

Article

Not peer-reviewed version

Air Pollution Monitoring using Cost-Effective Devices

[Yanis Colléaux](#), [Cédric Willaume](#), [Bijan Mohandes](#), [Jean-Christophe Nebel](#)^{*}, Farzana Rahman

Posted Date: 5 February 2025

doi: 10.20944/preprints202502.0210.v1

Keywords: air pollution monitoring; low-cost sensors; electrochemical sensors; non-dispersive infrared sensors; sensor performance variability; measurement correction; sensor calibration; datadriven correction; machine learning; multiple linear regression models



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Air Pollution Monitoring using Cost-Effective Devices

Yanis Colléaux¹, Cédric Willaume², Bijan Mohandes³, Jean-Christophe Nebel^{4,*} 
and Farzana Rahman⁴ 

¹ National School for Statistics and Data Analysis (ENSAI), Blaise Pascal BP37203, 35172 Bruz cedex, France; yanis.colleaux@eleve.ensai.fr

² National Graduate School of Engineering of Caen (ENSICAEN), 6 boulevard Maréchal Juin – CS 45053, 14050 Caen cedex 4, France; cedric.willaume@ecole.ensicaen.fr

³ Technocomm Consulting Ltd., The Barn, Highwood Farm, Long Lane, Newbury, Berkshire, RG14 2TB, United Kingdom; bijan@technocommconsulting.com

⁴ Kingston University, Holmwood House, Grove Crescent, Kingston upon Thames, Surrey, KT1 2EE, United Kingdom; farzana@kingston.ac.uk

* Correspondence: j.nebel@kingston.ac.uk

Abstract: Given the significant impact of air pollution on global health, continuous and precise monitoring of air quality in all populated environments is crucial. Unfortunately, even in the most developed economies, current air quality monitoring networks are largely inadequate. The high cost of monitoring stations has been identified as a key barrier to widespread coverage, making cost-effective air quality monitoring devices a potential game changer. However, the accuracy of measurements obtained from low-cost sensors is affected by many factors, including gas cross-sensitivity, environmental conditions, and production inconsistencies. Fortunately, machine learning models can capture complex interdependent relationships in sensor responses, and thus can enhance their readings and sensor accuracy. After gathering measurements from cost-effective air pollution monitoring devices placed alongside a reference station, the data were used to train such models. Assessment of their performance showed that models tailored to individual sensor units greatly improved measurement accuracy, boosting their correlation with reference-grade instruments by up to 10%. Nonetheless, the research also revealed that inconsistencies in the performance of similar sensor units can prevent the creation of a unified correction model for a given sensor type.

Keywords: air pollution monitoring; low-cost sensors; electrochemical sensors; non-dispersive infrared sensors; sensor performance variability; measurement correction; sensor calibration; data-driven correction; machine learning; multiple linear regression models

1. Introduction

Air pollution has significant implications for public health, climate change, and urban planning [1–4]. According to the World Health Organization, air pollution is responsible for nearly 7 million premature deaths yearly and is the second highest risk factor for noncommunicable diseases [5,7–12]. To tackle this critical global issue, it is essential to implement effective monitoring strategies and eventually take action to mitigate pollution [2,6,13–17]. Although the first air quality monitoring stations were established in the 1950s [18], and nowadays most major cities have such systems, the financial burden and limited accessibility of suitable instrumentation have posed significant barriers to these efforts [4,19]. The high cost of monitoring systems is attributed to the need to invest in expensive reference-grade instruments to ensure accurate measurements [20,21]. In addition, these instruments have suffered from limited spatial coverage due to their lack of portability, in addition to placement and availability issues [22].

In response to these challenges, cost-effective air quality monitoring devices have been developed using low-cost sensors [23]. The low-cost sensors are not only significantly cheaper than traditional stations, but also typically occupy less space, and can be mobile [24]. In recent years, affordable

devices have emerged as viable alternatives, facilitating wider deployment and frequently enabling near-real-time data collection. Nonetheless, these sensors still face challenges related to calibration, accuracy, reliability, and standardisation [25,26].

Although improving the hardware of affordable sensors is challenging without significantly increasing costs or compromising practicality, leveraging software to enhance their performance is a highly promising approach [27]. For instance, by comprehending how external factors affect sensor measurements, it becomes feasible to apply corrections that yield more reliable values [28,39]. Ultimately, the goal is to improve the observation quality of these devices to a level where they can perform tasks that are currently exclusive to traditional monitoring stations[29].

The objective of this study is to assess whether machine learning algorithms implemented on a microcontroller can improve readings from a set of low-cost sensors designed to measure CO, O₃, and CO₂ concentrations. This manuscript begins by reviewing the benefits of cost-effective air pollution monitoring devices, their current technical limitations, and methods to improve their measurement accuracy. It then details the sensors under investigation and the data collection process. Following this, a comprehensive description of the selected methodology to enhance measurement accuracy is given. Subsequently, experimental results are presented. Finally, the manuscript concludes with discussions and a summary of key conclusions.

2. Related Work

The development and availability of cost-effective air pollution monitoring devices have instigated a significant body of research dedicated to assessing their suitability for the intended applications. While the literature underscores their remarkable potential, it also highlights their inherent technical limitations. Fortunately, advancements in machine learning have enabled the development of more intelligent sensors, capable of accounting for factors such as sensor drift and cross-sensitivity, thereby delivering more accurate corrected readings.

2.1. Advantages of Cost-Effective Air Pollution Monitoring Devices

In large population centres, such as Shanghai, Hong Kong, Delhi, Dhaka, Dallas, and London, that suffer from severe air pollution challenges [36,37], often exacerbated by rapid urbanisation, industrial activities, and vehicular emissions [38], usage of traditional air quality monitoring stations is indispensable yet insufficient on their own [41,42]. As of today, London has only 83 air quality monitoring stations, fewer than one station per 100,000 inhabitants, which is clearly insufficient given that air quality can vary significantly between neighbouring streets [43]. Coverage is even lower in countries that are the most affected by air pollution, e.g., there are only 11 stations in Bangladesh. Indeed, a main obstacle to station deployment is their cost, which is typically over \$10,000 per unit without considering installation and maintenance charges. As low-cost sensors have shown promise in addressing the increasingly pressing challenge of urban air pollution [34], their deployment, either in tandem with or independently from existing networks [44], can achieve better spatial coverage, improved data granularity, and eventually deliver more timely interventions [4,37].

The primary appeal of these cost-effective devices lies in their scalability and the flexibility of their deployment. Indeed, their competitive cost allows for far more extensive deployment than would be feasible with expensive reference-grade monitors [27]. In sprawling metropolises, the availability of only a limited number of official monitoring stations cannot adequately capture localised pollution hotspots or micro-environments [41], resulting in data blind spots. Low-cost sensors, on the other hand, can be deployed in more significant numbers within residential districts, near industrial zones, and along congested traffic corridors [45]. This expanded network translates into a more precise understanding of how pollution varies across neighbourhoods, enabling policymakers to tailor interventions to the most affected areas [37].

Second, the portability and relatively small form factor of these devices make them far easier to install and maintain. While larger, reference-grade stations may require substantial ground space, stable power sources [4], and climate control units, the sensors can be placed on rooftops, street lights,

or moving vehicles [27]. This ease of placement is especially valuable in densely packed cities such as Hong Kong and London, where real estate is at a premium and infrastructural constraints are significant. Moreover, while traditional stations often require regular visits from specialised personnel, these sensors can often be serviced, calibrated, and updated remotely, reducing operational costs and minimising downtime [46].

Third, as cost-effective air pollution monitoring devices are usually integrated with an Internet of Things (IoT) platform, real-time or near-real-time data streams from a dense network can be invaluable for immediate decision-making. In cities prone to episodic pollution events, such as the winter smog in Delhi or the dust storms that occasionally affect Dallas, these data allow local authorities to promptly issue health advisories [47], divert traffic flows, or temporarily restrict polluting activities [29]. The ability to respond proactively can mitigate the severity of pollution events and reduce residents' exposure to harmful particulate matter and gases [37,48].

Fourth, the affordability and user-friendliness of these devices may encourage citizen engagement [4] and foster collaborative monitoring, particularly in regions where public awareness is crucial for policy support. Initiatives such as Breathe London [49] and a community-led effort in Hong Kong [50] demonstrate how crowd-source air quality data can help individuals manage air pollution exposure and related health risk at urban scale. Large scale deployment of low-cost sensors also allows tracking local pollution patterns in real time [27], and holding industries or municipal bodies accountable when levels exceed safe thresholds [51]. This democratisation of data creates a virtuous cycle of engagement [52], awareness, and policy responsiveness, thereby accelerating improvements in air quality management [37].

Finally, cost-effective air pollution monitoring aligns with several United Nations Sustainable Development Goals [53]. As large population centres face increasing pressure from rapid urbanisation, climate change, and public health crises, these sensors support cities in meeting international guidelines, reducing environmental inequalities, and safeguarding public health. Their ability to integrate seamlessly with digital platforms means local authorities, researchers, and citizens alike could access actionable insights, plan more sustainable infrastructure, and take steps to reduce pollution at its source.

2.2. Current Technical Limitations

Even if they have numerous advantages, cost-effective air pollution monitoring devices have limitations when compared against their reference-grade counterparts. Specifically, they depend on the technology of their individual sensors that can vary dramatically according to their target pollutants.

Whereas particulate matter is usually quantified by optical particle counters (OPCs) that focus a beam of light within the sensor and record its scattering using a photodetector, low-cost gas-sensing technologies can be classified into four main types [54]:

- Electrochemical (EC) sensors measure the current produced by electrochemical reactions with the target gas [55].
- Non-dispersive infrared (NDIR) sensors track reductions in infrared radiation when the gas passes through an active filter [56].
- Metal oxide semiconductor (MOS) sensors rely on gas-solid interactions that induce an electronic charge on the metal oxide surface.
- Photo-ionisation detection (PID) sensors employ ultraviolet light to ionise target molecules, converts the resulting ions into digital reading, and thus quantifies the chemical content [57].

Despite the variety of these technologies, these sensors experience inaccuracies caused by the following factors: sensor drift and ageing, cross-sensitivity and interferences, manufacturing inconsistencies and lack of standards, environmental factors, and limited dynamic range and saturation effects.

Due to the low-cost nature of these sensors, they are all subject to batch-to-batch variations in component quality, such as slight chemical deviations in sensor electrodes, which can shift baselines,

introduce offset misalignments, and cause inconsistencies in sensitivity from unit to unit [20,46]. Furthermore, many sensors are shipped with minimal or no factory calibration, leaving users to rely on in-field calibration that may be inconsistent or inadequate. Even when calibration is provided, manufacturers may use unspecified and/or proprietary algorithms [4]. Moreover, when they provide correction equations, they are based on tests conducted in laboratories, where conditions are controlled and simplified, which often does not reflect real-world conditions and fail to account for complex interactions between different pollutants. Compounding this problem is the absence of unified testing guidelines or standardised protocols. Thus, one manufacturer's 'validated' data might not be comparable to another's [22,58].

Another general characteristics of low-cost sensors is that their enclosure may lack adequate ingress protection, making them susceptible to water damage from rain or dew [59,60]. Furthermore, these enclosures often cannot endure prolonged exposure to ultraviolet light. In urban settings, high levels of electromagnetic interference can cause noise in the sensor's electronics, especially if sufficient shielding is not provided. Furthermore, corrosion and oxidation can damage the metal components, negatively impacting the sensor's electrical performance. Finally, low-cost sensors often produce raw signals rife with high-frequency variability, outliers, and artefacts brought on by electrical spikes or mechanical vibrations.

In addition to these general technical limitations, cost-effective air pollution monitoring devices suffer from a range of factors that are specific to their working mechanisms.

Optical particle counters

The main issue with optical particle counters is lens obstruction from dust or soot, skewing scattering intensity [61], and particulate accumulation in the sampling path compromising airflow, reducing consistency of particle detection [62]. Consequently, regular calibration is needed, sometimes monthly or quarterly, which increases operational costs and complexity [63]. Additionally, condensation within the sampling chamber can distort scattering profiles and a sudden surge of particulates can clog the measurement chamber leading to transient errors. Another issue is that particles of diverse sizes and compositions may scatter light differently, making it difficult to classify them accurately by size. Moreover, as aerosols can absorb moisture, artificially inflating particle diameter and varying refractive indices, high relative humidity can lead to overestimated concentrations [26].

Electrochemical sensors

Over time, electrochemical sensors experience degradation as their reactive electrolyte or electrode materials are consumed, altering their output. Whereas some sensors degrade within months, others maintain partial functionality for two or three years while accumulating biases [64]. Moreover, prolonged exposure to high or low temperatures and impurity can amplify drift by accelerating chemical reactions, depleting internal reagents or causing them to form condensation [65–68]. Finally, it has also been reported that ageing influences bias in voltage recording at certain environmental ozone concentrations [69].

In addition to ageing, EC sensors often register cross-reactants, such as NO₂ sensors also responding to O₃, Cl₂, and H₂S, which inflates or distorts measurements [27,70,71]. High pollutant concentrations can also saturate electrochemical sensors, leaving the electrode's reaction rate plateaued. Moreover, shifts in humidity and temperature can trigger spurious voltage changes or accelerate side reactions in the sensing chamber [72]. In particular, they are vulnerable to interference from water vapour, which, by occupying reactive sites on the metal oxide surface, leads to underestimations of pollutant levels [73–76].

Non-dispersive infrared sensors

Despite their growing popularity, non-dispersive infrared sensors face several inherent limitations that influence data fidelity, sensor lifespan, and overall utility in real-world deployments [77]. Depending on the quality of their filter design, they may be particularly vulnerable to interference from other infrared-absorbing gases. Such cross-sensitivity can inflate or depress readings, based on the overlap in absorption spectra. Furthermore, condensation or particulate deposits on the detector

degrade performance, diminishing the sensor's responsiveness and reliable range [78,79]. Moreover, these sensors are prone to calibration drift, which is exacerbated by environmental factors such as temperature and humidity fluctuations [80,81].

While high-end NDIR systems employ sophisticated temperature-compensation circuits and superior optical filters, low-cost NDIR sensors lack such refinements, resulting in lower measurement accuracy [81]. Due to their limited temperature compensation capability, even minimal temperature changes can introduce sizeable measurement errors over time. Therefore, if recalibration is not performed regularly, readings can drift far from reality. Eventually, in the absence of a robust enclosure and/or advanced compensation algorithms, their data quality may be questionable in real-world deployments, i.e., where environmental conditions can vary widely [73,74].

Photo-ionisation detection sensor

Due to their fast response times and broad-spectrum sensitivity, photo-ionisation detection sensors are widely used, especially for detecting volatile organic compounds (VOCs). However, selectivity is a major concern as they ionise all compounds with an ionisation potential below the lamp's energy (typically 10.6 eV) [82]. This leads to interference from multiple VOCs, such as benzene, toluene and xylene, which cannot be distinguished without additional separation techniques [57,83]. Additionally, sensor drift and stability degradation are key operational limitations, with reported drift values ranging between 2% and 15% per year [83]. Humidity and temperature variations also significantly affect PID performance, necessitating frequent recalibration [84]. Moreover, studies have shown that sensor accuracy degrades beyond acceptable error margins when exposed to relative humidity variations above 50% [67,82]. Furthermore, their limit of detection varies significantly from sub-ppb levels to above 2.5 ppm depending on the model and manufacturer [83]. Eventually, real-world performance validation has highlighted discrepancies between laboratory-calibrated PIDs and field-deployed versions, with studies indicating deviations of up to $\pm 25\%$ when compared with reference monitoring stations [67].

From an economic perspective, maintenance costs and power consumption present additional barriers to large-scale deployment. Indeed, they require periodic UV lamp replacement, increasing long-term operational costs beyond those of other LCS technologies such as MOS or EC sensors [83]. Additionally, power requirements are significantly higher due to the UV excitation mechanism, making them less suitable for battery-powered applications in mobile or remote monitoring systems.

Metal oxide semiconductor sensors

MOS sensors are known to deliver very high sensitivity, but usually lack selectivity. Thus, they tend to group chemically similar volatile organic compounds into a single reading [85]. These sensors can also experience overload under heavy pollutant exposure, leading to 'burn-in' that demands lengthy recovery. Furthermore, MOS sensors may suffer from surface poisoning when contaminants are adsorbed onto the metal oxide, irreversibly changing its conductivity. In addition, both extreme temperatures and high humidity impact their performance [86,87]. Also, when the surfaces of these sensors become saturated, their ability to absorb pollutants is hindered [40,88]. Finally, hysteresis effects can arise as these sensors transition between high and low pollutant concentrations, impairing consistency in the readings [89].

2.3. Measurement Corrections

It is essential to address the aforementioned limitations to ensure that cost-effective devices can fulfil their intended functions in real-world usage conditions and achieve the desired impact on air pollution monitoring. Among different approaches compared, advanced machine learning techniques like support vector regression and deep learning based methods were found to deliver better results [58,90–96]. Recent advances in software calibration, sensor miniaturisation, and machine learning have significantly narrowed the performance gap between low-cost sensors and their high-end counterparts [27]. Using sophisticated algorithms [33], particularly those powered by artificial intelligence, their readings can be adjusted for temperature, humidity, and cross-sensitivity to other pollutants [97]. Moreover, the integration of additional data sources, such as meteorological variables, have been

shown to enhance further sensor correction [27–29,98,99]. The continuous advancement in data analytics ensures that sensors become increasingly 'smarter' over time, adapting to new conditions and providing reliable trend insights.

Although initial models explored for correcting low-cost sensor measurements relied on the simple, but interpretable, linear regression, more advanced models, have generally been more reliable [95]. Among the many non-linear machine learning approaches, ensemble learning methods that use decision trees have proved particularly popular due to their ability to capture complex non-linear relationships in sensor responses, allowing them to deliver strong performance in addressing cross-sensitivity and environmental interference. Among them, gradient boosting regression trees (GBRT) have been widely used to model non-linear relationships between sensor readings and meteorological factors. Compared to linear regression, GBRT was shown to improve the coefficient of determination (R^2) from 0.36–0.51 to 0.68–0.76 for aerosol monitoring [98]. Another tree-based method, Random Forest (RF), has achieved significant popularity. Indeed, a recent study reported that RF-based corrections improved R^2 values for both particle (PM_{2.5}, and PM₁₀), and gaseous pollutants (SO₂, NO₂, CO, and O₃) to a range of 0.70–0.99, with RMSE values between 4.05 and 17.79 $\mu\text{g}/\text{m}^3$ for the gases [27]. Similarly, usage of a multi-stage approach that incorporates RF, baseline drift correction, and empirical filters, improved sensor accuracy to ± 2.6 ppb for NO₂, ± 4.4 $\mu\text{g}/\text{m}^3$ for PM₁₀, and ± 2.7 $\mu\text{g}/\text{m}^3$ for PM_{2.5} [28]. Additionally, RF models have been found to perform well over extended time periods, showing stability for up to 16 weeks, whereas more traditional models deteriorate more rapidly [99].

Another category of machine learning models, i.e., artificial neural networks (ANNs), including deep learning models, has been extensively studied to enhance measurements due to their effectiveness in handling non-linearities in sensor data [99–102]. Whereas dynamic neural network models were shown to achieve significant error reductions (Mean Absolute Error - MAE) to less than 2 ppb for NO₂, performance for O₃ proved disappointing [99]. Similarly, ANN-based models proved efficient for certain gases, i.e., O₃ and CO₂, but not for others, i.e., NO₂ [27]. More complex deep learning models, including hybrid ones, have also been designed to improve the precision of pollutant prediction. For example, a combination of convolutional neural networks with long- and short-term memory networks led to O₃ accuracy to increase to 3.58% [27,96]. Despite the performance of these models, they are often much less interpretable than others, making them less desirable for regulatory applications in particular [60].

The absence of standard datasets, the diversity of sensor technologies and models, the regular introduction of new devices, the significant variability of real-life conditions, and the lack of standardised evaluation protocols and metrics render comparisons between machine learning approaches impractical. Moreover, the evaluation of a large range of models on the same dataset has revealed that the optimal method depends on the type of sensor and pollutant of interest. Indeed, besides RF and artificial neural networks, other models, such as generalised additive models and support vector regression, were also found to be optimal under certain conditions [98].

In conclusion, selecting an appropriate methodology must be tailored to the specific sensor, intended usage, potential needs for real-time and on-device processing, and explainability requirements.

3. Dataset: Co-located Cost-Effective Device and Reference Station Measurements

3.1. Data Collection

This study evaluates the performance of cost-effective gas sensors and their enhancement using machine learning. Its focus is on measurements of carbon monoxide (CO), ozone (O₃), and carbon dioxide (CO₂), that are captured by two electrochemical sensors and one non dispersive infrared sensor, respectively. Whereas the NDIR sensors return CO₂ concentration in Parts Per Million (PPM), the electrochemical sensors respond to the gas being measured by either oxidising or reducing it generating a very small positive or negative current, i.e., tens of nanoamperes, proportional to the presence of the gas. To accurately measure these reactions, as recommended by the sensor manufacturers, these currents are first converted into voltages using a transimpedance amplifier, then these voltages are

amplified within a range of 0 to 5 volts, and finally they are digitised with a 16-bit resolution analogue to digital converter. The CO, O₃, and CO₂ sensors, along with temperature, humidity and atmospheric pressure sensors, were mounted on a printed circuit board, designed by Technocomm Consulting Limited. This setup was housed in a plastic box (20x9x6 cm), creating a compact and cost-effective air pollution monitoring device, EnviroSense™.

To compare their measurements with some 'ground-truth', these sensors are co-located with these of a reference air quality laboratory, i.e., the Weybourne Atmospheric Observatory (WAO). Established in 1992 and operated by the University of East Anglia, the WAO is a regional station of the Global Atmospheric Watch programme of the World Meteorological Organization. Located on the North Norfolk Coast, UK (52°57'02"N, 1°07'19"E, 15 m above sea level), it encounters with a broad spectrum of pollution levels, primarily due to south-westerly winds that carry polluted air from various parts of the UK, including London and the Midlands [103]. In particular, since 2008, it has been collecting every minute high-precision long-term in situ measurements of atmospheric carbon dioxide, oxygen, carbon monoxide and molecular hydrogen.

In order to be able to study sensor consistency, two sets of the cost-effective devices are co-located with this high-precision equipment, see Figure 1. The characteristics of all the gas sensors used in this study are summarised in Table 1.



Figure 1. View of the co-located equipment at WAO: cost-effective devices 5158 and 5178 are the two blue and grey boxes

Table 1. Specifications of the sensors used in this study. Note that PPB and PPM are Parts Per Billion and Million, respectively.

Sensor Type	Weybourne Atmospheric Lab.	Cost-Effective Solution
CO	Model	Ecotech Spectronus
	Technology	Honeywell AQ7CO
	Precision	Fourier Transform Infrared Spectrometer
	Unit cost	1 PPB
O ₃	Model	unknown
	Technology	> \$100,000 (it measures both CO and CO ₂)
	Precision	< \$150
	Unit cost	Thermo 49i Ozone Analyzer
CO ₂	Model	Honeywell AQ7OZ
	Technology	UV Absorption
	Precision	1 PPB
	Unit cost	> \$3,000
Others	Model	Ecotech Spectronus
	Technology	Sensirion SCD30
	Precision	Fourier Transform Infrared Spectrometer
	Unit cost	> \$100,000 (it measures both CO and CO ₂)
Others	Temperature (°C), relative humidity (%) and pressure (hPa)	Non Dispersive Infrared
		30,000 PPB (30 PPM)
		< \$50

These three sets of data measurements are referred to in this manuscript as from 'WAO', '5158' and '5178' (as they are the series numbers of the two cost-effective devices) data. Such data were retrieved continuously for 12 weeks from 20th May 2024 to 11th August 2024. Values were averaged over 30 minutes for every variable in order to have one observation every 30 minutes as a compromise between time resolution and measurement noise, leading to up to 4032 observations per individual sensor. Although the data from devices 5158 and 5178 are complete, the gas concentrations measured

by the WAO sensors have up to 10% missing values, which was due to various sensor calibration processes having taken place during the period of data collection. Table 2 shows a brief description of the collected dataset where the minimum, maximum, mean, standard deviation and number of missing values are provided for each individual sensor. It is important to highlight significant differences in terms of temperature, relative humidity, and pressure measurements between those reported by the WAO and the cost-effective devices (i.e., 5158 and 5178). They arise because the cost-effective sensors collected data within the boxes that house them, while the WAO sensor recorded outdoor measurements. Note that these data were collected inside the boxes because this is where the gas sensors, which they affect, are located and measure gas concentrations. Finally, Figure 2 shows CO, O₃ and CO₂ concentrations measured by the reference sensors from the Weybourne Atmospheric Observatory for the 12-week duration of this study.

Table 2. Description of the data collected by each sensor

Sensor	Unit	Min	Max	Mean	Std. Dev.	Missing Values
Temperature WAO	°C	7.56	28.07	14.93	3.20	2
Temperature 5158	°C	13.85	37.74	22.26	3.97	0
Temperature 5178	°C	12.99	38.07	21.69	4.08	0
Relative Humidity WAO	%	32.31	100.00	77.29	13.47	2
Relative Humidity 5158	%	27.22	68.41	52.03	9.04	0
Relative Humidity 5178	%	26.21	70.21	54.03	9.77	0
Pressure WAO	hPa	990.52	1023.53	1008.82	5.99	2
Pressure 5158	kPa	99.22	102.53	101.06	0.60	0
Pressure 5178	kPa	99.23	102.54	101.07	0.60	0
CO WAO	ppb	84.25	294.13	110.80	15.76	318
CO 5158	µV	1380450	2382109	2232868	66708.54	0
CO 5178	µV	1177112	2280425	2129110	85833.71	0
O ₃ WAO	ppb	7.72	65.69	28.83	8.34	395
O ₃ 5158	µV	2197081	2552181	2343918	49237.40	0
O ₃ 5178	µV	2164612	2589412	2341934	52920.40	0
CO ₂ WAO	ppm	406.88	468.03	423.47	9.22	330
CO ₂ 5158	ppm	94.58	180.73	121.17	11.52	0
CO ₂ 5178	ppm	217.99	329.56	256.31	15.89	0

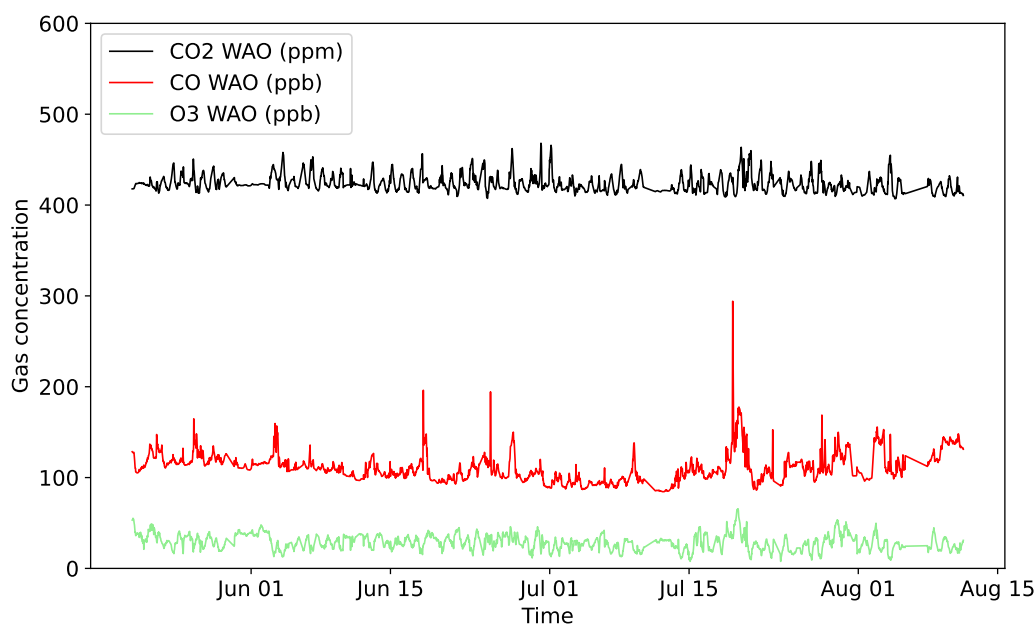


Figure 2. CO, O3 and CO2 measurements by WAO for the whole period (20/05/24 to 11/08/24)

3.2. Sensor Calibration

Calibrating low-cost sensors, especially electrochemical ones, is crucial to ensure accurate measurements. Typically, two-point calibration is used for each sensor: this involves calibrating the sensor at two known concentrations of the target gas, i.e., a zero point using pure air and a span point using a known concentration of the target gas. This process may be repeated a number of times to increase accuracy. As the way this process is performed and the quality of the reference gases affect calibration, in this study, an alternative optimised process is employed: linear regression is used to find the best fit between 11 weeks of data collected by a cost-effective gas sensor and its corresponding WAO sensor. Note that the same 11 weeks are used to train AI-based correcting models, while the remaining week is exploited to assess their performances. As Table 4 shows, correlations between WAO and calibrated cost-effective sensor readings vary between 72% and 84%, highlighting some limitations of the cost-effective solutions.

The outcome of the calibration process for the CO sensors is illustrated on Figure 3. It reveals that the measurements of the cost-effective CO sensors, shown in red and green, generally behave similarly to those of the reference WAO CO sensor. Still, it can be noticed that the readings of the cost-effective sensors are noisier, and often do not agree with each others.

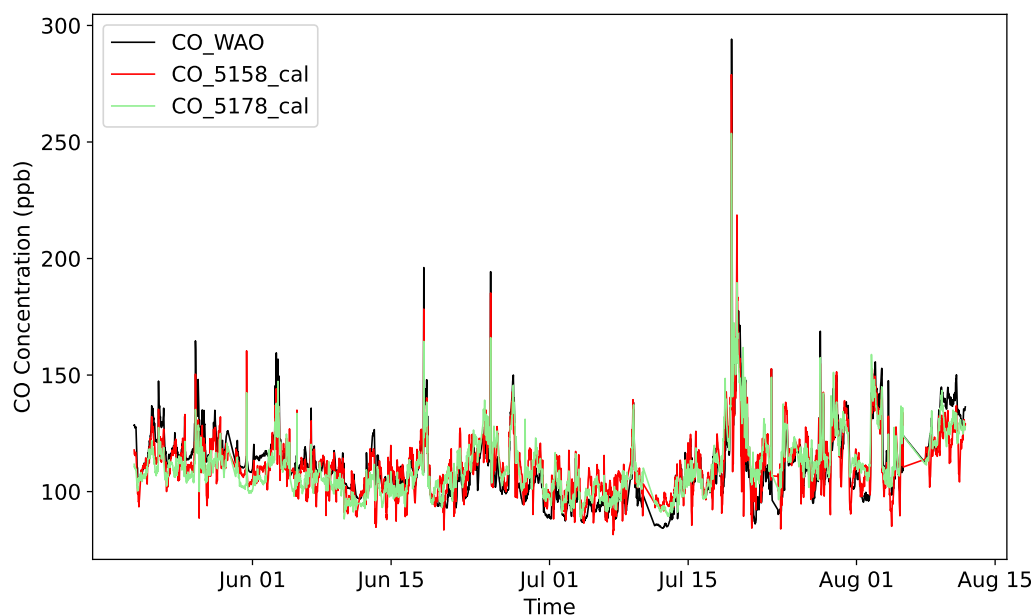


Figure 3. CO measurements after calibration of electrochemical sensors using linear regression

3.3. Data Analysis

To visualise the relationship between the measurements of the cost-effective gas sensors and those of the reference WAO CO sensor, a scatter plot is produced for each of the gases of interest, see Figure 4 for sensor 5158. For all three gases, a linear relationship is highlighted even if there is some scattering. Note that a similar outcome was observed for sensor 5178.

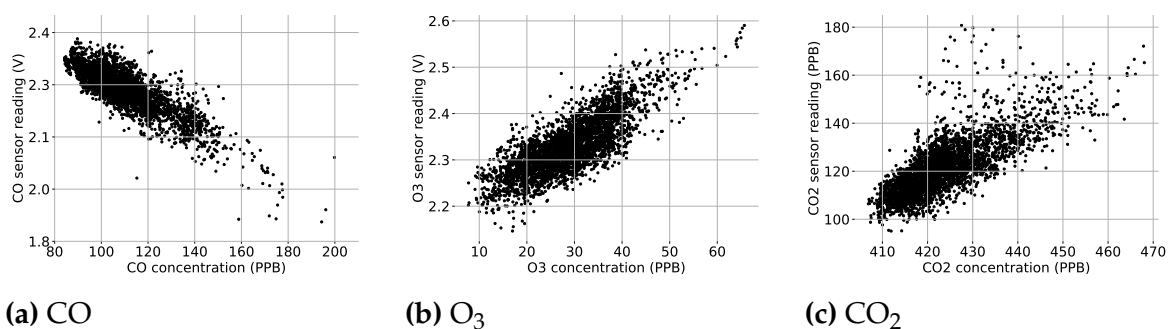


Figure 4. Scatter plots of gas measurements by sensor 5158 against those by the WAO sensor

To investigate this scattering, correlations between gas measurements made by each WAO sensor and all values return by one of the cost-effective devices are calculated. Figure 5 displays these correlation values for device 5158.

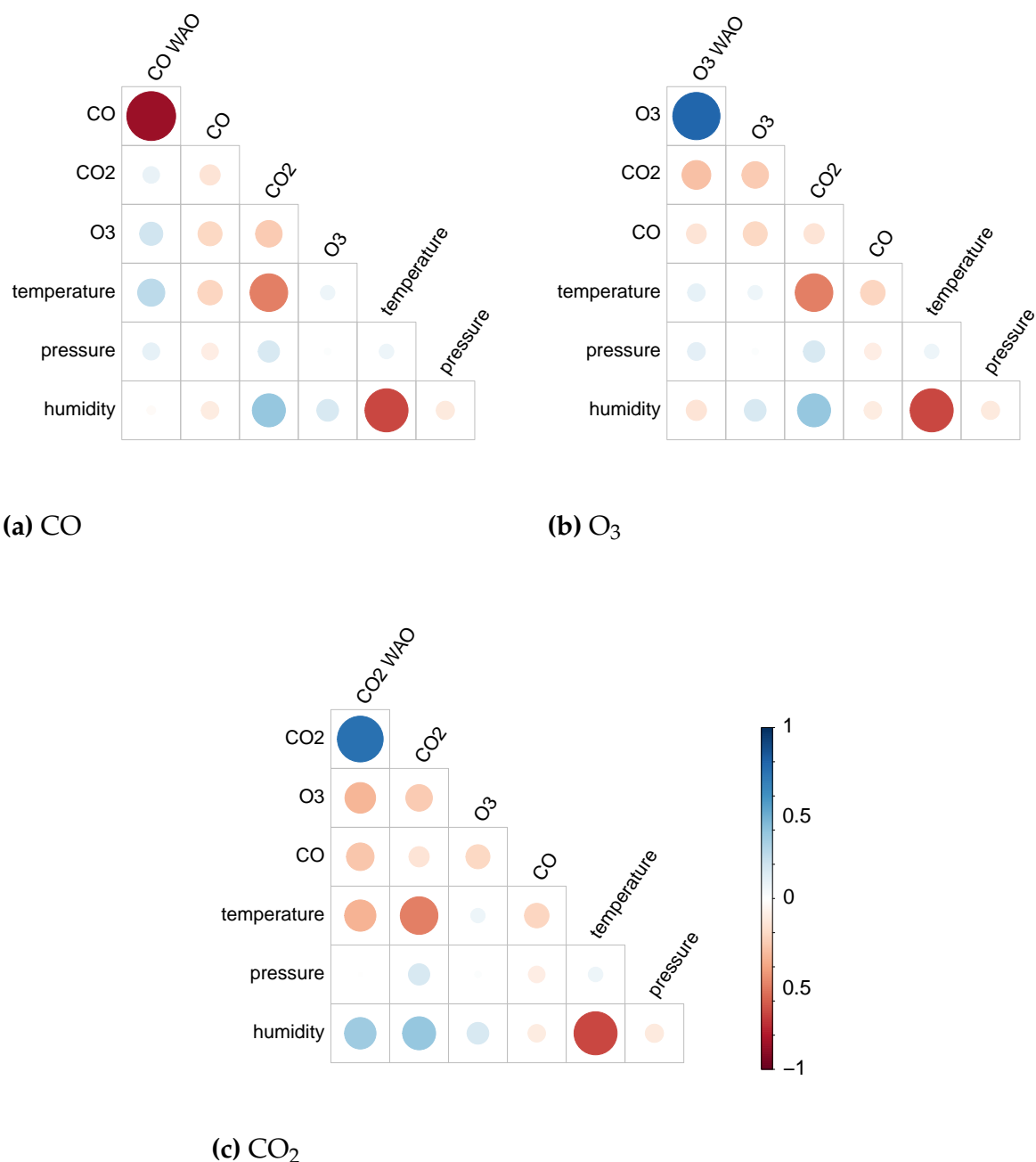


Figure 5. Correlation plots of measurements made by device 5158 with respect to gas readings from the WAO gas sensors

Figure 5a shows that the concentrations of CO measured by the cost-effective device are, as expected, highly correlated to the measurements of CO by the reference sensor. However, there are also important correlations with O₃ and temperature. This is consistent with the fact that electrochemical sensors are known to display gas cross-sensitivities [27,70,71]. More evidence of gas measurements potentially affected by the presence of other gases and meteorological conditions is provided in Figure 5b,c, where O₃ seems to be influenced by CO₂ and CO, and CO₂ readings are highly correlated to temperature, humidity, and O₃.

Finally, Figure 6 highlights slightly different behaviours between sensor 5158 and 5178 towards meteorological variables in particular. Indeed, whereas CO shows high correlation with temperature for sensor 5178, it is more moderate for sensor 5158.

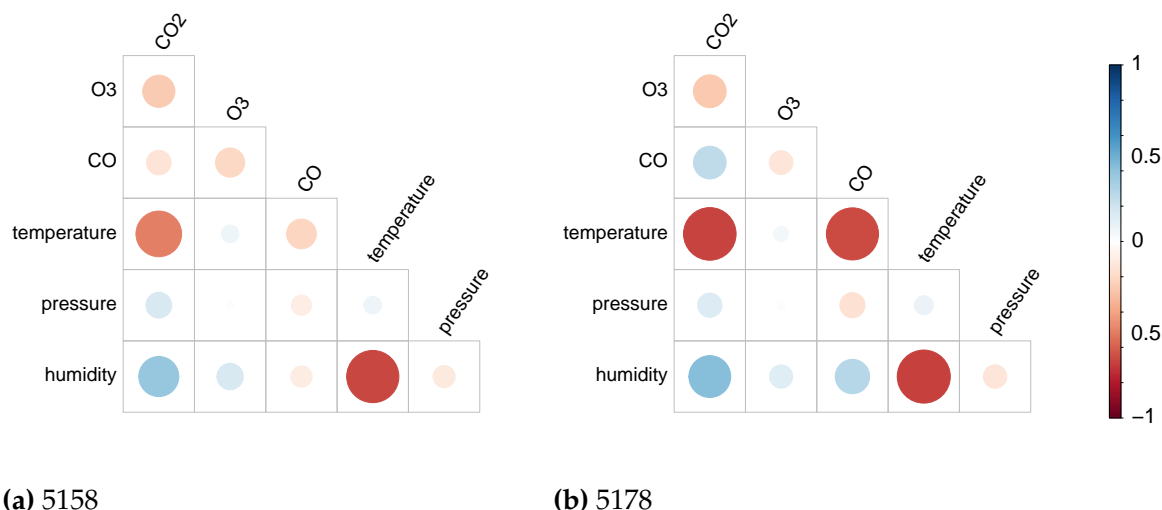


Figure 6. Correlation plots of measurements made by devices 5158 and 5178

4. Methodology to Enhance Measurement Accuracy

Since the low-cost sensors show accuracy limitations and the previous section suggests that they could be at least partially explained by cross gas sensitivity and meteorological conditions, it is proposed to design a methodology to learn the relationships between gas and meteorological measurements to correct these sensor readings. As each of the two cost-effective device hosts three gas sensors, i.e., CO, O₃ and CO₂, one can build a model for either each individual sensor, i.e., 6 models, or each sensor type, i.e., 3 models. Herein, both types of models are investigated following a similar methodology. After introducing the model selected in this study to correct data generated by the the low-cost sensors, the process of variable selection is described.

4.1. Model Choice

First, as in most practical case studies, it is expected that two-point calibration will be used to calibrate the low-cost sensors and no reference data will be available, models designed to rely on time series will not be suitable as they increasingly diverge. Second, as the processing power of cost-effective devices is likely to be provided by a micro-controller, model complexity must be relatively low. Third, models that provide a certain level of explainability, as opposed to the so-called 'black box' models, enable more informed decision-making and may reduce scepticism among some audiences. Finally, the linear relationships between measured and ground truth values revealed by Figure 4 suggests that usage of some linear regression model may be sufficient to enhance performance significantly. For these reasons, it is proposed to train a Multiple Linear Regression model (MLR) to correct the measurements performed by the cost-effective sensors. Moreover, to assess the trade-off made between model's complexity and its performance, results will be benchmarked using Support Vector Regression (SVR).

The MLR model for a given gas is defined as the following if one assumes p features:

$$Y = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p + \epsilon$$

With:

- Y : the reference observation;
- β_0 : the intercept;
- $(\beta_1, \dots, \beta_p)$: the coefficients of the model for each feature;
- x_k : the observation for the feature k , $k = 1, \dots, p$;
- ϵ : the residual term.

Whereas, for the sensor specific models, training and testing only involve data collected by the individual sensor of interest, for the models designed for a type of sensor, training and testing data combine the calibrated measurements from the two corresponding low-cost sensors from devices 5158 and 5178.

4.2. Feature Selection

As each low-cost device retrieves readings from temperature, humidity, pressure, CO, O3 and CO₂ sensors, 6 features are available for each model aiming to enhance gas measurements. As both the number of features and model complexity are quite low, a brute force approach was selected to determine the most suitable feature combination (among $2^6 - 1$, i.e., 63) for each model.

The Root Mean Squared Error (RMSE) was chosen as metric to assess performance of each generated model:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}}$$

where n is the number of observations, Y_i is the reference measurement and \hat{Y}_i is the enhanced measurement, i.e., $\hat{Y}_i = Y_i - \epsilon_i$.

The RMSE was computed using cross-validation (11 weeks for training and 1 week for testing) and the best combination was defined as the one that minimised the RMSE across all folds.

The outcome of this process is shown in Table 3, where selected features are highlighted for the three gas models of device 5158 and 5178, and the three combined gas models. For example, the best feature combination to correct the CO measurements by device 5158 are CO and humidity. One can highlight that apart from the CO₂ sensors, the two devices need different features for the best models to be delivered. In addition, humidity is the unique feature that enhances performance of all models. Regarding the three combined models, experiments concluded in the three cases that the best feature combination was usage of all the features. Note that a similar process was followed to determine the the best feature combination for the SVR models.

Table 3. Features selected for each model by minimising the RMSE.

	5158			5178			Combo (5158+5178)		
	CO	O3	CO2	CO	O3	CO2	CO	O3	CO2
CO	✓	✓	✓	✓		✓	✓	✓	✓
O3		✓	✓	✓	✓	✓	✓	✓	✓
CO2		✓	✓		✓	✓	✓	✓	✓
Temperature		✓		✓			✓	✓	✓
Humidity	✓	✓	✓	✓	✓	✓	✓	✓	✓
Pressure		✓	✓			✓	✓	✓	✓

To identify how important are the individual features to the measurement correction models, each feature was standardised before computing the MLR coefficients of each model:

$$Z = \frac{X - \bar{X}}{\sigma(X)}$$

where:

- Z : the standardised value of the feature;
- X : the initial value of the feature;
- \bar{X} : the mean value of the feature;
- $\sigma(X)$: the standard deviation of the feature.

Figures 7, 8 and 9 display, for the gas model of each individual sensor, the percentage of the absolute value of the MLR coefficient for each of the 6 features. Overall, as expected, the most important

feature for each model is the gas measurement associated to the model with values varying between 42% and 76%. Then, humidity and/or temperature are critical to the CO and O3 models, whereas CO and either O3 or humidity contribute to more than 13% of the corrections of the CO2 models. As already suggested by the different feature correlations displayed on Figure 6, the corresponding gas sensors of the two devices can show significantly different feature contributions.

Figure 10 shows similar information for the three combined gas models, where the importance of each feature is consistent with those displayed for the individual models.

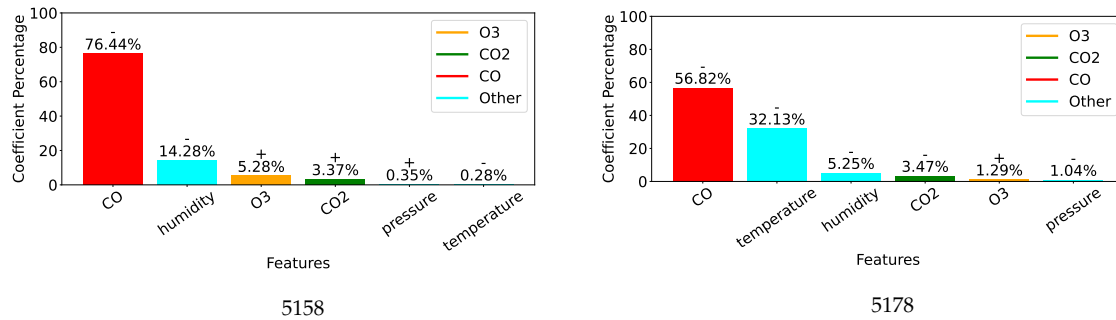


Figure 7. Coefficients for individual CO models

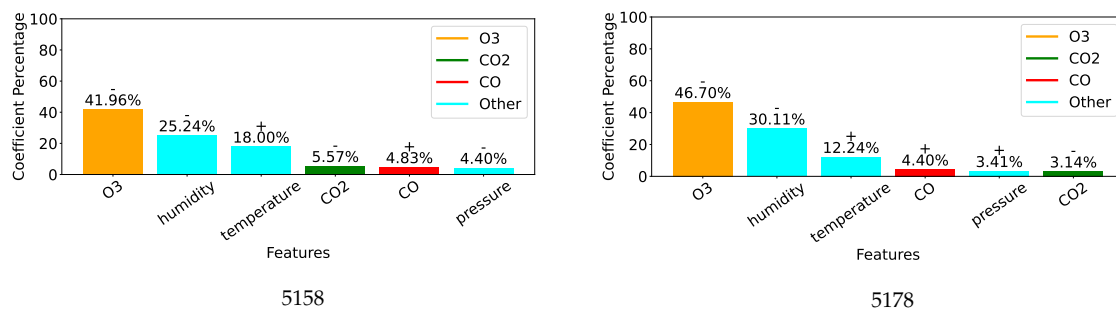


Figure 8. Coefficients for individual O3 models

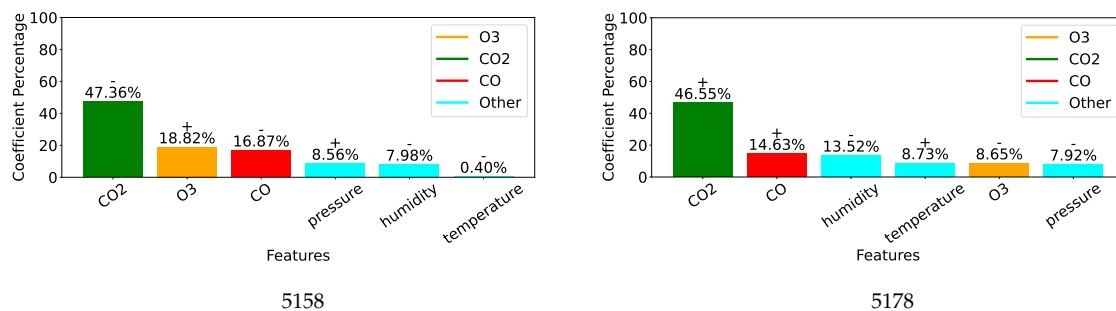


Figure 9. Coefficients for individual CO2 models

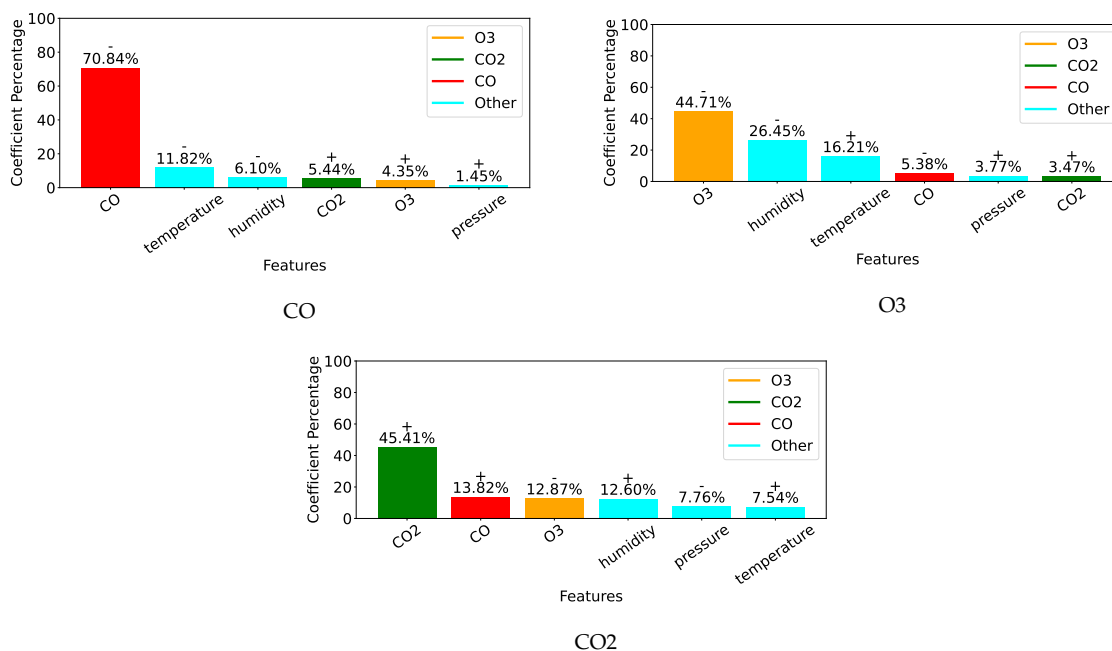


Figure 10. Coefficients for the combined models

5. Results

All experiments were conducted using cross-validation with an 11-week of training and 1-week of testing dataset. Reported results comprised measurement calibration and enhancement using MLR with both all features and best features (where available) and using SVR with best features. Note that for each SVR model, choice of regularization parameter (or cost), margin, and kernel (and their associated parameters) was optimised. Using WAO measurements as ground truth, performance was evaluated using three main metrics, i.e., Pearson's correlation coefficient (correlation), Mean Percentage Error (MPE), and Mean Average Error (MAE), including the associated Standard Deviation (STD), and Min and Max errors.

Two sets of experiments were conducted to evaluate the generated models: first, with models designed for each individual sensors, and second, with combined models associated with individual gases.

5.1. Models for individual sensors

Table 4 reports results obtained from models tailored to each individual sensor. Correlations between WAO and calibrated cost-effective sensor readings vary between 72% and 84%, which is usually considered as strong correlations. In addition, there are important correlation differences both between types of gas sensor, with CO sensors performing best, and within a type of gas sensor, especially O3.

In terms of enhanced measurements, all models using 'Best features' outperformed the calibration alone. For example, correlations increased to the range 81% and 91%, and MAE decreased by up to 46%. This confirms the added value of using additional features to that of the gas of interest. Although feature selection improves results overall, gains are generally quite limited. In addition, SVR models outperformed MLR models offering systematically the best MAE and reduced both standard deviation and max error.

Table 4. Evaluation of calibrated and enhanced measurements for individual sensors against ground truth

		Correlation	MPE	MAE	STD	Min	Max
CO 58 (ppb)	MLR (all features)	0.84	4.91 %	7.03	4.83	0.03	29.44
	MLR (best features)	0.84	4.66 %	6.86	4.76	0.05	30.10
	SVR (best features)	0.83	4.70 %	6.13	4.02	0.06	24.65
	Calibration	0.82	4.51 %	6.93	4.96	0.04	29.90
CO 78 (ppb)	MLR (all features)	0.90	2.95 %	4.80	3.76	0.02	29.76
	MLR (best features)	0.90	2.97 %	4.78	3.68	0.03	28.84
	SVR (best features)	0.91	3.07 %	4.17	2.89	0.04	20.43
	Calibration	0.84	5.71 %	7.66	4.80	0.06	26.87
O3 58 (ppb)	MLR (all features)	0.89	8.14 %	3.24	2.40	0.02	12.81
	MLR (best features)	0.89	8.14 %	3.24	2.40	0.02	12.81
	SVR (best features)	0.91	4.45 %	2.51	1.86	0.01	9.79
	Calibration	0.80	11.81 %	4.18	3.05	0.01	14.56
O3 78 (ppb)	MLR (all features)	0.82	13.44 %	4.93	3.15	0.09	18.83
	MLR (best features)	0.82	13.02 %	4.75	3.07	0.06	18.40
	SVR (best features)	0.85	8.40 %	3.54	2.48	0.03	14.14
	Calibration	0.72	15.10 %	5.19	3.70	0.03	18.57
CO2 58 (ppm)	MLR (all features)	0.80	0.33 %	4.18	3.57	0.02	18.85
	MLR (best features)	0.81	0.32 %	4.14	3.56	0.02	18.88
	SVR (best features)	0.82	0.38 %	3.93	3.24	0.02	16.50
	Calibration	0.74	0.33 %	4.70	3.78	0.01	20.35
CO2 78 (ppm)	MLR (all features)	0.85	0.42 %	3.84	3.39	0.03	19.35
	MLR (best features)	0.86	0.42 %	3.82	3.40	0.02	19.11
	SVR (best features)	0.87	0.41 %	3.56	3.13	0.03	16.05
	Calibration	0.79	0.51 %	4.60	3.72	0.02	19.38

5.2. Models for Combined Sensors

Results obtained by the combined models are displayed in Table 5. In terms of calibration performance, correlations are around the mean of the values obtained by the two corresponding sensors. As previously, both MLR and SVR enhance measurement quality for O3 and CO2, with SVR doing better than MLR. However, CO results are not improved as the various metrics show. This may be explained by the discrepancies between CO sensors 5158 and 5178 as shown in Figure 7. Indeed, it suggests very different behaviours especially towards temperature.

Comparison with performance reported in Table 4 clearly shows that models targeting single sensors better enhance measurements than the more general ones designed for a sensor type. Still, one can observe that for O3 and CO2 sensors, the combined models provide some added value when compared to performance delivered by individual sensor calibration.

Table 5. Evaluation of calibrated and enhanced measurements for combined sensors against ground truth

		Correlation	MPE	MAE	STD	Min	Max
CO (ppb)	MLR (all features)	0.81	4.27 %	6.83	4.99	0.03	35.85
	SVR (all features)	0.81	4.48 %	7.07	5.34	0.03	43.21
	Calibration	0.82	4.57 %	6.85	4.88	0.02	31.43
O3 (ppb)	MLR (all features)	0.85	10.79 %	4.05	2.89	0.01	18.74
	SVR (all features)	0.85	9.67 %	3.89	2.85	0.01	18.01
	Calibration	0.76	12.33 %	4.51	3.35	0.01	18.39
CO2 (ppm)	MLR (all features)	0.79	0.29 %	4.19	3.62	0.01	20.38
	SVR (all features)	0.80	0.32 %	4.01	3.60	0.01	20.52
	Calibration	0.75	0.32 %	4.56	3.73	0.01	20.85

6. Discussion

This study has provided valuable insights into the limitations of three types of low-cost sensors: two electrochemical sensors for CO and O3, and a non-dispersive infrared sensor for CO2. In addition, it has evaluated how usage of multiple linear regression models can be exploited to correct sensor readings.

Given the 12-week duration of data collection, this work was unable to assess the impact of sensor ageing. However, it provides evidence supporting several other causes of measurement inaccuracies described in sub-section 2.3. Although sensor manufacturers only provide correction curves for temperature, Figures 5 and 6 reveal significant correlations between individual gas measurements and not only temperature but also humidity and other gases. This is particularly evident for the non-dispersive infrared sensors measuring CO₂ and the electrochemical sensors reporting CO concentration. Inconsistent behaviour between two units of the same sensor is also noticeable, especially between the two CO sensors, which were affected very differently by temperature and humidity. This is clearly highlighted by the very different coefficient values of their individual models designed to correct their readings, see Figure 7.

For models trained on each individual sensor, the use of MLR models often significantly enhances performance. One should note that feature optimisation has shown very limited impact. Conversely, results using SVR suggest that employing more sophisticated models could further improve measurements if the constraint of processing data on the device was removed. Nonetheless, additional gains might be limited, as previous studies highlight cross-sensitivity with other gases, including Cl₂, NO₂, and H₂S [27,70,71], and the value of integrating meteorological data [27–29,98,99]. Thus, a promising strategy to enhance gas readings could involve feeding the models with more relevant features that they could exploit: air pollution monitoring devices could be equipped with additional sensors to detect other types of gases, and/or they could retrieve external meteorological information.

While ideally, a single correction model for each type of gas sensor would be preferred, the performance reported in Table 5 and cross-sensitivity of individual units highlighted in Figures 7–9 indicate that similarity between units' behaviour is crucial for the success of such models. In this study, the combined model proved very successful for the two O₃ sensors, which share very similar sensitivity profiles, see Figure 8. However, the model failed to deliver enhanced measurements for the two CO sensors, which exhibit very different sensitivity to other parameters, see Figure 7. Given that it may be unrealistic to expect significant improvements in the manufacturing consistency of low-cost sensors in the short term, a practical approach could be to design a small range of models for each sensor type and use factory calibration to associate each unit with one of these models.

7. Conclusions

The primary objective of this study was to assess how machine learning algorithms could improve readings from low-cost sensors, ensuring that these corrections could be performed on the monitoring device, i.e., on a simple microcontroller. To meet these requirements multiple linear regression models were trained and tested using data collected from two electrochemical sensors measuring CO and O₃ concentrations, as well as a non dispersive infrared sensor estimating CO₂ levels. These sensors were co-located with an air quality monitoring reference station to ensure validation against high-precision measurements.

Models designed for a specific sensor unit significantly enhanced measurements increasing their correlations with those from reference-grade instruments by up to 10%. Further improvements in accuracy are anticipated by incorporating additional features into the models, such as meteorological data and extra gas readings. Naturally, if the correction process could be performed online, more sophisticated machine learning models could be considered. However, the study also revealed that variations in the behaviour of different sensor units can hinder the development of a single correction model for each sensor type. To address production inconsistencies, if a factory-based calibration process was able to classify units according to sub-types, specific correction models could be developed for each of them.

While challenges persist, there is confidence that they will soon be overcome, enabling cost-effective devices to provide dense air quality monitoring networks. Such advancements would contribute significantly to several United Nations Sustainable Development Goals, promoting broader access to reliable environmental monitoring.

Author Contributions: Conceptualization, B.M., J.N. and F.R.; methodology, Y.C., C.W., J.N. and F.R.; software, Y.C. and C.W.; validation, Y.C. and C.W.; resources, B.M.; writing—original draft preparation, Y.C. and C.W.; writing—review and editing, Y.C., C.W., B.M., J.N. and F.R.; visualization, Y.C.; supervision, J.N. and F.R.; funding acquisition, B.M., J.N. and F.R.. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Innovate UK – Accelerated Knowledge Transfer - grant number 1726-AKT (2024), and UK government - UKSPF BIG Growth Programme - grant number P2687-100 (2024).

Data Availability Statement: Dataset available on request from the authors.

Acknowledgments: The authors would like to thank Dr Grant Forster, Senior Research Fellow in the School of Environmental Sciences at the University of East Anglia, UK, and coordinator of the Weybourne Atmospheric Observatory, for accepting to host their cost-effective sensors and giving access to the data collected by the laboratory.

Conflicts of Interest: The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Burnett, R. et al. (2018) 'Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter', *Proceedings of the National Academy of Sciences*, 115(38), pp. 9592–9597. doi:10.1073/pnas.1803222115.
2. 2022. Ambient (outdoor) air pollution, World Health Organization. [Accessed on 31 January 2025]
3. Forouzanfar, M.H. et al. (2016) 'Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990–2015: A systematic analysis for the global burden of disease study 2015', *The Lancet*, 388(10053), pp. 1659–1724. doi:10.1016/s0140-6736(16)31679-8.
4. Morawska, L. et al. (2018) 'Applications of low-cost sensing technologies for air quality monitoring and exposure assessment: How far have they gone?', *Environment International*, 116, pp. 286–299. doi:10.1016/j.envint.2018.04.018.
5. Fuller, R. et al. (2022) 'Pollution and health: A progress update', *The Lancet Planetary Health*, 6(6). doi:10.1016/s2542-5196(22)00090-0.
6. Noble, C.A. et al. (2001) 'Federal reference and equivalent methods for measuring fine particulate matter', *Aerosol Science and Technology*, 34(5), pp. 457–464. doi:10.1080/02786820121582.
7. Zhang, Xin, Chen, X. and Zhang, Xiaobo (2018) 'The impact of exposure to air pollution on cognitive performance', *Proceedings of the National Academy of Sciences*, 115(37), pp. 9193–9197. doi:10.1073/pnas.1809474115.
8. Manisalidis, I. et al. (2020) 'Environmental and health impacts of Air Pollution: A Review', *Frontiers in Public Health*, 8. doi:10.3389/fpubh.2020.00014.
9. Martenies, S.E., Wilkins, D. and Batterman, S.A. (2015a) 'Health impact metrics for Air Pollution Management Strategies', *Environment International*, 85, pp. 84–95. doi:10.1016/j.envint.2015.08.013.
10. Castell, N., Dauge, F. R., Schneider, P., Vogt, M., Lerner, U., Fishbain, B., Broday, D., & Bartonova, A. (2017). Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates? *Environment International* Vol. 99, pp. 293–302
11. Hegde, M., Nebel, J.-C., & Rahman, F. (2024). Cleaning up the Big Smoke: Forecasting London's Air Pollution Levels Using Energy-Efficient AI. *International Journal of Environmental Pollution and Remediation* Vol. 12, not yet published
12. Yang, B.-Y. et al. (2020) 'Ambient air pollution and diabetes: A systematic review and meta-analysis', *Environmental Research*, 180, p. 108817. doi:10.1016/j.envres.2019.108817.
13. Fuller, R. et al. (2022) 'Pollution and health: A progress update', *The Lancet Planetary Health*, 6(6). doi:10.1016/s2542-5196(22)00090-0.
14. www.epa.gov/indoor-air-quality-iaq/low-cost-air-pollution-monitors-and-indoor-air-quality
15. Snyder, E.G. et al. (2013) 'The changing paradigm of Air Pollution Monitoring', *Environmental Science & Technology*, 47(20), pp. 11369–11377. doi:10.1021/es4022602.
16. Taheri Shahraiyani, H. et al. (2015) 'The development of a dense urban air pollution monitoring network', *Atmospheric Pollution Research*, 6(5), pp. 904–915. doi:10.5094/apr.2015.100.

17. US EPA, O. (2016) EPA Scientists Develop and Evaluate Federal Reference & Equivalent Methods for Measuring Key Air Pollutants. <https://www.epa.gov/air-research/epa-scientists-develop-and-evaluate-federal-reference-equivalent-methods-measuring-key>
18. Hoekman, S.K. and Welstand, J.S. (2021) 'Vehicle emissions and air quality: The early years (1940s–1950s)', *Atmosphere*, 12(10), p. 1354. doi:10.3390/atmos12101354.
19. Apte, J.S. and Manchanda, C. (2024) 'High-resolution urban air pollution mapping', *Science*, 385(6707), pp. 380–385. doi:10.1126/science.adq3678.
20. Ródenas García, M. et al. (2022) 'Review of low-cost sensors for indoor air quality: Features and applications', *Applied Spectroscopy Reviews*, 57(9–10), pp. 747–779. doi:10.1080/05704928.2022.2085734.
21. Shairsingh, K. et al. (2023) 'Who air quality database: Relevance, history and future developments', *Bulletin of the World Health Organization*, 101(12), pp. 800–807. doi:10.2471/blt.23.290188.
22. Carotenuto, F. et al. (2023) 'Low-Cost Air Quality Monitoring Networks for long-term field campaigns: A Review', *Meteorological Applications*, 30(6). doi:10.1002/met.2161.
23. Bittner, A.S. et al. (2022) 'Performance characterization of low-cost air quality sensors for off-grid deployment in rural Malawi', *Atmospheric Measurement Techniques*, 15(11), pp. 3353–3376. doi:10.5194/amt-15-3353-2022.
24. Fowler, D. et al. (2020) 'A chronology of Global Air Quality', *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 378(2183), p. 20190314. doi:10.1098/rsta.2019.0314.
25. Carotenuto, F. et al. (2023) 'Low-Cost Air Quality Monitoring Networks for long-term field campaigns: A Review', *Meteorological Applications*, 30(6). doi:10.1002/met.2161.
26. Giordano, M.R. et al. (2021) 'From low-cost sensors to high-quality data: A summary of challenges and best practices for effectively calibrating low-cost Particulate Matter Mass Sensors', *Journal of Aerosol Science*, 158, p. 105833. doi:10.1016/j.jaerosci.2021.105833.
27. Wang, G. et al. (2024) 'Research of low-cost air quality monitoring models with different machine learning algorithms', *Atmospheric Measurement Techniques*, 17(1), pp. 181–196. doi:10.5194/amt-17-181-2024.
28. Bush, T. et al. (2022) 'Machine learning techniques to improve the field performance of low-cost air quality sensors', *Atmospheric Measurement Techniques*, 15(10), pp. 3261–3278. doi:10.5194/amt-15-3261-2022.
29. Ravindra, K. et al. (2024) Enhancing accuracy of air quality sensors with machine learning to augment large-scale monitoring networks, *Nature News*. Available at: <https://www.nature.com/articles/s41612-024-00833-9> (Accessed: 26 January 2025).
30. Kumar, P. et al. (2015) 'The rise of low-cost sensing for managing air pollution in cities', *Environment International*, 75, pp. 199–205. doi:10.1016/j.envint.2014.11.019.
31. Seesaard, T., Kamjornkittikoon, K. and Wongchoosuk, C. (2024) 'A comprehensive review on advancements in sensors for air pollution applications', *Science of The Total Environment*, 951, p. 175696. doi:10.1016/j.scitotenv.2024.175696.
32. Rodríguez-Trejo, A. et al. (2024) 'Air quality monitoring with low-cost sensors: A record of the increase of PM_{2.5} during Christmas and New Year's Eve celebrations in the city of Queretaro, Mexico', *Atmosphere*, 15(8), p. 879. doi:10.3390/atmos15080879.
33. Brugnone, F., Randazzo, L. and Calabrese, S. (2024) 'Use of low-cost sensors to study atmospheric particulate matter concentrations: Limitations and benefits discussed through the analysis of three case studies in Palermo, Sicily', *Sensors*, 24(20), p. 6621. doi:10.3390/s24206621.
34. Okorn, K. and Iraci, L.T. (2024) 'An overview of outdoor low-cost gas-phase air quality sensor deployments: Current efforts, trends, and limitations', *Atmospheric Measurement Techniques*, 17(21), pp. 6425–6457. doi:10.5194/amt-17-6425-2024.
35. Liu, X. et al. (2020) 'Low-cost sensors as an alternative for long-term air quality monitoring', *Environmental Research*, 185, p. 109438. doi:10.1016/j.envres.2020.109438.
36. Popoola, O.A.M. et al. (2018) 'Use of networks of low cost air quality sensors to quantify air quality in urban settings', *Atmospheric Environment*, 194, pp. 58–70. doi:10.1016/j.atmosenv.2018.09.030.
37. Low-cost sensors can improve air quality monitoring and People's Health (2024) World Meteorological Organization. Available at: <https://wmo.int/news/media-centre/low-cost-sensors-can-improve-air-quality-monitoring-and-peoples-health> (Accessed: 26 January 2025).
38. Bililign, S. et al. (2024) 'East African megacity air quality: Rationale and framework for a measurement and Modeling Program', *Bulletin of the American Meteorological Society*, 105(8). doi:10.1175/bams-d-23-0098.1.
39. Nalakarathi, N.V. et al. (2024) 'Challenges and opportunities in calibrating low-cost environmental sensors', *Sensors*, 24(11), p. 3650. doi:10.3390/s24113650.

40. Jin, Z. et al. (2024) 'Humidity-independent gas sensors in the detection of hydrogen sulfide based on ND2O3-loaded in2o3 porous nanorods', *Sensors and Actuators B: Chemical*, 403, p. 135237.
41. Khreis, H. et al. (2022) 'Evaluating the performance of low-cost air quality monitors in Dallas, Texas', *International Journal of Environmental Research and Public Health*, 19(3), p. 1647. doi:<https://www.doi.org/10.3390/ijerph19031647>.
42. Saeed, T. et al. (2024) Sustaining low-cost PM2.5 monitoring networks in South Asia: Technical challenges and solutions [Preprint]. doi:10.5194/egusphere-2024-1932.
43. 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. *Atmospheric Environment* Vol. 70, pp. 186–203
44. Kumar, P. et al. (2015) 'The rise of low-cost sensing for managing air pollution in cities', *Environment International*, 75, pp. 199–205. doi:10.1016/j.envint.2014.11.019.
45. Dharaiya, V.R. et al. (2023) 'Evaluating the performance of low-cost PM sensors over multiple coalesce network sites', *Aerosol and Air Quality Research*, 23(5), p. 220390. doi:10.4209/aaqr.220390.
46. Karagulian, F. et al. (2019) 'Review of the performance of low-cost sensors for air quality monitoring', *Atmosphere*, 10(9), p. 506. doi:10.3390/atmos10090506.
47. Bililign, S. et al. (2024) East African megacity air quality: Rationale and framework for a measurement and Modeling Program, AMETSOC. Available at: <https://journals.ametsoc.org/view/journals/bams/105/8/BAMS-D-23-0098.1.xml> (Accessed: 26 January 2025).
48. Thara Seesaard a et al. (2024) A comprehensive review on advancements in sensors for air pollution applications, *Science of The Total Environment*. Available at: <https://www.sciencedirect.com/science/article/abs/pii/S00489697240585> (Accessed: 26 January 2025).
49. Varaden, D., Leidland, E. and Barratt, B. (2019) *The Breathe London Wearables Study Engaging primary school children to monitor air pollution in London*. Greater London Authority. Available at: https://erg.ic.ac.uk/research/docs/Uploads_to_exposure_science_website/Final%20BLW%20Report_211019%20.pdf (Accessed: 26 January 2025).
50. Wenwei Che, et al. (2020). PRAISE-HK: A personalized real-time air quality informatics system for citizen participation in exposure and health risk management
51. Mahajan, S. et al. (2022) 'Translating citizen-generated air quality data into evidence for shaping policy', *Humanities and Social Sciences Communications*, 9(1). doi:10.1057/s41599-022-01135-2.
52. Lung, S.-C.C. et al. (2022) Research priorities of applying low-cost PM2.5 sensors in Southeast Asian countries, *International journal of environmental research and public health*. Available at: <https://pmc.ncbi.nlm.nih.gov/articles/PMC8835170/> (Accessed: 26 January 2025).
53. Bililign, S. et al. (2024a) East African megacity air quality: Rationale and framework for a measurement and Modeling Program, AMETSOC.
54. Higgins, C., Kumar, P. and Morawska, L. (2024) 'Indoor air quality monitoring and source apportionment using low-cost sensors', *Environmental Research Communications*, 6(1), p. 012001. doi:10.1088/2515-7620/ad1cad.
55. Liang Y et al. 2021 Field comparison of electrochemical gas sensor data correction algorithms for ambient air measurements *Sensors Actuators B* 327 128897, doi: 10.1016/j.snb.2020.128897
56. Air Quality Expert Group 2023 HM Government Department for Environment, Food and Rural Affairs
57. Coelho Rezende G, Le Calvé S, Brandner J J and Newport D 2019 Micro photoionization detectors *Sensors Actuators B* 287 86–94
58. Levy Zamora, M. et al. (2022) 'Evaluating the performance of using low-cost sensors to calibrate for cross-sensitivities in a multipollutant network', *ACS ES&T Engineering*, 2(5), pp. 780–793. doi:10.1021/acsestengg.1c00367.
59. Chai, H. et al. (2022) 'Stability of Metal Oxide Semiconductor Gas Sensors: A Review', *IEEE Sensors Journal*, 22(6), pp. 5470–5481.
60. Johnson, N. et al. (2018) 'Using a gradient boosting model to improve the performance of low-cost aerosol monitors in a dense, heterogeneous urban environment', *Atmospheric Environment*, 184, pp. 9–16. <https://doi.org/10.1016/j.atmosenv.2018.04.019>
61. Molaie, S. and Lino, P. (2021) 'Review of the newly developed, mobile optical sensors for real-time measurement of the Atmospheric Particulate Matter Concentration', *Micromachines*, 12(4), p. 416. doi:10.3390/mi12040416.

62. Lu, L. et al. (2022) 'Optical measurement method of non-spherical particle size and concentration based on high-temperature melting technique', *Measurement*, 198, p. 111375. doi:10.1016/j.measurement.2022.111375.
63. Hagan, D.H. and Kroll, J.H. (2020) 'Assessing the accuracy of low-cost optical particle sensors using a physics-based approach', *Atmospheric Measurement Techniques*, 13(11), pp. 6343–6355. doi:10.5194/amt-13-6343-2020.
64. Laref, R. et al. (2021) 'Empiric unsupervised drifts correction method of electrochemical sensors for in field nitrogen dioxide monitoring', *Sensors*, 21(11), p. 3581. doi:10.3390/s21113581.
65. Leo Hohenberger, T. et al. (2022) 'Assessment of the impact of sensor error on the representativeness of population exposure to urban air pollutants', *Environment International*, 165, p. 107329. doi:10.1016/j.envint.2022.107329.
66. Maag, B., Zhou, Z. and Thiele, L. (2018) 'A survey on sensor calibration in Air Pollution Monitoring deployments', *IEEE Internet of Things Journal*, 5(6), pp. 4857–4870. doi:10.1109/jiot.2018.2853660.
67. Wei, P., Ning, Z., Ye, S., Sun, L., Yang, F., Wong, K., Westerdahl, D., & Louie, P. (2018). Impact Analysis of Temperature and Humidity Conditions on Electrochemical Sensor Response in Ambient Air Quality Monitoring. *Sensors* Vol. 18, Issue 2, p. 59
68. Fine, G.F. et al. (2010) 'Metal oxide semi-conductor gas sensors in environmental monitoring', *Sensors*, 10(6), pp. 5469–5502.
69. Afshar-Mohajer, N. et al. (2017) 'Evaluation of low-cost electro-chemical sensors for environmental monitoring of ozone, nitrogen dioxide, and carbon monoxide', *Journal of Occupational and Environmental Hygiene*, 15(2), pp. 87–98. doi:10.1080/15459624.2017.1388918.
70. Pang, X. et al. (2018) 'The impacts of water vapour and co-pollutants on the performance of electrochemical gas sensors used for air quality monitoring', *Sensors and Actuators B: Chemical*, 266, pp. 674–684. doi:10.1016/j.snb.2018.03.144.
71. Zuidema, C. et al. (2019) 'Efficacy of paired electrochemical sensors for measuring ozone concentrations', *Journal of Occupational and Environmental Hygiene*, 16(2), pp. 179–190. doi:10.1080/15459624.2018.1540872.
72. Cross, E.S. et al. (2017) 'Use of electrochemical sensors for measurement of Air Pollution: Correcting interference response and validating measurements', *Atmospheric Measurement Techniques*, 10(9), pp. 3575–3588. doi:10.5194/amt-10-3575-2017.
73. Gherardi, S.; Astolfi, M.; Gaiardo, A.; Malagù, C.; Rispoli, G.; Vincenzi, D.; Zonta, G. Investigating the Temperature-Dependent Kinetics in Humidity-Resilient Tin–Titanium-Based Metal Oxide Gas Sensors. *Chemosensors* 2024, 12, 151. <https://doi.org/10.3390/chemosensors12080151>
74. Wang, C.; Yin, L.; Zhang, L.; Xiang, D.; Gao, R. Metal Oxide Gas Sensors: Sensitivity and Influencing Factors. *Sensors* 2010, 10, 2088–2106. <https://doi.org/10.3390/s100302088>
75. Kyungmin Kim, Jin Kuen Park, Jieon Lee, Yong Jung Kwon, Hyeunseok Choi, Seung-Min Yang, Jung-Hoon Lee, Young Kyu Jeong, Synergistic approach to simultaneously improve response and humidity-independence of metal-oxide gas sensors, *Journal of Hazardous Materials*, Volume 424, Part B, 2022, 127524, ISSN 0304-3894, <https://doi.org/10.1016/j.jhazmat.2021.127524>
76. Afshar-Mohajer, N., Zuidema, C., Sousan, S., Hallett, L., Tatum, M., Rule, A. M., ... Koehler, K. (2018). Evaluation of low-cost electro-chemical sensors for environmental monitoring of ozone, nitrogen dioxide, and carbon monoxide. *Journal of Occupational and Environmental Hygiene*, 15(2), 87–98. <https://doi.org/10.1080/15459624.2017.1388918>
77. Hayward, I.; Martin, N.A.; Ferracci, V.; Kazemimanesh, M.; Kumar, P. Low-Cost Air Quality Sensors: Biases, Corrections and Challenges in Their Comparability. *Atmosphere* 2024, 15, 1523. <https://doi.org/10.3390/atmos15121523>
78. Wastine, B.; Hummelgård, C.; Bryzgalov, M.; Rödjegård, H.; Martin, H.; Schröder, S. Compact Non-Dispersive Infrared Multi-Gas Sensing Platform for Large Scale Deployment with Sub-ppm Resolution. *Atmosphere* 2022, 13, 1789. <https://doi.org/10.3390/atmos13111789>
79. Jha, R.K. (2022) 'Non-dispersive Infrared Gas Sensing Technology: A Review', *IEEE Sensors Journal*, 22(1), pp. 6–15. doi:10.1109/jsen.2021.3130034.
80. Muller, M. et al. (2020) 'Integration and calibration of non-dispersive infrared (NDIR) CO₂ low-cost sensors and their operation in a sensor network covering Switzerland', *Atmospheric Measurement Techniques*, 13(7), pp. 3815–3834. doi:10.5194/amt-13-3815-2020.

81. Martin, C.R. et al. (2017) 'Evaluation and environmental correction of ambient CO₂; measurements from a low-cost Ndir Sensor', *Atmospheric Measurement Techniques*, 10(7), pp. 2383–2395. doi:10.5194/amt-10-2383-2017.
82. Wei Xu, Yunfei Cai, Song Gao, Shuang Hou, Yong Yang, Yusen Duan, Qingyan Fu, Fei Chen, Jie Wu, New understanding of miniaturized VOCs monitoring device: PID-type sensors performance evaluations in ambient air, *Sensors and Actuators B: Chemical*, Volume 330, 2021, 129285, ISSN 0925-4005, <https://doi.org/10.1016/j.snb.2020.129285>.
83. Spinelle, L. et al. (2017) 'Review of portable and low-cost sensors for the ambient air monitoring of benzene and other volatile organic compounds', *Sensors*, 17(7), p. 1520. doi:10.3390/s17071520.
84. Bilek, J.; Marsolek, P.; Bilek, O.; Bucek, P. Field Test of Mini Photoionization Detector-Based Sensors—Monitoring of Volatile Organic Pollutants in Ambient Air. *Environments* 2022, 9, 49. <https://doi.org/10.3390/environments9040049>
85. Schütze, A. et al. (2017) 'Highly sensitive and selective VOC sensor systems based on semiconductor gas sensors: How to?', *Environments*, 4(1), p. 20. doi:10.3390/environments4010020.
86. Nabil Abdullah, A. et al. (2020) 'Effect of environmental temperature and humidity on different metal oxide gas sensors at various gas concentration levels', *IOP Conference Series: Materials Science and Engineering*, 864(1), p. 012152. doi:10.1088/1757-899x/864/1/012152.
87. Mahdavi H, Rahbarpour S, Hosseini-Golgoo S, Jamaati H, Reducing the destructive effect of ambient humidity variations on gas detection capability of a temperature modulated gas sensor by calcium chloride, *Sensors and Actuators B: Chemical*, Volume 331, 2021, 129091, ISSN 0925-4005,
88. Isaac, N.A., Pikaar, I. & Biskos, G. Metal oxide semiconducting nanomaterials for air quality gas sensors: operating principles, performance, and synthesis techniques. *Microchim Acta* 189, 196 (2022).
89. Peterson, P. et al. (2017) 'Practical use of metal oxide semiconductor gas sensors for measuring nitrogen dioxide and ozone in urban environments', *Sensors*, 17(7), p. 1653. doi:10.3390/s17071653.
90. Nguyen, A.D. et al. (2024) 'Gamma: A universal model for calibrating sensory data of multiple low-cost air monitoring devices', *Engineering Applications of Artificial Intelligence*, 128, p. 107591. doi:10.1016/j.engappai.2023.107591.).
91. Mahajan, S. and Kumar, P. (2020) 'Evaluation of low-cost sensors for quantitative personal exposure monitoring', *Sustainable Cities and Society*, 57, p. 102076. doi:10.1016/j.scs.2020.102076.).
92. Rahi, P., Sood, S.P., Bajaj, R. et al. Air quality monitoring for Smart eHealth system using firefly optimization and support vector machine. *Int. j. inf. tecnol.* 13, 1847–1859 (2021). <https://doi.org/10.1007/s41870-021-00778-9>
93. Laref, R. et al. (2018) 'Support vector machine regression for calibration transfer between electronic noses dedicated to air pollution monitoring', *Sensors*, 18(11), p. 3716. doi:10.3390/s18113716.
94. Topalović, D.B. et al. (2019) 'In search of an optimal in-field calibration method of low-cost gas sensors for ambient air pollutants: Comparison of linear, multilinear and artificial neural network approaches', *Atmospheric Environment*, 213, pp. 640–658. doi:10.1016/j.atmosenv.2019.06.028.
95. Kang, J. and Choi, K. (2024) Calibration methods for low-cost particulate matter sensors considering seasonal variability, *MDPI*. doi: 1424-8220/24/10/3023
96. Rezaei, R., Naderalvojud, B., and Gullu, G.: A Comparative study of Deep Learning Models on Tropospheric Ozone Forecast- ing Using Feature Engineering Approach, *Atmosphere*, 14, 239, <https://doi.org/10.3390/atmos14020239>, 2023.
97. Nalakarathi, N.V.S.R. et al. (2024) Challenges and opportunities in calibrating low-cost environmental sensors, *Sensors* (Basel, Switzerland).
98. An Wang, Yuki Machida, Priyanka deSouza, Simone Mora, Tiffany Duhl, Neelakshi Hudda, John L. Durant, Fábio Duarte, Carlo Ratti, Leveraging machine learning algorithms to advance low-cost air sensor calibration in stationary and mobile settings, *Atmospheric Environment*, Volume 301, 2023, 119692, ISSN 1352-2310, <https://doi.org/10.1016/j.atmosenv.2023.119692>.
99. Zimmerman, N., Presto, A. A., Kumar, S. P. N., Gu, J., Haurlyliuk, 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. *Atmospheric Measurement Techniques* Vol. 11, Issue 1, pp. 291–313.
100. Clarity, Air quality monitoring 2.0 article, <https://www.clarity.io/blog/air-quality-monitoring-2-0-how-different-types-of-air-monitoring-technologies-are-contributing-to-a-more-holistic-understanding-of-air-pollution>.

101. Sofia Ahumada, Matias Tagle, Yeanice Vasquez, Rodrigo Donoso, Jenny Lindén, Fredrik Hallgren, Marta Segura and Pedro Oyola, Calibration of SO₂ and NO₂ Electrochemical Sensors via a Training and Testing Method in an Industrial Coastal Environment, *Sensors* 2022.
102. Eben S. Cross, Leah R. Williams, David K. Lewis, Gregory R. Magoon, Timothy B. Onasch, Michael L. Kaminsky, Douglas R. Worsnop, and John T. Jayne, Use of electrochemical sensors for measurement of air pollution: correcting interference response and validating measurements, *AMT* 2017.
103. Weybourne Atmospheric Observatory [Accessed on 31 January 2025].

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.