atmosphere

MDPI

Review

Low-Cost Air Quality Sensors: Biases, Corrections and Challenges in Their Comparability

Idris Hayward [1,2], Nicholas A. Martin [2], Valerio Ferracci [2], Mohsen Kazemimanesh [2] and Prashant Kumar [1,3,*]

[1] Global Centre for Clean Air Research (GCARE), School of Sustainability, Civil and Environmental Engineering, Faculty of Engineering and Physical Sciences, University of Surrey, Guildford GU2 7XH, UK; jh01724@surrey.ac.uk
[2] Air Quality and Aerosol Metrology Group, National Physical Laboratory (NPL), Teddington, London TW11 0LW, UK; nickmartin1959@me.com (N.A.M.); valerio.ferracci@npl.co.uk (V.F.); mohsen.kazemimanesh@npl.co.uk (M.K.)
[3] Institute for Sustainability, University of Surrey, Guildford GU2 7XH, UK
* Correspondence: p.kumar@surrey.ac.uk

Abstract: Low-cost air quality sensors are a promising supplement to current reference methods for air quality monitoring but can suffer from issues that affect their measurement quality. Interferences from environmental conditions such as temperature, humidity, cross-sensitivities with other gases and a low signal-to-noise ratio make them difficult to use in air quality monitoring without significant time investment in calibrating and correcting their output. Many studies have approached these problems utilising a variety of techniques to correct for these biases. Some use physical methods, removing the variability in environmental conditions, whereas most adopt software corrections. However, these approaches are often not standardised, varying in study duration, measurement frequency, averaging period, average concentration of the target pollutant and the biases that are corrected. Some go further and include features with no direct connection to the measurement such as the level of traffic nearby, converting the initial measurement into a modelled value. Though overall trends in performance can be derived when aggregating the results from multiple studies, they do not always match observations from individual studies, a phenomenon observed across many different academic fields and known as "Simpson's Paradox". The preference of performance metrics which utilise the square of the error, such as root mean squared error (RMSE) and r^2, over ones which use the absolute error, such as mean absolute error (MAE), makes comparing results between models and studies difficult. Ultimately, comparisons between studies are either difficult or unwise depending on the metrics used, and this literature review recommends that efforts are made to standardise the reporting of calibration and correction studies. By utilising metrics which do not use the square of the error (e.g., MAE), models can be more easily compared within and between studies. By not only reporting the raw error but also the error normalised by multiple factors (including the reference mean and reference absolute deviation), the variabilities induced by environmental factors such as proximity to pollution sources can be minimised.

Keywords: low-cost air quality sensor; electrochemical; optical particle counter; bias; drift; calibration; linear regression; random forest; artificial neural net; errors; Simpson's paradox

Citation: 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

Academic Editors: Dean Venables and Daniele Contini

Received: 9 November 2024
Revised: 12 December 2024
Accepted: 14 December 2024
Published: 20 December 2024

1. Introduction

Poor air quality (AQ) is one of the biggest global health concerns, with over 90% of people living in areas which exceed safe limits for air pollution according to the WHO [1]. Monitoring AQ is of the utmost importance, as it enables governments and regulatory bodies to assess the levels of pollution a population is exposed to in a geographical area, the sources of emissions and the effects of different policies. Equipment traditionally used to measure air pollution is bulky, expensive and requires skilled technicians to maintain. As a result, monitoring sites tend to be sparse and far apart, with individual stations

often covering areas of 1–10 km^2. In the worst cases, some countries have no monitoring capabilities at all [2].

To combat the sparseness of outdoor AQ measurements, a large focus has recently been put on low-cost AQ monitors that could be used to supplement current measurement practices. While not yet capable of matching the quality of reference instrumentation, the increased spatio-temporal resolution of AQ data they enable make them potentially ideal for personal exposure monitoring and more general research applications.

The concentration of air pollutants can significantly change over as short a distance as a few meters and as short a timespan as a few seconds [2–6]. Therefore, an increase in spatio-temporal coverage not only provides researchers and end-users with more granular information on air pollutants but can also be used for applications such as determining the proportionate contributions of different sources to current pollution levels (i.e., pollution source apportionment) [7].

However, the measurement quality of these low-cost sensors is affected by numerous external factors, from changing meteorological and environmental conditions to the environment where they were calibrated. The factors affecting the measurements differ between component types, manufacturers and even between two sensing components manufactured in the same batch [8–16]. By contrast, reference-grade instrumentation does not generally suffer biases from external factors due to internal components that maintain constant pressure, humidity and temperature.

If the measurement quality of low-cost sensors is to improve, the effects of these external factors need to be quantified, a task made more difficult by the intra-sensor variability causing differences in the individual sensors' responses to the same conditions. In an effort to minimise these effects, many different corrections have been implemented and are detailed in this review. The most popular countermeasure by far is in software, using multivariate machine learning algorithms, both parametric and non-parametric in nature, to attempt to remove the biases via calibration. However, the number of algorithms tested in published work is often limited, with only a few algorithms tested in a single collocation campaign. As the environmental conditions and pollution ranges that a low-cost sensor is exposed to during a campaign can vary substantially depending on location [2–6], time of year and external events such as wildfires and fireworks [17], comparing the performance of calibration techniques between different studies can be difficult.

This work reviews electrochemical sensors and low-cost optical particle counters, the two components most commonly used in low-cost sensors in recent literature, the advantages they provide over reference and equivalent monitors and their physical limitations. This includes biases that occur as the result of changes in environmental conditions and interferences from other pollutants. Many other low-cost sensing components are available, including metal oxide and non-dispersive infrared sensors. As this study focuses on results found in recent literature, they were not included in this review.

We then focus on the different algorithms used to calibrate and correct low-cost sensor components, including a collation of results from many published studies, and make suggestions based on common data science practices that can help improve the ability to compare results between studies and that would ultimately lead to better-quality data from low-cost AQ sensors.

We focus on the calibration of NO_2 and O_3 electrochemical sensors for the measurement of gaseous pollutants and the correction of low-cost OPC sensors measuring the concentration of particular matter (PM) fractions. We specifically examine the measurement of PM below 2.5 μm (PM2.5) in diameter and 10 μm (PM10) in diameter. These were chosen as they are the four most common pollutants found during the literature survey which are relevant to ambient outdoor air quality.

2. Methods

Figure 1 shows a variation of a prisma diagram that explains our methodology for selecting studies for this review. The literature review was performed by first searching for studies relevant to low-cost air quality sensor calibration from sources including Google Scholar and JSTOR. The search terms included "low-cost", "air quality", "electrochemical", "optical particle counter", "drift", "bias", "calibration" and "regression". The studies were then examined for relevance to outdoor ambient air quality monitoring using low-cost sensors. Papers which focus on correcting the measurements of pre-built devices instead of the raw sensor signals were discarded, as they were out of scope.

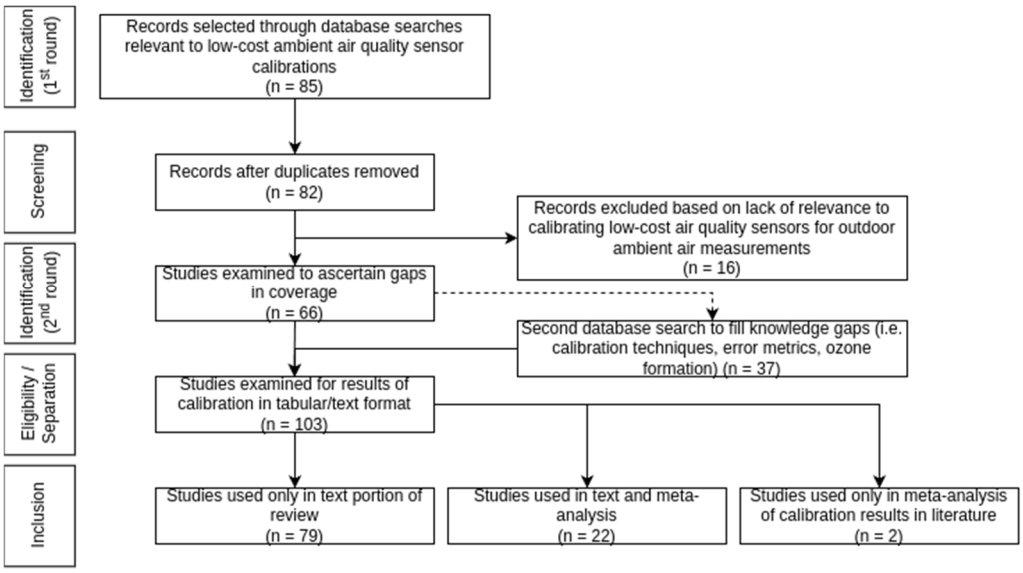

Figure 1. Literature Review Methodology.

Once the studies were collated, a second wave was conducted to collect papers relevant to the most common techniques and error metrics encountered in the studies. The information from these was then formatted to first describe the issues with the sensing components used in low-cost air quality sensors and the techniques used to correct them and assess their performance before discussing how some studies veer into the "modelling" of air quality rather than a measurement, as well as how the lack of standardisation of the methodology makes comparing results across two studies difficult.

A meta-analysis of results provided in papers was also conducted, using results that were provided in the text of the paper. This analysis summarised the previous results in the literature to attempt to draw a conclusion as to the best techniques to use when calibrating low-cost air quality sensors but was ultimately found to be inconclusive.

3. Component Types

A range of different low-cost sensing components can be used to measure AQ-relevant species, each with their own advantages and drawbacks. Some are specific to certain pollutants (e.g., electrochemical sensors monitoring NO), others are more general (e.g., OPCs measuring the count of particulate matter as a whole rather than specific particulates such as nitrates and oxides, (NO_2 + O_3, commonly referred to as O_x)).

3.1. Electrochemical Sensors

Electrochemical (EC) sensors are the most popular low-cost sensor component for measuring the concentration of any redox-capable gases (e.g., NO, CO), particularly in recent years. They are more sensitive than metal oxide sensors [2,18,19] and have faster response times (around 1 min) [4,5,14,20,21], though this depends on the flow rate of air across the sensor [21]. Their response is linear with respect to the target gas [4,13–15,17,19–26], though

the linearity is lost with fluctuating temperature and humidity [19,23,25,26] and also at high temperatures (>30 °C) [17]. The limit of detection for EC sensors is very low (few parts per billion), allowing them to be used for ambient air measurements [13,15,20–22,27]. This high sensitivity is the reason the measurements are so susceptible to biases from environmental factors [13]. The lifetime of EC sensors is around 1–2 years; the sensing component is generally considered to have reached its end of life when its reading varies by 50% from when it was new due to the loss of sensitivity and baseline drift [2,4,18,21,27,28].

3.1.1. Operating Principle

EC sensors consist of an electrochemical cell which is covered by a gas porous membrane, containing an electrolyte solution and three electrodes (working, counter and reference) [29,30]. The working and reference electrodes have a potential difference between them that corresponds to the redox potential of the target gas, allowing for selectivity in the measured pollutant. The potential difference is kept constant to ensure the redox reaction goes to completion. Atmospheric gases diffuse through the membrane into the electrolyte solution. When a molecule of the target gas comes into contact with the working electrode, a redox reaction occurs which is balanced out by the counter electrode [31,32]. The current measured across the working and counter electrodes is proportional to the concentration of the target gas, though it is also affected by changes in environmental conditions such as temperature. Some EC sensors are supplied with a fourth, auxiliary electrode which is not exposed to atmospheric gases and theoretically only changes current in response to external environmental factors, thus enabling correction for these effects (See Section 3.1.2) [29].

3.1.2. Biases

EC sensors are very susceptible to changes in their environments, with a range of factors affecting their measurements. Changes in temperature have a large impact [2,5,6,13–16,20–22,24–27,29,33–42]. It has been suggested that a change in temperature has a more significant impact on the measurement than the temperature itself [26]. The temperature affects the measurement by altering the sensitivity of the sensor to the target gas [13,14,25,35,38,40], which has a non-linear effect on the measurement [26,27,41]. As the temperature increases, the sensitivity decreases linearly until 30 °C. Past this point, an EC sensor's dependence on temperature becomes non-linear while its sensitivity to the target pollutant remains low. The sensitivity increases again above 40 °C [14].

Relative humidity (RH) also has a large influence on measurements [2,5,6,13–17,20–22,25–27,29,30,35–39,41–43]. It has a non-linear effect [26,41], typically only significant at very low or very high humidity (<15% or >90%) [22,30,35,38]. The rate of change in relative humidity has been known to cause large spikes in the measurement [14,17,26]. These spikes were prominent in Sun et al., 2017 [17], where, when switching from ambient air to air passed through a Nafion dryer, a sharp temporary spike in the EC sensor response from the sudden drop in RH was recorded.

Wind speed has also been identified as a potential bias in EC sensor measurements [2,6,20,22,33,43], most likely due to higher wind speeds preventing gases homogenising in the vicinity of the sensor and limiting diffusion across the EC membrane [22,33,44].

Additionally, air pressure has been noted to have an effect on EC sensor measurements [2,6]. However, a study that took place 1500 m above sea level reported no degradation in measurement quality as a result of the low pressure, as it was calibrated in a low-pressure environment. This indicates that, as long as the sensor is calibrated for the appropriate conditions, pressure should not pose an issue. However, fluctuations in pressure can still affect measurement quality [6].

Changes in the surface area of the electrolyte have been identified as a potential source of bias in measurements made by EC sensors. When the shape of the meniscus on the electrolyte's surface fluctuates, the coverage of the electrode changes, modifying the reaction rate between the working electrode and the target gas. These fluctuations can be caused by changes in temperature, relative humidity, wind speed and pressure [6]. Electronic noise from the circuits can also interfere with the measurement by biasing the measured current with a random offset [16].

EC sensors are generally selective [2,4,13,33,45]. However, they still suffer from some cross-sensitivities with other atmospheric trace gases [5,15,16,20,24,29,36,37,46,47]. Previous studies have tested these in detail, most notably Lewis et al. [46]. Most cross-sensitivities only have a minor effect on the signal, so they only impact the measurement at low concentrations of the target gas, effectively increasing the LOD [46]. However, some can have a significant impact.

The worst cross-sensitivity for EC sensors is the one between NO_2 and O_3, two key AQ pollutants which have the same reducing potential at the working electrode [6,23] and are therefore indistinguishable by EC sensors [2,6,8,13–15,21–23,29,43,48]. EC sensors are also affected by long-term constant drifts in the baseline of the measurement, occurring in a random direction at a random magnitude [2,6,14–17,19,21,22,25,27,29,36,37,49,50]. This permanent change can be caused by the aging and degradation of electrodes [6,14], electrolyte evaporation over time [27] and dust clogging pores of the membrane [6,50]. Baseline drift can cause significant bias in the measurements in as little as a few weeks, requiring frequent recalibration [2,19,22,25].

3.2. Optical Particle Counters

Low-cost OPCs (optical particle counters) are less accurate than high-cost reference and equivalent monitors when measuring particulate matter (PM) fractions but have been shown to be a usable cheaper alternative for selected applications [9,51,52]. They are the most widely used low-cost PM sensors. Detailed comparisons between different OPCs have already been carried out [53].

3.2.1. Operating Principle

OPCs collect and measure the intensity of light scattered by a single particle as it crosses a focused beam of light [54]. A photodetector collects the scattered light, with its response dependent on the particle's size, refractive index, shape and surface roughness [4,9–11,30,51,53–57].

OPCs cannot detect particles below 0.3 µm in size as they are limited by the wavelength of light used (typically in the 600 nm range). If the particles are smaller than the wavelength, they will not scatter the light. Therefore, ultrafine particle measurements are not possible with low-cost OPCs [11,13,48,58,59]. OPC measurements also assume that all particles are spherical and of uniform density and shape, which does not necessarily represent the real-world composition [52]. As a result of variability between low-cost OPCs during manufacturing, different OPCs have different responses to the same conditions [9–12].

3.2.2. Biases

There are conflicting reports about the effects of temperature on OPC measurements, with some studies identifying temperature-driven biases [10,56,57,60] and others not finding any significant effects [9,55,61].

However, the magnitude of bias from temperature was not significant, particularly compared to the effects of RH [10,60]. One study also found that temperature was only a useful predictor when relative humidity was not accounted for; once relative humidity was also incorporated as a predictor for potential bias, the contribution of temperature was insignificant [57]. This effect could potentially be related to the effect of temperature on relative humidity, as RH is a function of the absolute humidity and temperature [56].

RH causes significant bias in OPC measurements [10,12,30,36,40,52,53,56,59–62]. At high RH, biases occur for two main reasons. Inorganic particulates (such as NaCl, nitrates and sulfates) show large hygroscopic growth above 85% RH. This increased growth changes their refractive index, biasing the measurement high [36,52,56,59]. Water droplets also condense at high RH levels and can scatter the OPC laser if they enter the sensor, thus biasing the measurement high [30,59].

As the OPC response changes with the refractive index of the particles measured, changes in particle composition can cause biases in the measurement. Secondary inorganic particulates (e.g., ammonium nitrate, ammonium sulfate, ammonium chloride) can cause significant overestimation [9]. As aerosol properties vary over time and by location, both short- and long-term biases can be introduced by seasonal changes, local events and/or a change in location. Local events can include, but are not limited to, forest fires [9,18], fireworks [8,18] and changes in both wind direction and speed [10,56,58,60].

Only a few publications reported baseline drift in OPCs [9,10,53,63], and even where it was noted, it sometimes appeared to be a malfunction in individual units [9,10]. Baseline drift in OPCs can occur due to the degradation of components in the instrument, an altered flow rate or contamination [53,63]. The buildup of dust alone is not the cause, as cleaning the instrument does not return it to its original baseline [63].

3.3. Chosen Environments and Their Influence

Low-cost sensors can be calibrated in laboratories with the use of environmental chambers [53]. These chambers can control the environmental conditions within them (e.g., temperature, RH), as well as the concentration of different pollutants. This level of control has the advantage of enabling accurate tests of LCS systems over a wide range of conditions. However, as the tests needed to thoroughly characterise the response of sensors to specific conditions can be lengthy, resource- and time-intensive and comprehensive laboratory tests may not always be practical [8,15,18,27,38,41,43,49,53,64].

A more viable alternative is to run selected tests in the laboratory followed by a field calibration by co-location with reference-grade instruments, or simply running all tests in the field for a long enough period that a significant range of conditions are observed.

The most practical way of calibrating sensors is to perform the calibration in ambient air, attempting to cover as wide a range of conditions as possible [8,15,18,27,38,41,43,49,53,64,65]. The wider the range of conditions in which the sensor is tested, the more information can be used to determine how the sensor responds to different conditions. However, it may not be possible to cover the full range of changes in parameters that concern the measurement and its biases with a single field calibration [15,16,20,24,25,41,49].

Though field calibrations are preferred to laboratory calibrations in terms of both time and practicality, it is recommended that low-cost sensors are calibrated in a similar environment to the one they are deployed in, as it has been shown that calibration parameters that work well in one geographical region can perform worse in others [14,41,43,44,49,53].

4. Common Calibration Techniques

4.1. Manufacturer Calibration

LCS system manufacturers often provide factory calibrations for their sensing components, with several studies utilising factory calibrations for EC sensors provided by Alphasense [29]. If the sensor component communicates digitally, the device will generally transmit a factory-calibrated measurement when requested. If the component is analog, the manufacturer may provide an equation and coefficients to convert the voltage output to a concentration measurement [29,40].

Some studies have had success with these calibrations when comparing against reference-grade measurements [29,40,43], but given the variations in performance when sensors are moved from one environment to another and baseline drift, caution should be taken before assuming these calibrations represent an accurate measurement.

4.2. Linear Regression

Linear regression is by far the most common regression technique found across the studies examined in this review, with 10 utilising it for NO_2 (Figure 2a), 8 for O_3 (Figure 2b) and 7 for PM2.5 (Figure 2d). Linear regression appears in two forms in the literature. Univariate linear regression (ULR) models a single variable (x) that has a linear relationship with a response (y) [66,67]. ULR equations typically take the form of Equation (1). Linear regression appears in two forms in the literature. Univariate linear regression (ULR) models a single variable (x) that has a linear relationship with a response (y) [66,67]. ULR equations typically take the form of Equation (1).

$$y = a_0 + a_1 x + \epsilon \quad (1)$$

x represents the uncorrected measurement and y represents the corrected. a_0 represents the intercept and a1 represents the gradient. ϵ represents a random error component. Random errors are assumed to be uncorrelated, have a mean of 0 and have an unknown variance of σ^2.

MLR (multivariate linear regression) models multiple variables (x_z) that have linear relationships with respect to their effect on the response (y) [68,69].

MLR equations typically take the form of Equation (2).

$$y = a_0 + a_1 x_1 + a_2 x_2 + \ldots + a_z x_z + \epsilon \quad (2)$$

x_n represents the uncorrected measurement and other measurements such as temperature. y represents the corrected measurement. a_0 represents the intercept. a_n where $n > 0$ represents the intercept. ϵ represents a random error component. The coefficients are calculated via least squares regression, and their values are determined by minimising the sum of the squared differences between ground truth observations ($\hat{y}$) and the modelled straight line ($a_0 + a_1 x$) [68].

MLR is the most common calibration technique found in the literature, with many publications at the very least employing it as a baseline for a comparison with other calibration techniques [8,12,14–16,24,25,27,29,38,41,42,47,56,57,69]. Liang et al. tested several variations on the linear formula to try and capture the non-linear nature of input variables such as temperature and RH [29].

Many variables have been used in different MLR calibrations, including

- Output of EC sensor working electrode [14,27,38,42]
- Output of MOx sensor [8,42,69]
- Output of multiple MOx sensors [69]
- Output of OPC [56,57]
- Output of EC sensor auxiliary electrode [14,27,42]
- Temperature [8,14,25,27,30,38,41,42,57,69]
- Temperature [2,41]
- Humidity [14,25,30,38,41,42,56,57,69]
- Humidity [2,41]
- Temperature × Relative Humidity [41]
- Time [8]
- Solar radiation [8]
- Cross-sensitive gases [14,42,69]
- Wind speed [30]

The use of time is particularly interesting as it could potentially be used to calibrate against baseline drift [8,49], though seasonal variations may be difficult to separate from baseline drift over short time periods. However, a long-term calibration campaign may be cost- and time-prohibitive, as well as significantly eat into the expected lifetime of EC sensors (1–2 years) [2,4,19,22,27,28].

The performance of linear regression is highly variable across applications to NO_2, O_3 and $PM_{2.5}$ data (Figures 3–5). This can partially be explained by the different ranges of features used. Multiple studies utilise ULR with only the raw sensor measurement as a feature [31,41,42,68,70–73], while deSouza et al. [70] utilise up to 15 features when correcting $PM_{2.5}$ measurements. The performance of the 15 feature MLR correction varied drastically depending on the size of the train:test set split, with significantly greater performance when 2 weeks of sequential measurements from winter and 2 weeks of sequential measurements from spring are used as a training set (r 0.765, RMSE 6.74 μg m^{-3}) than when only the winter measurements are used (r 0.324, RMSE 32.95 μg m^{-3}).

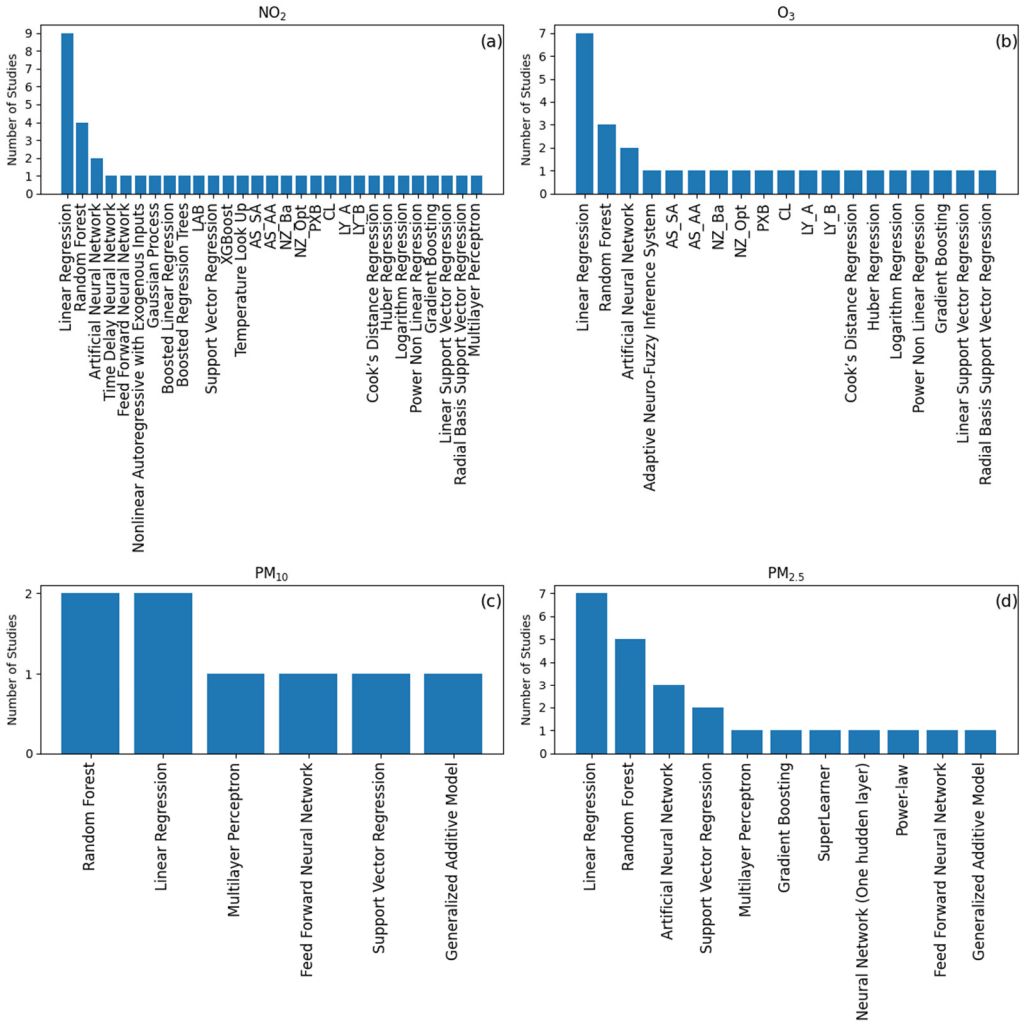

Figure 2. Summary of the different calibration techniques used in the studies reviewed, separated into (**a**) NO_2, (**b**) O_3, (**c**) PM_{10} and (**d**) $PM_{2.5}$.

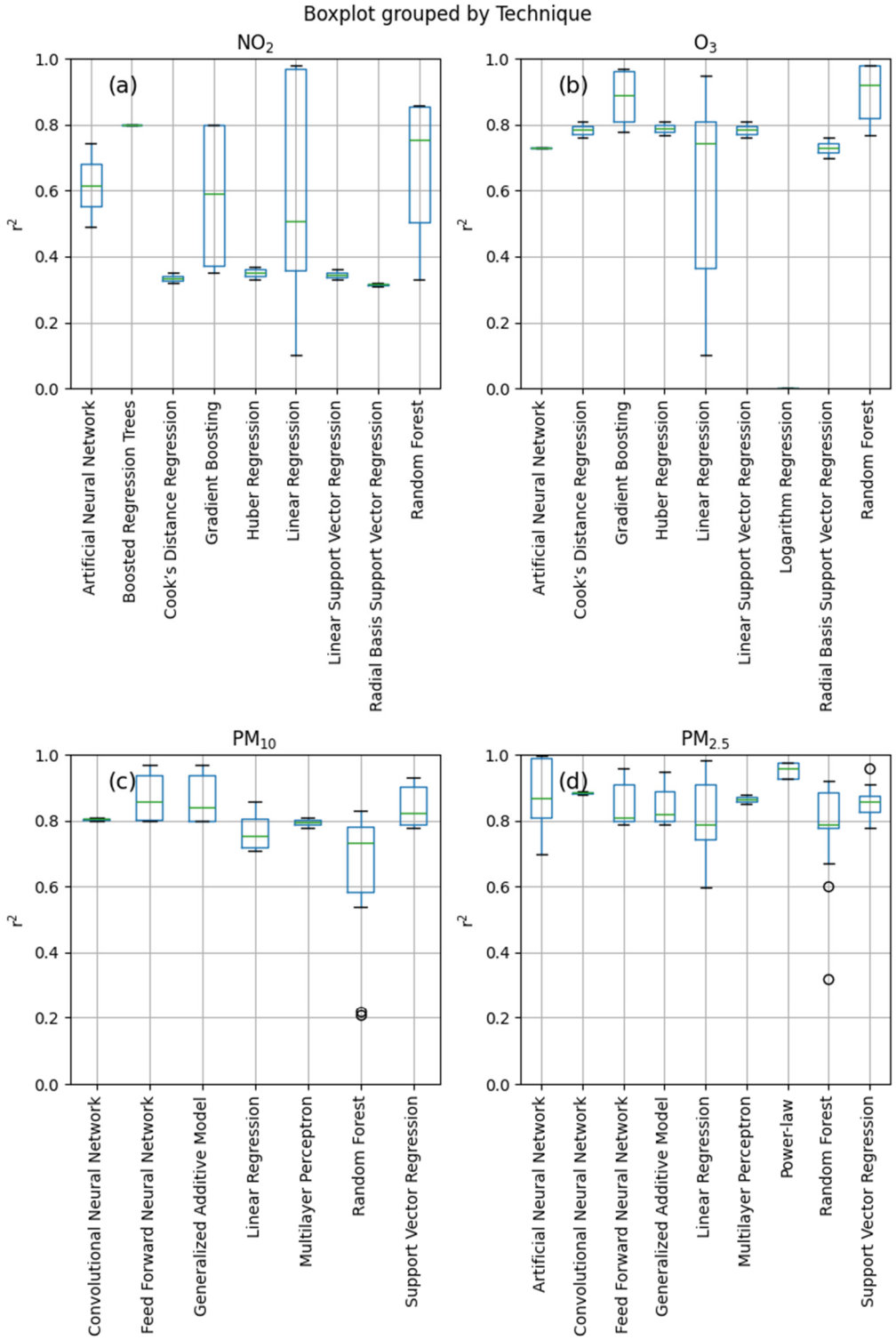

Figure 3. Summary of the r^2 results in the studies reviewed, separated into (**a**) NO_2, (**b**) O_3, (**c**) PM_{10} and (**d**) $PM_{2.5}$. The green line represents the median for each regression technique, the box represents the interquartile range and the whiskers represent the lowest and highest points within the median plus/minus 1.5 times the interquartile range. Dots represent outliers outside of the whiskers.

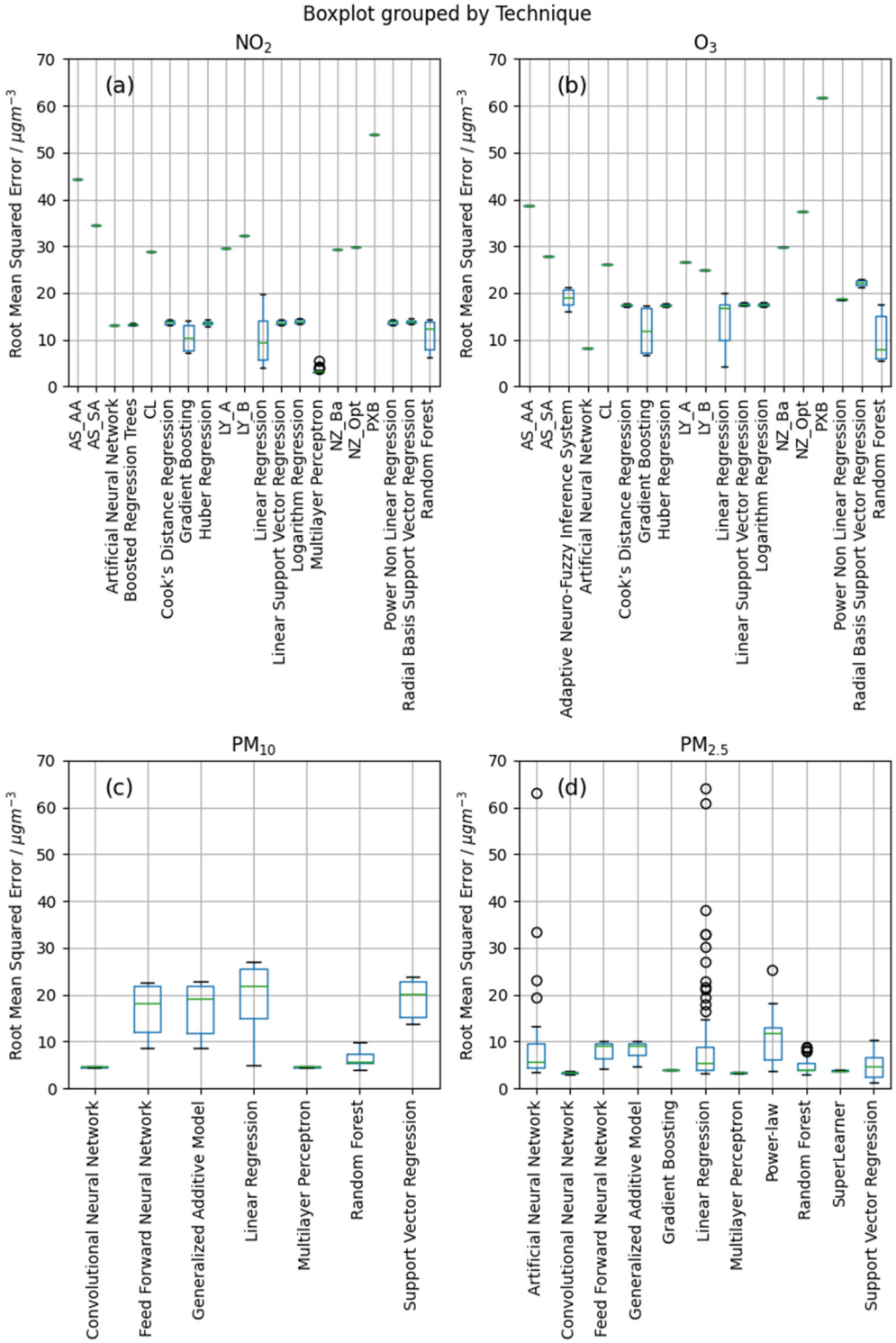

Figure 4. Summary of the RMSE results in the studies reviewed, separated into (**a**) NO_2, (**b**) O_3, (**c**) PM_{10} and (**d**) $PM_{2.5}$. The green line represents the median for each regression technique, the box represents the interquartile range and the whiskers represent the lowest and highest points within the median plus/minus 1.5 times the interquartile range. Dots represent outliers outside of the whiskers.

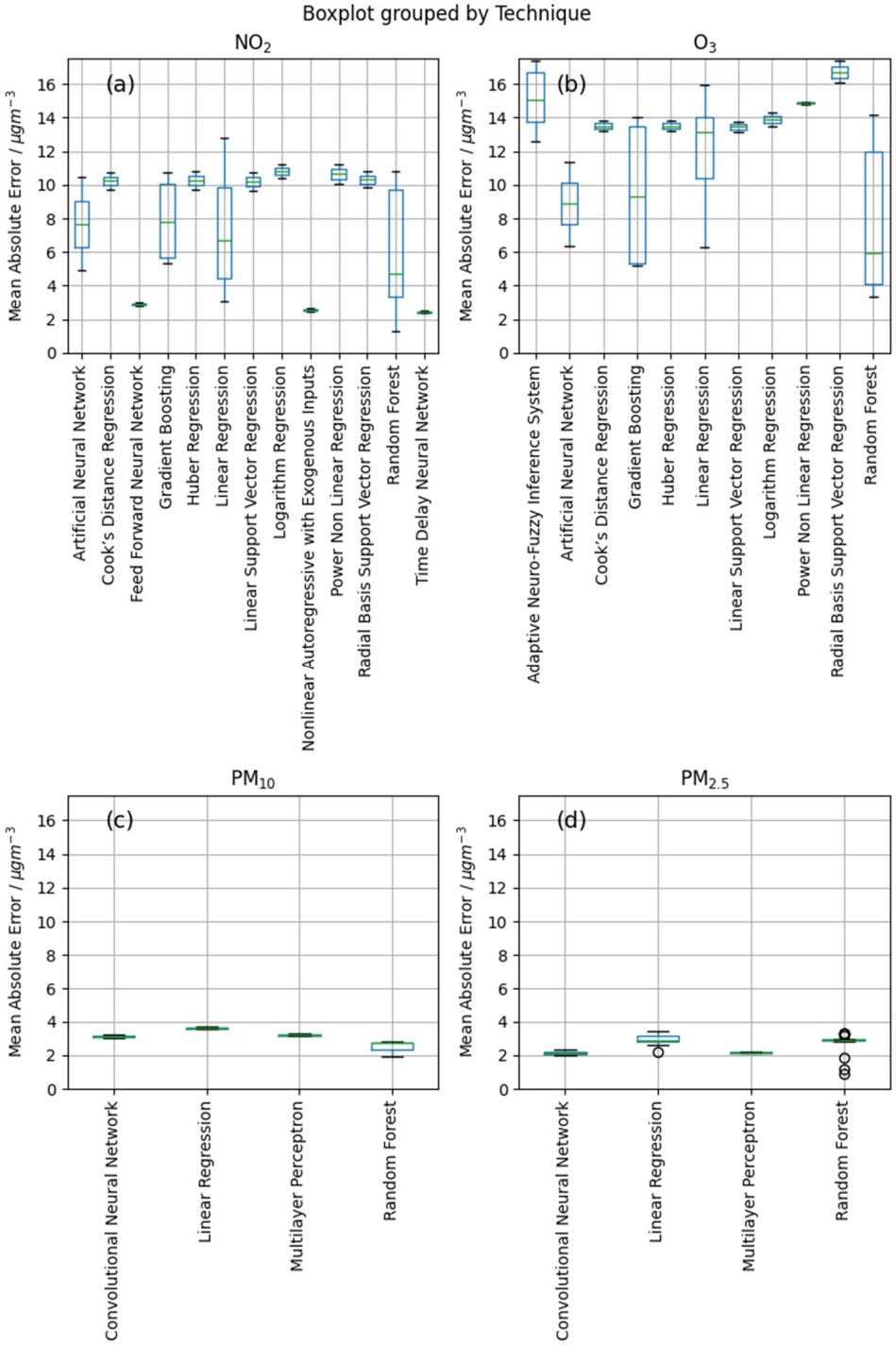

Figure 5. Summary of the MAE results in the studies reviewed, separated into (**a**) NO_2, (**b**) O_3, (**c**) PM_{10} and (**d**) $PM_{2.5}$. The green line represents the median for each regression technique, the box represents the interquartile range and the whiskers represent the lowest and highest points within the median plus/minus 1.5 times the interquartile range. Dots represent outliers outside of the whiskers.

4.3. Random Forest

Random Forest (RF) was the second most common regression technique found in our review of the literature, with five studies utilising it for NO_2 (Figure 2a), four for O_3 (Figure 2b) and six for $PM_{2.5}$ (Figure 2d). For the purpose of low-cost ambient AQ sensor calibration, RF is used to predict the measurement value from a set of inputs, including the sensor output and environmental measurements [24,25,38,41,68,69]. However, it performs

poorly outside the range it was trained in, resulting in poor predictions [25]. RF has been used to describe several different algorithms [74] but generally refers to the one described in Breiman, 2001 [75]. "Forests" are constructed from a collection of decision trees. Decision trees take one or more inputs and test them against a series of true/false statements based on the inputs provided. For example, one branch could test whether an input was greater than or equal to a specified value determined in the training stage. These branches continue until a final output value is reached, and the number of branches is decided by the user before training. Each output is then used as a "vote" for the calibrated value [74,75].

It is a more consistent performer than linear regression in terms of its RMSE (Figure 5), MAE (Figure 6) and r^2 (Figure 4), with the exception of RMSE for NO_2, where it performs, on average, worse than linear regression and r^2 for $PM_{2.5}$ and PM_{10}, where there are significant outliers that perform worse than the average (these performance metrics are covered extensively in Section 5.2).

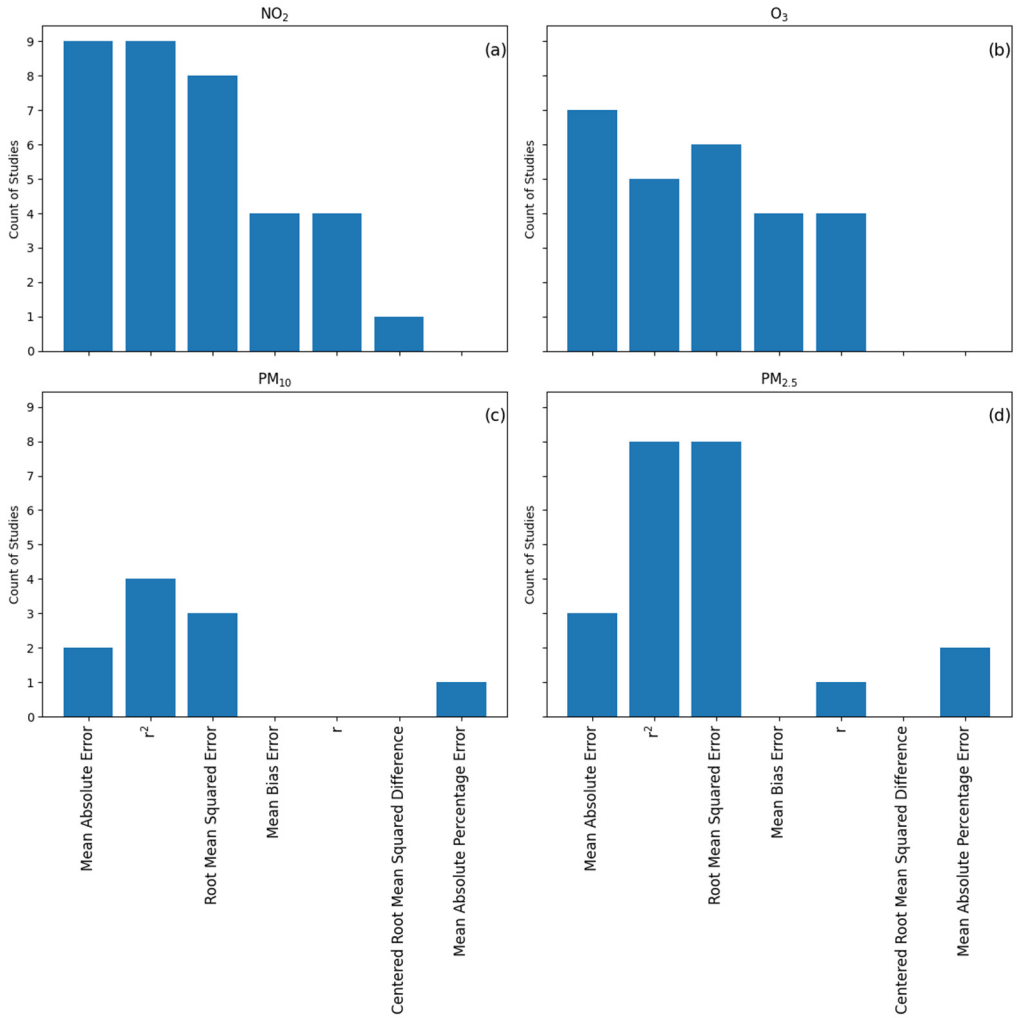

Figure 6. Summary of the different metrics used in the studies reviewed, separated into (**a**) NO_2, (**b**) O_3, (**c**) PM_{10} and (**d**) $PM_{2.5}$.

The average RMSE value being higher for RF than linear regression in NO_2 calibrations is difficult to explain. In the four studies that utilise both techniques for NO_2 calibration [25,38,71,76], four out of six RF calibrations result in a lower RMSE than linear regression (Zimmerman et al., 2018 does not report RMSE). The two that perform worse only use NO_2 as features, so they do not account for biases from environmental factors such as temperature and RH. The introduction of other features in RF calibration more than halved the RMSE, as reported by Cavaliere et al., 2023, while providing less of a

performance improvement for LR, suggesting that single-feature calibrations are not the optimal way to use RF.

The low r^2 outliers in $PM_{2.5}$ and PM_{10} can be explained in the same way as those in NO_2. The outlier r^2 values come from Veiga et al., 2021, where only raw OPC measurements were included as a feature (−0.13 for $PM_{2.5}$ and 0.22 for PM_{10}), or raw OPC measurements, internal temperature and internal RH measurements (0.32 for $PM_{2.5}$ and 0.21 for PM_{10}).

The larger variance in r^2 scores can also be explained by its inappropriateness as a metric in comparing different models [77] and models trained on different sizes of datasets [78], as discussed in Section 5.2.1.

4.4. Artificial Neural Network

For the purposes of low-cost ambient AQ sensor calibration, Artificial Neural Networks (ANNs) are used to predict the non-linear relationships between input and output variables. They appear under several different names in the literature depending on their specific operations. Artificial Neural Networks [25,31,41,68–70,72], Convolutional Neural Networks [79], Multilayer Perceptrons [79,80] and Feed Forward Neural Networks [81,82] were all encountered during the literature search.

By modelling highly complex non-linear interactions, very accurate measurements can be obtained. However, they require large amounts of training data, which is impractical in many situations [8,12,24,25,41,68].

ANNs are made up of multiple layers of "neurons", which perform non-linear transformations on their inputs. Each neuron scales one or more inputs before passing it to a functional neuron which applies linear or non-linear transformations. The output is then passed to the next layer. Some outputs can feed back to neurons in the same layer. The final layer is known as the output layer [83–85].

ANNs consistently provide the lowest RMSE and MAE scores on average (Figures 3 and 4) but are outperformed by RF for both NO_2 and O_3 when looking at the r^2 score. Feed Forward Neural Networks are the best performers, on average, for PM_{10} corrections when looking at r^2 scores (Figure 4c). Convolutional Neural Networks are the second-best performers, on average, for $PM_{2.5}$ corrections when looking at r^2 scores, beaten by Power Law regression (Figure 4d).

5. Model Validation

5.1. Model Selection

5.1.1. Train–Test Split

In order to properly assess the calibration and correction models used, the dataset must be split into a training set, where the model is fit to minimise the error between the features provided and the "ground truth" reference data, and a test set, where the validation metrics are calculated. This avoids "overfitting", a problem with machine learning models resulting from models fitting to idiosyncrasies in the dataset such as noise [86]. Any model that performs significantly better in the training set than in the test set is likely overfitted to variables irrelevant to the actual measurement and should be discarded.

This approach works well when testing a single model, but when it comes to model selection, testing multiple combinations of features, such as Ko et al., 2024 [73] and Ali et al., 2023 [87], the models should be optimised using results from the test set and the final result calculated using a separate validation set [88].

There are incredibly large variations in train:test splits in the literature, with very few [31,69,72,88] utilising validation sets. Train:test proportions range from 92:8 in Koziel et al., 2024 [80] to 3:97 in Liang et al., 2021 [29].

5.1.2. k-Folds Cross Validation

Not all measurements are made equally. Some can be significantly impacted by noise or other biases, and if a disproportionate number of these are used in the test or training sets, the model performance will be significantly lowered. The most common way of

accounting for this in the literature is by utilising *k*-folds cross validation, where the dataset, after optionally removing a validation set, is split into *k* equally sized datasets, all split from the whole dataset and all using unique test sets. For example, $k = 5$ would result in 5 sets split by an 80:20 train:test split, and $k = 10$ would result in 10 sets split by a 90:10 [86]. Several studies utilised this technique [25,38,69], though it is far less common than a single train:test or train:validation:test split. It is therefore difficult to determine whether their results would apply to the real world or whether they are the result of overfitting. The perceived lack of a split in the data may be the result of the low number of measurements (154) compared to other studies (262080 in Liang et al., 2021 [29]), potentially making splitting the data into separate sets difficult.

5.2. Metrics

5.2.1. r^2

r^2 is the most common metric found in the literature, aside from O_3 studies, where MAE is more popular, and r is more commonly used as a measure of correlation than other pollutants (Figure 6). It can be calculated using Equation (3).

It is a measure of the "goodness of fit" and is regularly used to compare competing regression models within studies, though it has often been described as inappropriate when comparing techniques that transform or otherwise model the data in different ways [77,89], such as RF and MLR.

It has been reported to vary substantially depending on the number of variables introduced. Fisher, 1924 stated that the probability of scoring an r^2 value over a specified amount increases significantly with the number of variables used in the model [90,91]. Therefore, any study that uses multiple variables (e.g., temperature, relative humidity) to predict air quality measurements using an LCS system cannot be compared to one which only uses the response of the sensing component measuring the target pollutant.

The performance of the technique can also decrease with an increase in the number of measurements, making comparisons between studies of varying lengths (Kang et al., 2024's [82] 2 years vs. Mahajan et al., 2020's [68] 15 h) and varying sampling frequencies (Agrawal et al., 2024's [72] 1 min vs. Ko et al., 2024's [73] 1 h) difficult [80]. It also does not capture the full performance of a model and should only be used in tandem with other statistics [92,93].

$$r^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \overline{y})^2} \quad (3)$$

where N represents the total number of measurements, y_i represents the ground-truth measurement at time i, $\overline{y}$ represents the mean of ground-truth measurements and $\hat{y}_i$ represents the predicted i [89].

5.2.2. Root Mean Squared Error

Root Mean Squared Error (RMSE) is the second most common metric found in the literature and the most common error metric (Figure 6). It can be calculated using Equation (4). It follows the formula for general error given in Equation (5), where $\gamma = 2$.

RMSE calculations follow three steps. First, the total squared error is calculated by summing the squared difference between the predicted measurement and the ground truth. As the errors are squared, large errors have more influence on the overall result than smaller errors, giving large outliers a disproportionate influence. The total SE is then divided by the number of measurements to obtain the MSE, and then the square root of the MSE is taken to get the RMSE.

The RMSE results (Figure 4) are far more varied than the MAE results (Figure 5), largely due to Liang et al., 2021 [29], which contained a significant number of high measurement errors in NO_2 and O_3 reporting RMSE and not MAE.

The use of RMSE over MAE has raised some criticism in previous literature as it does not just vary with the total error but also with variance in error magnitudes. Willmott

& Matsura, 2005 [94] posited that, given that RMSE does not just increase with the total error, it is an inappropriate metric to compare the results of different regressions. RMSE's relationship to MAE follows the following rule: $MAE \leq RMSE \leq \frac{1}{N} \cdot MAE$, and the fact that RMSE does not increase proportionately with MAE suggests that it is unsuitable for comparing two models evaluated over different lengths of time [94], similar to r^2 [78].

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \quad (4)$$

where N represents the total number of measurements, y_i represents the "ground-truth" measurement at time i and $\hat{y}_i$ represents the predicted measurement at time i.

$$E^{\frac{1}{\gamma}} = \left[\frac{\sum_{i=1}^{N}\omega_i|y_i - \hat{y}_i|^{\gamma}}{\sum_{i=1}^{N}\omega_i}\right]^{\frac{1}{\gamma}}; \gamma > 0 \quad (5)$$

where N represents the total number of measurements, y_i represents the "ground-truth" measurement at time i, $\hat{y}_i$ represents the predicted measurement at time i, γ represents the order of the error, with 1 corresponding to MAE and 2 to RMSE, and ω_i corresponds to a weighting given to a measurement, where higher values give it more significance in the error calculation. It defaults to 1 [95].

5.2.3. Mean Absolute Error

Mean Absolute Error (MAE) is the third most common metric found in the literature and the second most common error metric (Figure 6). It follows the formula for general error (Equation 5), where $\gamma = 1$. It can be calculated using Equation (6). The use of MAE was found to be preferable to RMSE by Willmott & Masura, 2005 [94], as the addition of another measurement can only vary the MAE by the magnitude of the error, without the potential increase from an increase in the variance between error magnitudes. Despite this, it is very uncommon to see it used for $PM_{2.5}$ and PM_{10} corrections, with RMSE being far more prevalent. It is, however, just as or more common when validating NO_2 and O_3 calibrations, making the performance of different techniques easier to compare.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \quad (6)$$

where N represents the total number of measurements, y_i represents the "ground-truth" measurement at time i, and $\hat{y}_i$ represents the predicted measurement at time i [93].

5.2.4. Mean Bias Error

Mean Bias Error (MBE) is less common than RMSE or MAE and is not used at all for $PM_{2.5}$ and PM_{10} (Figure 3). It is similar to MAE but does not convert the error to an absolute number, meaning it sums negative and positive numbers. It shows the average bias in a model's predicted measurements, positive or negative. It can be calculated using Equation (7).

If a model suffers from significant random error centered at 0, this will not be captured by the MBE. However, the overall bias can still be useful when comparing techniques. Figure 7 shows the MBE for NO_2 and O_3 calibrations. Techniques commonly found in readily available packages (e.g., scikit-learn [96]), such as Linear Regression, Gradient Boosting and Random Forest, tend to have an MBE close to 0, whereas the more bespoke techniques showcased in Liang et al., 2021 tend to show significant bias in one direction, particularly AS_SA, PXB and NZ_Ba.

$$MBE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i) = \overline{y}_i - \hat{\overline{y}}_i \quad (7)$$

where N represents the total number of measurements, y_i represents the predicted measurement at time i and $\hat{y}_i$ represents the "ground-truth" measurement at time i [93].

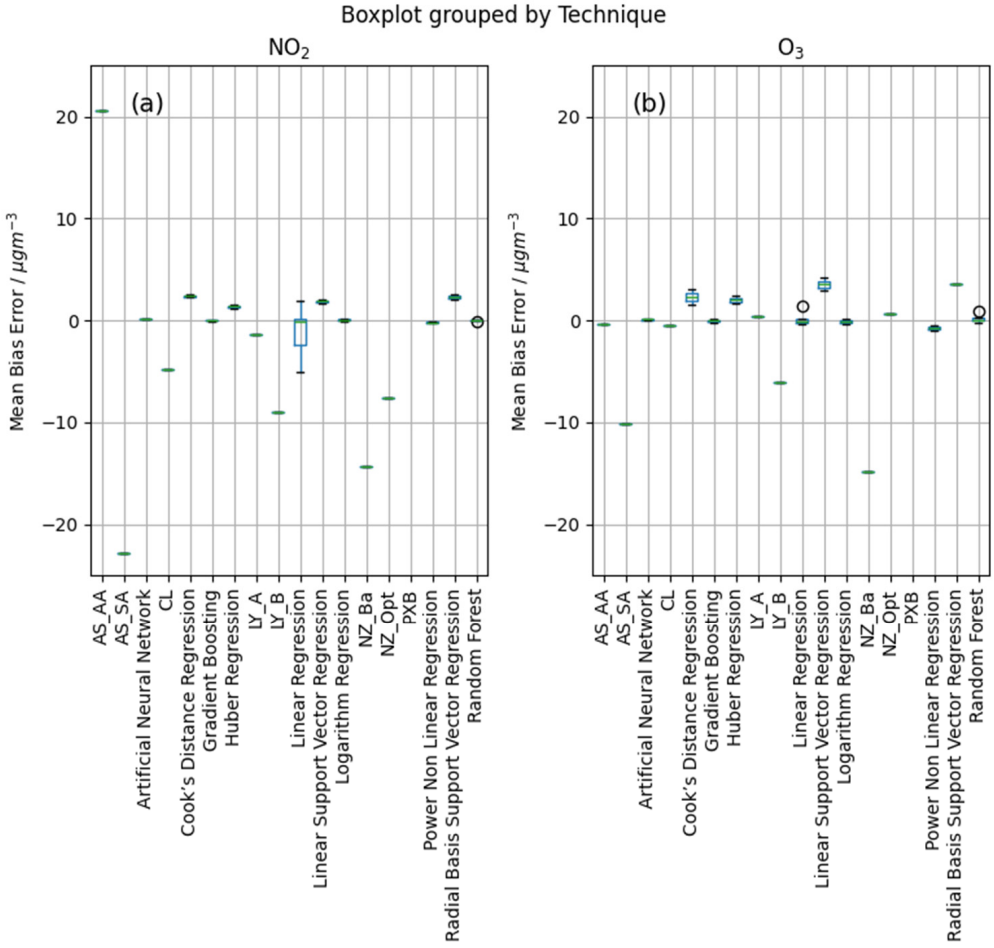

Figure 7. Summary of the MBE results in the studies reviewed, separated into (**a**) NO_2 and (**b**) O_3. The green line represents the median for each regression technique, the box represents the interquartile range and the whiskers represent the lowest and highest points within the median plus/minus 1.5 times the interquartile range. Dots represent outliers outside of the whiskers.

5.3. Features Used

The number of variables used in the regression algorithms that calibrate or correct LCS measurements is highly variable across the literature. The assumption is that by including features that are known to affect the measurement, such as temperature and RH, these biases can be corrected for. Some studies go further, introducing features such as solar radiation [8], wind speed and direction [76], traffic level [97,98] and "the angle at which the sun reaches its zenith in the sky at a specific location" [82], even though these features directly impact the concentration of the target pollutant instead of the measurement of the pollutant, blurring the lines between measurements and a model (See Section 5.5).

Cavaliere et al., 2023 [71] test multiple variable combinations for both NO_2 and O_3 (Figures 8 and 9). They mostly test single-feature calibrations (i.e., only NO_2 or O_3), testing a variety of techniques across two different LCS. They also test the inclusion of the square and the cube of the raw NO_2 and O_3 measurement using a variant of linear regression called polynomial regression, which includes polynomial versions of the features to correct for non-linear trends.

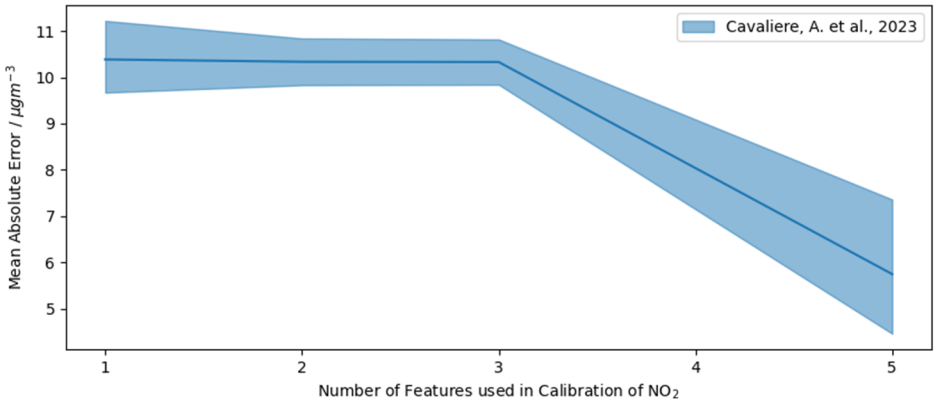

Figure 8. MAE by the number of features used in the correction of NO_2 in Cavaliere, A. et al., 2023 [71]. The shaded area represents the minimum and maximum values, and the solid line represents the mean.

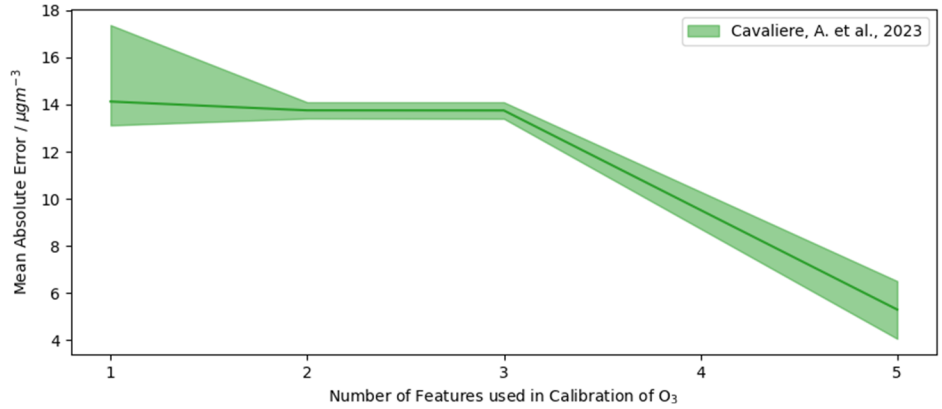

Figure 9. MAE by number of features used in the correction of O_3 in Cavaliere, A. et al., 2023 [71]. The shaded area represents the minimum and maximum value, and the solid line represents the mean.

Finally, they test, using the measurements from both sensors, the internal and external temperature of the LCS and the RH. Whereas the polynomial regression does not affect the MAE much (Figures 8 and 9), the inclusion of the additional sensor, temperature and RH significantly improves the MAE, particularly for O_3, which has an average MAE of 14 µg m^{-3} when using one feature vs. 6 µg m^{-3} for 5. This matches observations that there is significant cross-sensitivity between NO_2 and O_3 when measuring O_3. Though NO_2 sensors can have this bias mitigated with a filter that removes O_3, there is no equivalent filter that removes NO_2 for O_3 measurements [2,6,8,13–15,21–23,29,43,48].

Both Ko et al., 2024 [73] and Kureshi et al., 2022 [79] test multiple variables for $PM_{2.5}$ and assess them with MAE. However, Kureshi et al., 2022 [79] do not report the number of variables used in each correction, making it difficult to assess how performance changes with the number of features.

Ko et al., 2022 tested the inclusion of multiple features in both LR and RF corrections of $PM_{2.5}$, including $PM_{2.5}$, T, RH, NO_2 and second- and third-order interaction features based off these initial features (e.g., $NO_2 \times T$, $PM_{2.5} \times T \times RH$). The average MAE when compared to the number of features used follows an exponential decay pattern, with the elbow of the curve occurring at four features used (Figure 10). The minimum MAE for four features occurs when using $PM_{2.5}$, T, RH and NO_2 with RF, showing that the inclusion of interaction variables achieves little improvement [73].

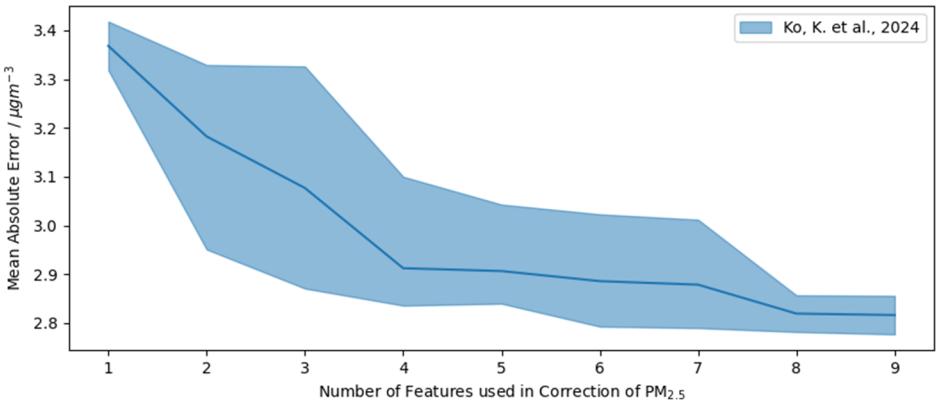

Figure 10. MAE by the number of features used in the correction of $PM_{2.5}$ in Ko, K. et al., 2024 [73]. The shaded area represents the minimum and maximum values, and the solid line represents the mean.

5.4. Variations in Performance and Simpson's Paradox

Simpson's Paradox describes phenomena where outcomes from results from individual experiments are inconsistent with outcomes when the results are aggregated [99]. For example, Cavaliere et al., 2023 [71] compare LR and GB for both NO_2 and O_3 and found that GB has a lower MAE than LR every time when equivalent conditions are used (Number of variables and device). deSouza et al., 2022 [70] also utilise both GB and LR but do not use equivalent conditions, making comparison difficult. However, when looking at the MAE across all studies for NO_2 (Figure 6a), LR has a lower average MAE than GB, suggesting LR performs better. However, it is one of the worst performers in nearly all studies.

A similar trend can be seen in $PM_{2.5}$ when comparing LR and RF. In Ko et al., 2024, RF achieved a lower MAE than LR in every situation that did not use NO_2 as a feature. The results are more mixed when including NO_2 as a feature. Given there is no reason for NO_2 levels to physically affect $PM_{2.5}$ measurement based on first principles, it is instead likely acting as a proxy for environmental conditions such as traffic. This potentially leads to overfitting in the RF model, suggesting one possible reason why it does not perform as well. Kureshi et al., 2022 [79] also compare RF and LR models for correcting $PM_{2.5}$, with RF consistently achieving a lower MAE than LR. Several other studies also compared LR and RF but only reported RMSE and/or r^2, making them difficult to compare. However, the average MAE for RF is higher than that for LR when correcting $PM_{2.5}$, despite some RF models achieving significantly lower MAEs than the lowest LR MAE.

The reason for this inconsistency between individual results and the aggregated results is due to the wide range of varying experimental conditions used in these studies. For example, some may choose to calibrate with very-high-frequency measurements (e.g., 1 min) to obtain the widest range of conditions possible, increasing the data density at the expense of increased noise [15,27,29]. However, some calibrate with low-frequency measurements (e.g., 1 h). This is carried out to reduce noise in the measurement or because the collocated reference-grade monitor outputs hourly averages, or both [40,47,57,69]. The campaign length is one of the biggest variances between publications, ranging from 15 h [68] to 2 years [73,82]. As seasonal variations in temperature, humidity and particle composition can all affect the response of the sensor, shorter campaigns only show a snapshot of the capabilities of a calibration algorithm. Mahajan et al.'s [68] study may have achieved good $PM_{2.5}$ correction results utilising only the raw OPC measurement as a feature for the LR and RF correction models, but it is unlikely that environmental conditions would have varied significantly in the 15 h the study ran for. A longer study that includes a wide range of environmental changes in factors including T and RH may provide more insights into the biases from these features.

The number of sensor systems used differs between studies. Most use one sensor system [12,15,16,25,29,40,57,69]. Smith et al., 2018 and 2019 [15,16] utilised one sensor

system, but it contained clusters of sensors within it. Malings et al. [41] used the most sensors, utilising 62 in total [41].

As the environmental conditions in which these calibrations are performed vary quite drastically, the reported errors often do not provide enough information. Smith et al., 2019 reported not only the RMSE scores for the four models (ULR, BLR, BRT and GP) they tested but also assessed the models by splitting the test set into four quartiles based on the reference concentration and reporting the RMSE for each. The RMSE scores between quartiles varied drastically, with ULR (presented as SLR in the publication) achieving scores of 11.7 ppb, 13.3 ppb, 30.8 ppb and 17.4 ppb in ascending order of quartile. BRT had a higher variation, achieving 5.6 ppb, 15.2 ppb, 72.5 ppb and 120 ppb in ascending order of quartile. This shows that even when the same models using the same type of sensing components are applied to different environments, the performance of the model can vary substantially. Even when normalising the errors using the mean reference concentration, the variances were still present. This could partially be due to the use of RMSE, which varies with three characteristics of the errors (distribution of error magnitudes, square root of number of errors and average error magnitude), instead of MAE, which varies with one (average error magnitude) [94].

Several researchers have realised this and either reported both standard and normalised errors or just normalised ones [16,29,47]. Smith et al. [16] went further and included separate errors for different concentration quartiles to show how the calibration techniques performed at different concentrations.

5.5. Are We Calibrating Measurements or Models?

Concerns were raised in Hagler et al., 2018 [36] about the use of parameters that have no proven influence on the measurement, including error and bias. Those concerns are not shared by all, with many publications using inputs which have no proven direct impact on the measurement. An example of this is Sayahi et al., 2020 [8], who used solar radiation as an input to predict the O_3 concentration, along with a metal oxide (MO_x) sensor and temperature. The use of solar radiation is a problem, as it has no proven direct impact on the measurement itself, but it was chosen, as it improved the measurement output when included. By including parameters that correlate with the measurement but do not necessarily cause a bias in it, one cannot ensure that that specific parameter will continue to affect the measurement in the same positive manner. Solar radiation is likely to make a significant contribution to the measurement, as the rate of generation of O_3 at the ground level is related to the intensity of incident UV light [100,101], but this affects the concentration of O_3 in the air, not the measurement of O_3 by the sensor. Using this parameter as a predictor blurs the line between a measurement and a model.

Si et al., 2020 utilised reference measurements as input parameters for their calibrations as well as for their ground truth measurements, e.g., CO reference measurements used in their low-cost $PM_{2.5}$ calibration [12]. This was carried out as these measurements were algorithmically determined to improve the measurement at that time in that location. This is incredibly impractical, as the large-scale deployment of low-cost sensor systems using this calibration technique would need to be co-located with reference instruments, defying the purpose of low-cost systems in the first place.

By using parameters that algorithmically improve the measurement at the time of calibration in the location they were tested in, one may reduce the uncertainty in the calibrated measurement. However, by using parameters based exclusively on their correlation with the measurement, and not on how it may physically affect it, there is no assurance the effects of these additional inputs will not vary over time. A short-term improvement in results could lead to an uncharacterised uncertainty in the future.

6. Conclusions

A range of sensing components are used in low-cost AQ sensor systems, all with their own advantages and drawbacks. The most prominent sensing component in recent

literature, electrochemical sensors, have proven suitable for ambient AQ monitoring despite being biased by multiple external factors. The most common external biases are caused by changes in temperature and relative humidity, though cross-interferences with other trace gases can also affect the measurements—most notably, a 100% cross-sensitivity between NO_2 and O_3.

Optical particle counters were the only low-cost particle sensor type encountered in the literature. When calibrated in the same environment they are deployed in, they tend to perform very well compared to reference-grade instruments. This performance quickly degrades if the environment they are monitoring varies from where they were calibrated, either by being moved a long distance or by seasonal variations in particulate compositions.

Two distinct calibration environments are identified in the literature, either in a controlled environmental chamber or deployed in the field. However, nearly all calibrations found in the literature use field calibrations due to the more expensive nature of laboratory calibrations, as well as their difficulty in reproducing the wide range of conditions encountered in the field. Field calibrations allow for exposing the sensor to a wide range of training variables, but it remains important to ensure the sensor is calibrated in a similar environment to the one it will be deployed in; otherwise, the calibration may perform poorly.

Despite a wide range of calibration techniques being tested in the literature with a range of different input variables, direct comparisons between publications is difficult, as there is no standardised protocol for calibration approaches. There are huge variations in campaign lengths, measurement averaging windows and environmental conditions that make comparisons very challenging. Of particular note are the reporting of unnormalised errors outside of the context of the environment in which they were calculated. To combat this, several papers report normalised errors, which help make comparisons easier, but until a standardised reporting method is developed, similar to high-cost reference instrumentation, the comparison of different techniques and systems will continue to be difficult.

Due to the lack of standardisation, the studies are closer to disparate individual activities rather than part of a larger picture. Comparisons between them are unfortunately limited due to the varying lengths of studies, the model selection methodology such as varying sizes of train:test splits and the error metrics used. This is particularly highlighted with the use of unnormalised metrics, potentially making the reported errors as much a function of the environment in which they were calibrated as the performance of the calibrations themselves. The use of metrics like r^2 and RMSE also hinders comparisons, as variations in the number of observations used to validate calibrations, the number of independent variables used in the calibration and the magnitudes of the errors all affect the reported performance, not just the error of the calibration. This unfortunately limits any conclusions being drawn by comparison, making recommendations of technique and correction variables different without a comprehensive, standardised study.

Author Contributions: Conceptualisation, I.H., N.A.M. and P.K.; Methodology, I.H.; Study collection, filtering and review, I.H.; software and visualisations, I.H.; validation, V.F., M.K. and P.K.; writing—original draft preparation, I.H. and P.K.; writing—review and editing, I.H., N.A.M., V.F., M.K. and P.K.; supervision, N.A.M., V.F. and P.K.; funding acquisition, N.A.M. and P.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the University of Surrey, the National Physical Laboratory and EPSRC via a Ph.D. studentship (2018–2021; grant reference #2126120), awarded through the University Research Scholarship scheme, to support I.H.'s PhD research. P.K. acknowledges the support received through the UKRI (NERC, EPSRC, AHRC)-funded RECLAIM Network Plus (EP/W034034/1) and the NERC-funded GreenCities (NE/X002799/1) projects.

Data Availability Statement: All individual results collected for this review are available upon request.

Acknowledgments: We would like to express our gratitude to the staff at the University of Surrey and the National Physical Laboratory for their support during I.H.'s PhD.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Gumy, S.; Prüss-Üstün, A. *Ambient Air Pollution: A Global Assessment of Exposure and Burden of Disease*; World Health Organisation: Geneva, Switzerland, 2016.
2. McKercher, G.R.; Salmond, J.A.; Vanos, J.K. Characteristics and applications of small, portable gaseous air pollution monitors. *Environ. Pollut.* **2016**, *223*, 102–110. [CrossRef]
3. Dobre, A.; Arnold, S.J.; Smalley, R.J.; Boddy, J.W.D.; Barlow, J.F.; Tomlin, A.S.; Belcher, S.E. Flow field measurements in the proximity of an urban intersection in london, UK. *Atmos. Environ.* **2005**, *39*, 4647–4657. [CrossRef]
4. Yi, W.Y.; Lo, K.M.; Mak, T.; Leung, K.S.; Leung, Y.; Meng, M.L. A survey of wireless sensor network based air pollution monitoring systems. *Sensors* **2015**, *15*, 31392–31427. [CrossRef] [PubMed]
5. Kim, J.; Shusterman, A.A.; Lieschke, K.J.; Newman, C.; Cohen, R.C. The BErkeley atmospheric CO_2 observation network: Field calibration and evaluation of low–cost air quality sensors. *Atmos. Meas. Tech. Discuss.* **2018**, *11*, 1937–1946. [CrossRef]
6. Weissert, L.; Alberti, K.; Miskell, G.; Pattinson, W.; Salmond, J.; Henshaw, G.; Williams, D.E. Low–cost sensors and microscale land use regression: Data fusion to resolve air quality variations with high spatial and temporal resolution. *Atmos. Environ.* **2019**, *213*, 285–295. [CrossRef]
7. Popoola OA, M.; Carruthers, D.; Lad, C.; Bright, V.B.; Mead, M.I.; Stettler, M.E.J.; Saffell, J.R.; Jones, R.L. Use of networks of low cost air quality sensors to quantify air quality in urban settings. *Atmos. Environ.* **2018**, *194*, 58–70. [CrossRef]
8. Sayahi, T.; Garff, A.; Quah, T.; Lê, K.; Becnel, T.; Powell, K.M.; Gaillardon, P.-E.; Butterfield, A.E.; Kelly, K.E. Long term calibration models to estimate ozone concentrations with a metal oxide sensor. *Environ. Pollut.* **2020**, *267*, 115363. [CrossRef] [PubMed]
9. Sayahi, T.; Butterfield, A.; Kelly, K.E. Long–term field evaluation of the plantower PMS low–cost particulate matter sensors. *Environ. Pollut.* **2018**, *245*, 932–940. [CrossRef] [PubMed]
10. Bulot, F.M.J.; Johnston, S.J.; Basford, P.J.; Easton, N.H.C.; Apetroaie-Cristea, M.; Foster, G.L.; Morris, A.K.R.; Cox, S.J.; Loxham, M. Long-term field comparison of multiple low-cost particulate matter sensors in an outdoor urban environment. *Sci. Rep.* **2019**, *9*, 7497. [CrossRef] [PubMed]
11. Johnston, S.J.; Basford, P.J.; Bulot, F.M.J.; Apetroaie-Cristea, M.; Easton, N.H.C.; Davenport, C.; Foster, G.L.; Loxham, M.; Morris, A.K.R.; Cox, S.J. City scale particulate matter monitoring using LoRaWAN based air quality IOT devices. *Sensors* **2019**, *19*, 209. [CrossRef] [PubMed]
12. Si, M.; Xiong, Y.; Du, S.; Du, K. Evaluation and calibration of a low–cost particle sensor in ambient conditions using machine–learning methods. *Atmos. Meas. Tech.* **2020**, *13*, 1693–1707. [CrossRef]
13. Baron, R.; Saffell, J. Amperometric gas sensors as a low cost emerging technology platform for air quality monitoring applications: A review. *ACS Sens.* **2017**, *2*, 1553–1566. [CrossRef] [PubMed]
14. Mijling, B.; Jiang, Q.; de Jonge, D.; Bocconi, S. Field calibration of electrochemical NO_2 sensors in a citizen science context. *Atmos. Meas. Tech.* **2018**, *11*, 1297–1312. [CrossRef]
15. Smith, K.R.; Edwards, P.M.; Evans, M.J.; Lee, J.D.; Shaw, M.D.; Squires, F.; Wilde, S.; Lewis, A.C. Clustering approaches to improve the performance of low cost air pollution sensors. *Faraday Discuss.* **2017**, *200*, 621–637. [CrossRef] [PubMed]
16. Smith, K.R.; Edwards, P.M.; Ivatt, P.D.; Lee, J.D.; Squires, F.; Dai, C.; Peltier, R.E.; Evans, M.J.; Sun, Y.; Lewis, A.C. An improved low–power measurement of ambient NO_2 and O_3 combining electrochemical sensor clusters and machine learning. *Atmos. Meas. Tech.* **2019**, *12*, 1325–1336. [CrossRef]
17. Sun, L.; Westerdahl, D.; Ning, Z. Development and evaluation of a novel and cost–effective approach for low–cost NO_2 sensor drift correction. *Sensors* **2017**, *17*, 1916. [CrossRef]
18. Miskell, G.A.; Salmond, J.; Williams, D.E. Solution to the problem of calibration of low–cost air quality measurement sensors in networks. *ACS Sens.* **2018**, *3*, 832–843. [CrossRef] [PubMed]
19. 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. The changing paradigm of air pollution monitoring. *Environ. Sci. Technol.* **2013**, *43*, 11369–11377. [CrossRef] [PubMed]
20. Tian, B.; Hou, K.M.; Diao, X.; Shi, H.; Zhou, H.; Wang, W. Environment–adaptive calibration system for outdoor low-cost electrochemical gas sensors. *IEEE Access* **2019**, *7*, 62592–62605. [CrossRef]
21. Spinelle, L.; Gerboles, M.; Aleixandre, M. Performance evaluation of amperometric sensors for the monitoring of O_3 and NO_2 in ambient air at ppb level. *Procedia Eng.* **2015**, *120*, 480–483. [CrossRef]
22. Afshar-Mohajer, N.; Zuidema, C.; Sousan, S.; Hallett, L.; Tatum, M.; Rule, A.M.; Thomas, G.; Peters, T.M.; Koehler, K. Evaluation of low–cost electro–chemical sensors for environmental monitoring of ozone, nitrogen dioxide, and carbon monoxide. *J. Occup. Environ. Hyg.* **2018**, *15*, 87–98. [CrossRef] [PubMed]
23. Hossain, M.; Saffell, J.; Baron, R. Differentiating NO_2 and O_3 at low cost air quality amperometric gas sensors. *ACS Sens.* **2016**, *1*, 1291–1294. [CrossRef]
24. Cordero, J.M.; Borge, R.; Narros, A. Using statistical methods to carry out in field calibrations of low cost air quality sensors. *Sens. Actuators B Chem.* **2018**, *267*, 245–254. [CrossRef]

25. Wei, P.; Sun, L.; Anand, A.; Zhang, Q.; Huixin, Z.; Deng, Z.; Wang, Y.; Ning, Z. Development and evaluation of a robust temperature sensitive algorithm for long term NO_2 gas sensor network data correction. *Atmos. Environ.* **2020**, *230*, 117509. [CrossRef]
26. Ouyang, B. First–principles algorithm for air quality electrochemical gas sensors. *ACS Sens.* **2020**, *5*, 2742–2746. [CrossRef] [PubMed]
27. Hagan, D.H.; Isaacman-VanWertz, G.; Franklin, J.P.; Wallace, L.M.M.; Kocar, B.D.; Heald, C.L.; Kroll, J.H. Calibration and assessment of electrochemical air quality sensors by co-location with regulatory-grade instruments. *Atmos. Meas. Tech.* **2018**, *11*, 315–328. [CrossRef]
28. Han, P.; Mei, H.; Liu, D.; Zeng, N.; Tang, X.; Wang, Y.; Pan, Y. Calibrations of low-cost air pollution monitoring sensors for CO, NO_2, O_3, and SO_2. *Sensors* **2021**, *21*, 256. [CrossRef]
29. Liang, Y.; Wu, C.; Jiang, S.; Li, Y.J.; Wu, D.; Li, M.; Cheng, P.; Yang, W.; Cheng, C.; Li, L.; et al. Field comparison of electrochemical gas sensor data correction algorithms for ambient air measurements. *Sens. Actuators B Chem.* **2021**, *327*, 128897. [CrossRef]
30. Li, H.; Zhu, Y.; Zhao, Y.; Chen, T.; Jiang, Y.; Shan, Y.; Liu, Y.; Mu, J.; Yin, X.; Wu, D.; et al. Evaluation of the performance of low–cost air quality sensors at a high mountain station with complex meteorological conditions. *Atmosphere* **2020**, *11*, 212. [CrossRef]
31. Spinelle, L.; Gerboles, M.; Villani, M.G.; Aleixandre, M.; Bonavitacola, F. Field calibration of a cluster of low-cost available sensors for air quality monitoring. Part a: Ozone and nitrogen dioxide. *Sens. Actuators B Chem.* **2015**, *215*, 249–257. [CrossRef]
32. Szulczyński, B.; Gębicki, J. Currently commercially available chemical sensors employed for detection of volatile organic compounds in outdoor and indoor air. *Environments* **2017**, *4*, 21. [CrossRef]
33. Castell, N.; Viana, M.; Minguillón, M.C.; Guerreiro, C.; Querol, X. *Real–World Application of New Sensor Technologies for Air Quality Monitoring*; ETC/ACM: Bilthoven, The Netherlands, 2013.
34. Borrego, C.; Costa, A.m.; Ginja, J.; Amorim, M.; Coutinho, M.; Karatzas, K.; Sioumis, T.; Katsifarakis, N.; Konstantinidis, K.; De Vito, S.; et al. Assessment of air quality microsensors versus reference methods: The EuNetAir joint exercise. *Atmos. Environ.* **2016**, *147*, 246–263. [CrossRef]
35. Phala, K.S.E.; Kumar, A.; Hancke, G.P. Air quality monitoring system based on ISO/IEC/IEEE 21451 standards. *IEEE Sens. J.* **2016**, *16*, 5037–5045. [CrossRef]
36. Hagler, G.S.W.; Williams, R.; Papapostolou, V.; Polidori, A. Air quality sensors and data adjustment algorithms: When is it no longer a measurement? *Environ. Sci. Technol.* **2018**, *52*, 5530–5531. [CrossRef] [PubMed]
37. Kizel, F.; Etzion, Y.; Shafran-Nathan, R.; Levy, I.; Fishbain, B.; Bartonova, A.; Broday, D.M. Node-to-node field calibration of wireless distributed air pollution sensor network. *Environ. Pollut.* **2018**, *233*, 900–909. [CrossRef] [PubMed]
38. Zimmerman, N.; Presto, A.A.; Kumar, S.P.N.; Gu, J.; Hauryliuk, A.; Robinson, E.S.; Robinson, A.L.; Subramanian, R. A machine learning calibration model using random forests to improve sensor performance for lower–cost air quality monitoring. *Atmos. Meas. Tech.* **2018**, *11*, 291–313. [CrossRef]
39. Curto, A.; Donaire-Gonzalez, D.; Barrera-Gómez, J.; Marshall, J.; Nieuwenhuijsen, M.; Wellenius, G.; Tonne, C. Performance of low–cost monitors to assess household air pollution. *Environ. Res.* **2018**, *163*, 53–63. [CrossRef]
40. Shindler, L. Development of a low–cost sensing platform for air quality monitoring: Application in the city of rome. *Environ. Technol.* **2019**, *42*, 618–631. [CrossRef] [PubMed]
41. Malings, C.; Tanzer, R.; Hauryliuk, A.; Kumar, S.P.N.; Zimmerman, N.; Kara, L.B.; Presto, A.A.; Subramanian, R. Development of a general calibration model and long–term performance evaluation of low–cost sensors for air pollutant gas monitoring. *Atmos. Meas. Tech.* **2019**, *12*, 903–920. [CrossRef]
42. Suriano, D.; Cassano, G.; Penza, M. Design and development of a flexible, plug–and–play, cost–effective tool for on–field evaluation of gas sensors. *J. Sens.* **2020**, *2020*, 8812025. [CrossRef]
43. Gerboles, M.; Buzica, D. *Evaluation of Micro–Sensors to Monitor Ozone in Ambient Air*; European Communities: Luxembourg, 2009.
44. Peterson, P.J.D.; Aujla, A.; Grant, K.H.; Brundle, A.G.; Thompson, M.R.; Hey, J.V.; Leigh, R.J. Practical use of metal oxide semiconductor gas sensors for measuring nitrogen dioxide and ozone in urban environments. *Sensors* **2017**, *17*, 1653. [CrossRef] [PubMed]
45. Albert, K.J.; Lewis, N.S.; Schauer, C.L.; Sotzing, G.A.; Stitzel, S.E.; Vaid, T.P.; Walt, D.R. Cross–reactive chemical sensor arrays. *Chem. Rev.* **2000**, *100*, 2595–2626. [CrossRef] [PubMed]
46. Lewis, A.C.; Lee, J.D.; Edwards, P.M.; Shaw, M.D.; Evans, M.J.; Moller, S.J.; Smith, K.R.; Buckley, J.W.; Ellis, M.; Gillot, S.R.; et al. Evaluating the performance of low cost chemical sensors for air pollution research. *Faraday Discuss.* **2016**, *189*, 85–103. [CrossRef] [PubMed]
47. Topalovi, D.B.; Davidovi, M.D.; Jovanovi, M.; Bartonova, A.; Ristovski, Z.; Jovaevi-Stojanovi, M. 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. *Atmos. Environ.* **2019**, *213*, 640–658. [CrossRef]
48. Buehler, C.; Xiong, F.; Zamora, M.L.; Skog, K.M.; Kohrman-Glaser, J.; Colton, S.; McNamara, M.; Ryan, K.; Redlich, C.; Bartos, M.; et al. Stationary and portable multipollutant monitors for high–spatiotemporal–resolution air quality studies including online calibration. *Atmos. Meas. Tech.* **2021**, *14*, 995–1013. [CrossRef]
49. Zoest, V.; Osei, F.B.; Stein, A.; Hoek, G. Spatio-temporal regression kriging for modelling urban NO_2 concentrations. *Atmos. Environ.* **2019**, *210*, 66–75.
50. Kendler, S.; Zuck, A. The challenges of pronlonged gas sensing in the modern urban environment. *Sensors* **2021**, *20*, 189.

51. Northcross, A.L.; Edwards, R.J.; Johnson, M.A.; Wang, Z.-M.; Zhu, K.; Allen, T.; Smith, K.R. A low–cost particle counter as a realtime fine–particle mass monitor. *Environ. Sci. Process. Impacts* **2012**, *15*, 433–439. [CrossRef] [PubMed]
52. Crilley, L.R.; Marvin, S.; Ryan, P.; Kramer, L.J.; Robin, P.; Stuart, Y.; Lewis, A.C.; Pope, F.D. Evalutation of a low–cost optical particle counter (alphasense OPC–N_2) for ambient air monitoring. *Atmos. Meas. Tech.* **2018**, *11*, 709–720. [CrossRef]
53. Alfano, B.; Barretta, L.; Del Giudice, A.; De Vito, S.; Di Francia, G.; Esposito, E.; Formisano, F.; Massera, E.; Miglietta, M.L.; Polichetti, T. A review of low–cost particulate matter sensors from the developers' perspectives. *Sensors* **2020**, *20*, 6819. [CrossRef]
54. Njallson, T.; Novosselov, I. Design and optimization of a compact low–cost optical particle sizer. *J. Aerosol Sci.* **2018**, *119*, 1–12. [CrossRef]
55. Johnson, K.K.; Bergin, M.H.; Russell, A.G.; Hagler GS, W. Field test of several low–cost particulate matter sensors in high and low concentration urban environments. *Aerosol Air Qual. Res.* **2018**, *18*, 565–578. [CrossRef] [PubMed]
56. Hua, J.; Zhang, Y.; de Foy, B.; Mei, X.; Shang, J.; Zhang, Y.; Sulaymon, I.D.; Zhou, D. Improved $PM_{2.5}$ concentration estimates from low–cost sensors using calibration models categorized by relative humidity. *Aerosol Sci. Technol.* **2021**, *55*, 600–613. [CrossRef]
57. Liu, X.; Zhao, Q.; Zhu, S.; Peng, W.; Yu, L. An experimental application of laser–scattering sensor to estimate the traffic–induced $PM_{2.5}$ in beijing. *Environ. Monit. Assess.* **2020**, *192*, 450. [CrossRef] [PubMed]
58. Yuval; Molho, H.M.; Zivan, O.; Broday, D.M.; Raz, R. Application of a sensor network of low cost optical particle counters for assessing the impact of quarry emission on its vicinity. *Atmos. Environ.* **2019**, *211*, 29–37.
59. Samad, A.; Mimiaga FE, M.; Laquai, B.; Vogt, U. Investigating a low–cost dryer designed for low–cost PM sensors measuring ambient air quality. *Sensors* **2021**, *21*, 804. [CrossRef] [PubMed]
60. Borghi, F.; Spinazzè, A.; Campagnolo, D.; Rovelli, S.; Cattaneo, A.; Cavallo, D.M. Precision and accuracy of a direct–reading miniaturized monitor in $PM_{2.5}$ exposure assessment. *Sensors* **2018**, *18*, 3089. [CrossRef] [PubMed]
61. Spinazzè, A.; Fanti, G.; Borghi, F.; Del Buono, L.; Campagnolo, D.; Rovelli, S.; Cattaneo, A.; Cavallo, D.M. Field comparison of instruments for exposure assessment of airborne ultrafine particles and particulate matter. *Environ. Sci. Technol.* **2017**, *154*, 274–284. [CrossRef]
62. Sally Liu, L.J.; Slaughter, J.C.; Larson, T.V. Comparison of light scattering devices and impactors for particulate measurements in indoor, outdoor, and personal environments. *Environ. Sci. Technol.* **2002**, *36*, 2977–2986.
63. Collingwood, S.; Zmoos, J.; Pahler, L.; Wong, B.; Sleeth, D.; Handy, R. Investigating measurement variation of modified low–cost particle sensors. *J. Aerosol Sci.* **2019**, *135*, 21–32. [CrossRef]
64. Piedrahita, R.; Xiang, Y.; Masson, N.; Ortega, J.; Collier, A.; Jiang, Y.; Li, K.; Dick, R.P.; Lv, Q.; Hannigan, M.; et al. The next generation of low–cost personal air quality sensors for quantitative exposure monitoring. *Atmos. Meas. Tech.* **2014**, *7*, 3325–3336. [CrossRef]
65. Zauli-Sajani, S.; Marchesi, S.; Pironi, C.; Barbieri, C.; Poluzzi, V.; Colacci, A. Assessment of air quality sensor system performance after relocation. *Atmos. Pollut. Res.* **2021**, *12*, 282–291. [CrossRef]
66. Montgomery, D.C.; Peck, E.A.; Vining, G.G. *Introduction to Linear Regression Analysis*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2021; pp. 12–68.
67. Weisberg, S. *Applied Linear Regression*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2005; pp. 19–46.
68. Mahajan, S.; Kumar, P. Evaluation of low-cost sensors for quantitative personal exposure monitoring. *Sustain. Cities Soc.* **2020**, *57*, 102076. [CrossRef]
69. Astudillo, G.D.; Garza–Castañon, L.E.; Minchala Avila, L.I. Design and evaluation of a reliable low–cost atmospheric pollution station in urban environment. *IEEE Access* **2020**, *8*, 51129–51144. [CrossRef]
70. Desouza, P.; Kahn, R.; Stockman, T.; Obermann, W.; Crawford, B.; Wang, A.; Crooks, J.; Li, J.; Kinney, P. Calibrating networks of low-cost air quality sensors. *Atmos. Meas. Tech.* **2022**, *15*, 6309–6328. [CrossRef]
71. Cavaliere, A.; Brilli, L.; Andreini, B.P.; Carotenuto, F.; Gioli, B.; Giordano, T.; Stefanelli, M.; Vagnoli, C.; Zaldei, A.; Gualtieri, G. Development of low-cost air quality stations for next-generation monitoring networks: Calibration and validation of NO_2 and O_3 sensors. *Atmos. Meas. Tech.* **2023**, *16*, 4723–4740. [CrossRef]
72. Agrawal, D.; Saini, A.K.; Rai, A.C.; Kala, P. In chamber calibration and performance evaluation of air quality low-cost sensors. *Atmos. Pollut. Res.* **2024**, *15*, 102299. [CrossRef]
73. Ko, K.; Cho, S.; Rao, R.R. Evaluation of calibration performance of a low-cost particulate matter sensor using collocated and distant NO_2. *Atmos. Meas. Tech.* **2024**, *17*, 3303–3322. [CrossRef]
74. Biau, G.; Scornet, E. A random forest guided tour. *TEST* **2016**, *25*, 5–32. [CrossRef]
75. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
76. Ionascu, M.E.; Castell, N.; Boncalo, O.; Marcu, M. Calibration of CO, NO_2, and O_3 using airify: A low-cost sensor cluster for air quality monitoring. *Sensors* **2021**, *21*, 7977. [CrossRef]
77. Scott, A.; Wild, C. Transformations and r. *Am. Stat.* **1991**, *45*, 127–129. [CrossRef]
78. Draper, N.R. The box-wetz criterion versus R. *J. R. Stat. Soc. Ser. A (Gen.)* **1984**, *147*, 100–103. [CrossRef]
79. Kureshi, R.R.; Mishra, B.K.; Thakker, D.; John, R.; Walker, A.; Simpson, S.; Thakkar, N.; Wante, A.K. Data-driven techniques for low-cost sensor selection and calibration for the use case of air quality monitoring. *Sensors* **2022**, *22*, 1093. [CrossRef]
80. Koziel, S.; Pietrenko-Dabrowska, A.; Wojcikowski, M.; Pankiewicz, B. On memory-based precise calibration of cost-efficient NO2 sensor using artificial intelligence and global response correction. *Knowl.-Based Syst.* **2024**, *290*, 111564. [CrossRef]

81. Esposito, E.; De Vito, S.; Salvato, M.; Bright, V.; Jones, R.; Popoola, O. Dynamic neural network architectures for on field stochastic calibration of indicative low cost air quality sensing systems. *Sens. Actuators B-Chem.* **2016**, *231*, 701–713. [CrossRef]
82. Kang, J.; Choi, K. Calibration methods for low-cost particulate matter sensors considering seasonal variability. *Sensors* **2024**, *24*, 3023. [CrossRef]
83. Hagan, M.T.; Demuth, H.B.; Beale, M.H.; De Jesús, O. *Neural Network Design*; Martin Hagan: San Francisco, CA, USA, 2014; pp. 14–35.
84. Galushkin, A.I. *Neural Networks Theory*; Springer: Berlin/Heidelberg, Germany, 2007; pp. 1–32.
85. Hristev, R.M. *The ANN Book*; Self-Published; 1998; pp. 3–8.
86. Lever, J.; Krzywinski, M.; Altman, N. Points of significance: Model selection and overfitting. *Nat. Methods* **2016**, *13*, 703–705. [CrossRef]
87. Ali, S.; Alam, F.; Arif, K.M.; Potgieter, J. Low-cost CO sensor calibration using one dimensional convolutional neural network. *Sensors* **2023**, *23*, 854. [CrossRef]
88. Alhasa, K.M.; Nadzir, M.S.M.; Olalekan, P.; Latif, M.T.; Yusup, Y.; Faruque, M.R.I.; Ahamad, F.; Hamid, H.H.A.; Aiyub, K.; Ali, S.H.M.; et al. Calibration model of a low-cost air quality sensor using an adaptive neuro-fuzzy inference system. *Sensors* **2018**, *18*, 4380. [CrossRef] [PubMed]
89. Anderson-Sprecher, R. Model comparisons and r. *Am. Stat.* **1994**, *48*, 113–117. [CrossRef]
90. Fisher, R.A., III. The influence of rainfall on the yield of wheat at rothamsted. Philosophical Transactions of the Royal Society of London. *Ser. B Contain. Pap. A Biol. Character* **1924**, *213*, 89–142.
91. Healy, M. The use of r 2 as a measure of goodness of fit. *J. R. Stat. Soc. Ser. A Stat. Soc.* **1984**, *147*, 608–609. [CrossRef]
92. Kvålseth, T.O. Cautionary note about r. *Am. Stat.* **1985**, *39*, 279–285. [CrossRef]
93. Willett, J.B.; Singer, J.D. Another cautionary note about r 2: Its use in weighted least-squares regression analysis. *Am. Stat.* **1988**, *42*, 236–238. [CrossRef]
94. Willmott, C.J.; Matsuura, K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Clim. Res.* **2005**, *30*, 79–82. [CrossRef]
95. Willmott, C.J.; Ackleson, S.G.; Davis, R.E.; Feddema, J.J.; Klink, K.M.; Legates, D.R.; O'Donnell, J.; Rowe, C.M. Statistics for the evaluation and comparison of models. *J. Geophys. Res.* **1985**, *90*, 8995–9005. [CrossRef]
96. Swami, A.; Jain, R. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
97. Veiga, T.; Munch-Ellingsen, A.; Papastergiopoulos, C.; Tzovaras, D.; Kalamaras, I.; Bach, K.; Votis, K.; Akselsen, S. From a low-cost air quality sensor network to decision support services: Steps towards data calibration and service development. *Sensors* **2021**, *21*, 3190. [CrossRef] [PubMed]
98. Bush, T.; Papaioannou, N.; Leach, F.; Pope, F.D.; Singh, A.; Thomas, G.N.; Stacey, B.; Bartington, S. Machine learning techniques to improve the field performance of low-cost air quality sensors. *Atmos. Meas. Tech.* **2022**, *15*, 3261–3278. [CrossRef]
99. Haunsperger, D.B.; Saari, D.G. The lack of consistency for statistical decision procedures. *Am. Stat.* **1991**, *45*, 252–255. [CrossRef]
100. Monks, P.S.; Archibald, A.T.; Colette, A.; Cooper, O.; Coyle, M.; Derwent, R.; Fowler, D.; Granier, C.; Law, K.S.; Mills, G.E.; et al. Tropospheric ozone and its precursors from the urban to the global scale from air quality to short–lived climate forcer. *Atmos. Chem. Phys.* **2015**, *15*, 8889–8973. [CrossRef]
101. Lightfoot, P.; Cox, R.; Crowley, J.; Destriau, M.; Hayman, G.; Jenkin, M.; Moortgat, G.; Zabel, F. Organic peroxy radicals: Kinetics, spectroscopy and tropospheric chemistry. *Atmos. Environ. Part A Gen. Top.* **1992**, *26*, 1805–1961. [CrossRef]