., 2015. On the use of small and cheaper sensors and devices for indicative citizen-based monitoring of respirable particulate matter. Environ. Pollut. 206, 696–704.

Karagulian, F., Barbiere, M., Kotsev, A., Spinelle, L., Gerboles, M., Lagler, F., Redon, N., Crunaire, S., Borowiak, A., 2019. Review of the performance of low-cost sensors for air quality monitoring. Atmosphere 10, 506.

Kelly, K.E., Whitaker, J., Petty, A., Widmer, C., Dybwad, A., Sleeth, D., Martin, R., Butterfield, A., 2017. Ambient and laboratory evaluation of a low-cost particulate matter sensor. Environ. Pollut. 221, 491–500.

Kularatna, N., Sudantha, B.H., 2008. An environmental air pollution monitoring system based on the IEEE 1451 standard for low cost requirements. IEEE Sens. J. 8, 415–422.

Kumar, P., Morawska, L., Martani, C., Biskos, G., Neophytou, M., Sabatino, S.D., Bell, M., Norford, L., Britter, R., 2015. The rise of low-cost sensing for managing air pollution in cities. Environ. Int. 75, 199–205.

Kumar, P., Skouloudis, A.N., Bell, M., Viana, M., Carotta, M.C., Biskos, G., Morawska, L., 2016. Real-time sensors for indoor air monitoring and challenges ahead in deploying them to urban buildings. Sci. Total. Environ. 560, 150–159.

Liu, Y., Daum, P.H., 2000. The effect of refractive index on size distributions and light scattering coefficients derived from optical particle counts. J. Aerosol Sci. 31, 945–957.

Makri, A., Stilianakis, N.I., 2008. Vulnerability to air pollution health effects. Int. J. Hyg. Environ. Health 211, 326–336.

Malings, C., Tanzer, R., Hauryliuk, A., Kumar, S.P.N., Zimmerman, N., Kara, L.B., Presto, A.A., 2019a. Development of a general calibration model and long-term performance evaluation of low-cost sensors for air pollutant gas monitoring. Atmos. Meas. Tech. 12, 903–920.

Malings, C., Tanzer, R., Hauryliuk, A., Saha, P.K., Robinson, A.L., Presto, A.A., Subramanian, R., 2019b. Fine particle mass monitoring with low-cost sensors: corrections and long-term performance evaluation. Aerosol Sci. Technol. 54, 160–174.

McKercher, G.R., Jennifer, A.S., Vnos, J.K., 2017. Characteristics and applications of small, portable gaseous air pollution monitors. Environ. Pollut. 223, 102–110.

Mead, M.I., Popoola, O.A.M., Stewart, G.B., Landshoff, P., Calleja, M., Hayes, M., Baldovi, J.J., McLeod, M.W., Hodgson, T.F., Dicks, J., Lewis, A., Cohen, J., Baron, R., Saffell, J.R., Jones, R.L., 2013. The use of electrochemical sensors for monitoring urban air quality in low-cost, high-density networks. Atmos. Environ. 70, 186–203.

Moore, D.S., Notz, W.I., Flinger, M.A., 2013. The Basic Practice of Statistics. WH Freeman, New York.

Morawska, L., Thai, P.K., Liu, X., Asumadu-Sakyi, A., Ayoko, G., Bartonova, A., Bedini, A., Chai, F., Christensen, B., Dunbabin, M., Gao, J., Hagler, G.S.W., Jayaratne, R., Kumar, P., Lau, A.K.H., Louie, P.K.K., Mazaheri, M., Ning, Z., Motta, N., Mullins, B., Rahman, M.M., Ristovski, Z., Shafiei, M., Tjondronegoro, D., Westerdahl, D., Williams, R., 2018. Applications of low-cost sensing technologies for air quality monitoring and exposure assessment: how far have they gone? Environ. Int. 116, 286–299.

Mukherjee, A., Stanton, L.G., Graham, A.R., Roberts, P.T., 2017. Assessing the utility of low-cost particulate matter sensors over a 12-week period in the Cuyama Valley of California. Sensors (Basel) 17, 1805.

Pal, R., 2017. Validation methodologies. Predictive Modeling of Drug Sensitivity.

Pearce, N., 2016. Analysis of matched case-control studies. BMJ 352, i969.

Rai, A.C., Kumar, P., Pilla, F., Skouloudis, A.N., Di Sabatino, S., Ratti, C., Yasar, A., Rickerby, D., 2017. End-user perspective of low-cost sensors for outdoor air pollution monitoring. Sci. Total Environ. 607–608, 691–705.

Sayahi, T., Butterfield, A., Kelly, K.E., 2019a. Long-term field evaluation of the plantower PMS low-cost particulate matter sensors. Environ. Pollut. 245, 932–940.

Sayahi, T., Kaufman, D., Becnel, T., Kaur, K., Butterfield, A.E., Collingwood, S., Zhang, Y., Gaillardon, P.E., Kelly, K.E., 2019b. Development of a calibration chamber to evaluate the performance of low-cost particulate matter sensors. Environ. Pollut. 255, 113131.

Snyder, E.G., Watkins, T.H., Solomon, P.A., Thoma, E.D., Williams, R.W., Hagler, G.S.W., Shelow, D., Hindin, D.A., Kilaru, V.J., Preuss, P.W., 2013. The changing paradigm of air pollution monitoring. Environ. Sci. Technol. Lett. 47, 11369–11377.

Sohn, J.H., Atzeni, M., Zeller, L., Pioggia, G., 2008. Characterisation of humidity dependence of a metal oxide semiconductor sensor array using partial least squares. Sens. Actuators B Chem. 131, 230–235.

Sousan, S., Koehler, K., Hallett, L., Peters, T.M., 2016a. Evaluation of the alphasense optical particle counter (OPC-N2) and the grimm portable aerosol spectrometer (PAS-1.108). Aerosol Sci. Technol. 50, 1352–1365.

Sousan, S., Koehler, K., Thomas, G., Park, J.H., Hillman, M., Halterman, A., Peters, T.M., 2016b. Inter-comparison of low-cost sensors for measuring the mass concentration of occupational aerosols. Aerosol Sci. Technol. 50, 462–473.

Sousan, S., Koehler, K., Hallett, L., Peters, T.M., 2017. Evaluation of consumer monitors to measure particulate matter. J. Aerosol Sci. 107, 123–133.

Spinelle, L., Gerboles, M., Aleixandre, M., 2015a. Performance evaluation of amperometric sensors for the monitoring of O 3 and NO 2 in ambient air at ppb level. Procedia Eng. 120, 480–483.

Spinelle, L., Gerboles, M., Villani, M.G., Aleixandre, M., Bonavitacola, F., 2015b. Field calibration of a cluster of low-cost available sensors for air quality monitoring. Part a: ozone and nitrogen dioxide. Sensors Actuators B Chem. 215, 249–257.

Sun, L., Wong, K.C., Wei, P., Ye, S., Huang, H., Yang, F., Westerdahl, D., Louie, P.K., Luk, C.W., Ning, Z., 2016. Development and application of a next generation air sensor network for the Hong Kong marathon 2015 air quality monitoring. Sensors (Basel) 16, 211.

US Environmental Protection Agency, 2017. List of Designated Reference and Equivalent Methods.

Vercellino, R.J., Sleeth, D.K., Handy, R.G., Min, K.T., Collingwood, S.C., 2018. Laboratory evaluation of a low-cost, real-time, aerosol multi-sensor. J. Occup. Environ. Hyg. 15, 559–567.

Wang, Y., Li, J., Jing, H., Zhang, Q., Jiang, J., Biswas, P., 2015. Laboratory evaluation and calibration of three low-cost particle sensors for particulate matter measurement. Aerosol Sci. Technol. 49, 1063–1077.

Williams, D.E., Salmond, J., Yung, Y.F., Akaji, J., Wright, B., Wilson, J., Laing, G., 2009. Development of low-cost ozone and nitrogen dioxide measurement instruments suitable for use in an air quality monitoring network. Sensors, 2009 IEEE.

Williams, R., Kilaru, V., Snyder, E., Kaufman, A., Dye, T., Rutter, A., Russell, A., Hafner, H., 2014. Air Sensor Guidebook. US Environmental Protection Agency.

Zamora, M.L., Rice, J., Koehler, K., 2020. One year evaluation of three low-cost PM2.5 monitors. Atmos. Environ. 235 (117615).

Zimmerman, N., Presto, A.A., Kumar, S.P.N., Gu, J., Hauryliuk, A., Robinson, E.S., Robinson, A.L., 2018. A machine learning calibration model using random forests to improve sensor performance for lower-cost air quality monitoring. Atmos. Meas. Tech. 11, 291–313.