https://doi.org/10.5194/amt-2024-130
Preprint. Discussion started: 15 August 2024

Atmospheric Measurement Techniques Discussions

Impact and Optimization of Calibration Conditions for Air Quality Sensors in the Long-term Field Monitoring

Han Mei[1], Peng Wei[2*], Meisam Ahmadi Ghadikolaei[1], Nirmal Kumar Gali[1], Ya Wang[1], Zhi Ning[1,*]

[1]Division of Environment and Sustainability, The Hong Kong University of Science and Technology, Hong Kong, China
[2]College of Geography and Environment, Shandong Normal University, Jinan, China

Correspondence to: Zhi Ning (zhining@ust.hk), Peng Wei (pengwei@sdnu.edu.cn)

Abstract. The rapid expansion of low-cost sensor networks for air quality monitoring necessitates rigorous calibration to ensure data accuracy. Despite numerous published field calibration studies, a universal and comprehensive assessment of factors affecting sensor calibration remains elusive, leading to potential discrepancies in data quality across different networks. To address these challenges, this study deployed eight sensor-based monitors equipped with electrochemical sensors for NO_2, NO, CO, and O_3 measurement in strategically chosen locations within Hong Kong, Macau, and Shanghai, covering a wide range of climatic conditions: Hong Kong's subtropical climate, Macau's similar yet distinct urban environment, and Shanghai's more variable climate. This strategic deployment ensured that the sensors' performance and calibration processes were tested across diverse atmospheric conditions. Each monitor employed a patented dynamic baseline tracking method for the gas sensors, which isolates the concentration signals from temperature and humidity effects, enhancing the sensors' accuracy and reliability. The tests, which involved evaluating the validation performance by analyzing randomly selected calibration sample subsets ranging from 1 to 15 days, indicated that the length of the calibration period, pollutant concentration range, and time averaging period are pivotal for sensor calibration quality. We determined that a 5–7 days calibration period minimizes calibration coefficient errors, and a wider concentration range improves the validation R^2 values for all sensors, suggesting the necessity of setting specific concentration range thresholds. Moreover, a time averaging period of at least 5 minutes for data with 1-minute resolution was recommended to enable optimal calibration in field operation. This study emphasizes the need for a comprehensive calibration assessment and the importance of considering environmental variability in sensor calibration condition. These findings offer methodological guidance for the calibration of other sensor types, providing a reference for future research in the field of sensor calibration.

1 Introduction

Rapid advancements in low-cost air sensor technology have led to a significant increase in their applications across various fields. These sensors offer a promising and cost-effective solution for monitoring air pollution at finer spatial scales and in novel locations compared to traditional monitoring methodologies. This has resulted in a growing demand for high-quality sensor data. Calibration is an indispensable component of the air sensor operational paradigm, pivotal for securing accurate

https://doi.org/10.5194/amt-2024-130
Preprint. Discussion started: 15 August 2024


Atmospheric Measurement Techniques Discussions

and dependable data. By establishing a relationship between the raw sensor output and the corresponding reference measurement, calibration enhances the accuracy and precision of sensor data.

Common calibration methods include multi-point calibration with standard gases, controlled chamber calibration (Papapostolou et al., 2017; Sousan et al., 2016), on-site probe gas calibrations, and field side-by-side calibration (Bisignano et al., 2022; Holstius et al., 2014; Spinelle et al., 2015, 2017). Multi-point calibration allows sensors to undergo multiple concentration points and zero checks in the laboratory, while controlled chamber calibration involves creating a laboratory chamber to simulate ambient conditions with variable concentrations, temperature, and humidity levels (Papapostolou et al., 2017). Additionally, on-site calibration through probe gas calibration is another approach where the gas sensor is calibrated directly in the field using probe gases of known standard gas concentrations. As all three methods are laboratory-based methods or rely on standard gas, they inherently possess constraints and may not fully capture the intricate interactions of multiple pollutants and environmental factors encountered in situ. This raises questions regarding the representativeness of laboratory setup in relation to actual monitoring locales and the application of calibration results obtained through standard gas to field conditions (Castell et al., 2017). The limitations of the laboratory-based or standard gas methods underscore the advantages of an alternative: the side-by-side calibration, which involves the co-locating sensor systems with reference analyzers in real-world environmental settings for a designated duration. This approach leverages the natural fluctuation of pollutant concentrations and environmental factors to accurately calibrate the sensors' sensitivity and baseline response. It is advantageous due to its procedural simplicity, negligible consumable usage, and cost efficiency compared to laboratory assessments (Castell et al., 2017). Consequently, it has become as a preferred method for calibration in various scenarios (Spinelle et al., 2015, 2017).

Despite the widespread application of field side-by-side calibration, several critical concerns persist regarding the process. The primary issue is the selection of appropriate calibration conditions. Factors like the calibration duration (Levy Zamora et al., 2023), the pollutant concentrations distribution (Levy Zamora et al., 2023), sensor ageing(Li et al., 2021), interference from non-target gases (Cross et al., 2017), the impacts of temperature and relative humidity (Ariyaratne et al., 2023), and various gas sampling methods can significantly influence the calibration results. Determining the optimal conditions is crucial for achieving accurate and reliable calibration results. Extensive research has focused on the calibration period, the most frequently reported in recent studies (Datta et al., 2020; Gao et al., 2015; Kim et al., 2018; Mukherjee et al., 2019; Pinto et al., 2014; Spinelle et al., 2015, 2017; Topalovic et al., 2019). One study by Zamora et al. (2023) evaluated the impact of calibration period on calibration quality using calibration periods of up to 6 months from one year of $PM_{2.5}$, CO, NO, NO_2, and O_3 data in Maryland, US. Their results indicated diminishing improvements in median root-mean-square error (RMSE) for calibration periods longer than six weeks for all sensors. Zamora et al. (2023) also highlighted the importance of considering environmental conditions during the calibration period that are similar to those encountered during the evaluation period to achieve the best calibration performance. Another study by Okorn et al. (2021) reported that longer calibration periods (i.e., six weeks) resulted in fits with a reduced bias compared to fits obtained from shorter calibration periods (1 week), while the one-week calibrations yielded the best R^2 (coefficient of determination) values. While these studies have offered valuable

https://doi.org/10.5194/amt-2024-130
Preprint. Discussion started: 15 August 2024


Atmospheric Measurement Techniques Discussions

insights into sensor field calibration conditions, more discussion is needed on other calibration factors, particularly the range of pollutant concentrations during the calibration period and the selection of time averaging length for raw data before calibration, which are more easily to be standardized and quantifiable compared to other factors.

In addition to investigating calibration conditions, an equally crucial aspect to address is the development of an effective calibration model that can accommodate these optimized sensor calibration conditions. Most studies have adopted generic multiple linear regression (MLR) or machine learning models to calibrate raw sensor data, taking into account various complex variables such as temperature, relative humidity (RH), their gradient and cross-sensitivity to other pollutants (Datta et al., 2020; Han et al., 2021; Levy Zamora et al., 2023; Si et al., 2020; Topalovic et al., 2019; Wei et al., 2020; Zimmerman et al., 2018). These models, while comprehensive, often face limitations such as the risk of over-fitting, extensive training requirements, restricted applicability, and difficulties in replicating and scaling up for large sensor numbers. Furthermore, the complexity of machine learning models can pose significant barriers for everyday users.

Addressing these challenges, this study employed a simplified yet effective approach by establishing a linear calibration model and identifying the critical factors that influence calibration quality, thus optimizing calibration conditions for NO2, NO, CO, and O3 electrochemical sensors. We investigated this using a patented dynamic baseline tracking method designed to mitigate temperature and humidity effects on sensor signals, allowing the sensor devices, Mini Air Stations (MASs), to observe data most directly related to the concentration signal. This approach enabled the development and use of a refined linear calibration model. Our research uncovers three pivotal factors that significantly impact sensor calibration and validation performance: calibration period, concentration range, and time averaging. By examining these factors' effects on the variation of sensor's calibration coefficients, we aim to deepen the understanding of sensor calibration processes and enhance the performance of low-cost electrochemical air sensors. This methodology not only simplifies the calibration process but also ensures that the calibration model remains robust and applicable in varied and long-term field conditions.

2 Material and methods

2.1 Data collection

2.1.1 Sensor devices

Eight microsensor-based Mini Air Stations (MAS-AF300, Sapiens, China), hereinafter referred to as 'MAS', shown in Figure 1, were utilized in this study for continuous measurements of the air pollutants NO_2, NO, O_3, and CO under field conditions. Each MAS unit included three or four gas sensors along with a combined RH and temperature sensor (SHT-75, Sensirion AG). This study focuses on electrochemical gas sensors for NO_2 (Alphasense NO_2-B43F), NO (Alphasense NO-B4), CO (Alphasense CO-B4), and O_3 (Alphasense OX-B431). It should be noted that Alphasense OX-B431 sensor is designed to detect oxidizing gases (O_3 + NO_2) rather than O_3 alone. Therefore, to accurately measure O_3 concentration, it is necessary to pair the NO_2 sensor (NO_2-B43F) with the oxidizing gas sensor (OX-B431). By calculating the difference between the two sensors, the

https://doi.org/10.5194/amt-2024-130
Preprint. Discussion started: 15 August 2024

Atmospheric Measurement Techniques Discussions

O_3 concentration can be determined. Furthermore, the MAS system incorporates numerous sophisticated functionalities. It is equipped with an active air sampler, ensuring a flow rate of 0.8 L min^{-1}. The sample air undergoes filtration through a Teflon dust filter before directly entering the sensor module, without the implementation of any temperature or humidity control measures. The Teflon dust filter for each MAS will be replaced regularly every month to prevent dust from entering the gas module and causing measurement errors and shortening the sensor life. Moreover, to mitigate potential drift during long-term deployment, the MAS gas model incorporates an auto-zeroing function. During the zeroing process, the gaseous pollutant measurement module receives air samples from a separate zero module, from which NO, NO_2, and O_3 have been eliminated. The data collected during the zeroing period is subsequently analyzed to rectify any drift effects during the long-term deployment phase, as part of the data cleaning procedure. A comprehensive description of this technology and its functional advantages can be found in a paper by Sun(Sun et al., 2017). All these incorporated functionalities in the MAS system are aimed at optimizing sensor performance, enhancing measurement accuracy, and ensuring their long-term stability.

Figure 1. Structure diagram of MAS monitoring devices (dimensions: 420 × 320 × 180 mm, H × W × D; weight: 12 kg; power consumption: 15W).

2.1.2 Measurement campaign details

To assess sensor performance under varying ambient conditions, these MASs were deployed in three distinctively different urban and climatic settings: Hong Kong's humid subtropical climate, Macau's somewhat similar yet distinct urban environment, and Shanghai's more variable climatic conditions. Each city featured a co-location campaign with an AQMS, as detailed in Table 1, and the AQMSs were equipped with Federal Equivalent Method (FEM) reference analyzers.

https://doi.org/10.5194/amt-2024-130
Preprint. Discussion started: 15 August 2024

Atmospheric Measurement Techniques Discussions

The first co-location campaign in Hong Kong involved the four MASs, each equipped with all four types of gas sensors (NO_2, NO, CO, and O_3), which were placed at the Tseung Kwan O AQMS (22.3716°E,114.1148°N) regulated by the Hong Kong Environmental Protection Department, showcasing a wide range of urban air quality conditions. In the second co-location campaign, two MASs were located at the Taipa Air Quality Monitoring Station (22.15896°E, 113.56882°N) in Macau, focusing on NO_2, NO, and O_3 to capture the general urban background conditions unique to the region. The third campaign took place in Shanghai, where two MASs, monitoring NO_2, NO, and CO, were placed separately alongside two sets of reference analyzers at the Waigaoqiao Port 2 site (31.36662°E, 121.57242°N) and Port 4 site (31.33302°E, 121.65496°N). This campaign was also the longest co-location campaign, lasting 22 months, offering a prolonged observation of the diverse and more polluted air quality conditions typical of a major industrial hub. These locations were chosen to ensure a comprehensive analysis across a spectrum of urban pollution levels and environmental conditions.

All eight MAS units were designed to automatically transmit the measured raw sensor signals and concentration data of the pollutants from the MAS to a secure cloud server in real-time at 1-minute resolution. The reference analyzer in Hong Kong provided 1-minute time resolution pollutant concentration data, while those in Macau and Shanghai provided hourly averaged data, enabling us to conduct calibration analysis at varying time resolutions.

Table 1. Details of MAS devices in co-location calibrations.

Location	MAS ID	Reference analyzer data time resolution	Co-location periods	Monitoring pollutants and concentration range (5^{th} to 95^{th} percentile range)	MAS inside temperature and RH range
Hong Kong	MAS1	Minute	2021-07-27 00:00 to 2022-10-10 00:00 (15 months)	NO_2: 3.7 ppb - 34.6 ppb NO: 0.4 ppb - 18.0 ppb CO: 152 ppb - 643 ppb O_3: 4.3 ppb - 69.1 ppb	Temp: 10 ˚C - 43˚C RH: 17% - 85 %
	MAS2	Minute	2021-12-24 00:00 to 2022-10-10 00:00 (10 months)		Temp: 10 ˚C - 46˚C RH: 16% - 86 %
	MAS3, MAS4	Minute	2021-07-10 00:00 to 2022-10-10 00:00 (15 months)		Temp: 10 ˚C - 45˚C RH: 16% - 93 %
Macau	MAS5, MAS6	Hourly	2021-04-04 13:00 to 2022-04-26 05:00 (13 months)	NO_2: 0 ppb - 26.3 ppb NO: 0 ppb - 17.6 ppb O_3: 0 ppb - 68.8 ppb	Temp: 10 ˚C - 47˚C RH: 21% - 89 %

https://doi.org/10.5194/amt-2024-130
Preprint. Discussion started: 15 August 2024

Atmospheric Measurement Techniques Discussions

Shanghai	MAS7, MAS8	Hourly	2019-10-12 01:00 to 2021-07-31 23:00 (22 months)	NO_2: 14.1 ppb - 63.4 ppb NO: 3.2 ppb - 142.5 ppb CO: 258 ppb - 862 ppb	Temp: -8 ˚C - 51˚C RH: 0% - 90 %

2.2 Dynamic baseline tracking method to mitigate environmental effects on sensors

The design of the MAS enables the isolation of the concentration signal from environmental variables of temperature and RH through a dynamic baseline tracking method, which operates by differentiating between the environmental and pollutant concentration induced sensor signals using a dual-sensor module. This gas sensor system comprises a primary sensor - an electrochemical gas sensor exposed directly to the air, capturing the original signal (designated as ORG) influenced by pollutants, temperature, and RH, and a reference sensor - an identical electrochemical gas sensor paired with a patented pair differential filter (designated as PDF) allowing only water molecules to pass through. This setup prevents the target gas pollutants from reaching the reference sensor, thereby isolating and measuring the dynamic environmental impact on the sensor baseline response. This process is referred to as "dynamic baseline tracking method" in this study. Prior to initiating the co-location campaign, a 15-day pre-test under field conditions and a laboratory test in the environmental chamber were conducted to demonstrate the method's capability to enhance the sensor performance under varying temperature and humidity conditions. Figure 2 shows the layout of each MAS sensor module and illustrates how the dynamic baseline tracking method works under laboratory and ambient conditions. Each MAS sensor module produces four distinct outputs for a specific pollutant: (i) the ORG sensor signal in volts, V_{ORG}, (ii) the PDF sensor signal in volts, V_{PDF}, (iii) the voltage output from the difference of the ORG and PDF sensor signals in volts, V_{DIFF}, and (iv) the concentration output of target gas in ppb, *Conc*. Each MAS has an onboard algorithm capability that converts sensor signals to concentration, with the conversion automatically performed onboard the MAS for real-time concentration output. Eq. (1) presents the conversion equation for NO_2, NO, and CO, where '*a*' denotes the slope of the equation, which is also indicative of the sensitivity (ppb mV^{-1}) of the electrochemical sensors, and *'b'* represents the intercept of the equation. For the gas sensors exhibit cross-sensitivity with non-target gases, an interfering gas correction component can be incorporated. Eq. (2) presents the equation for calculating O_3 concentrations using the Alphasense OX-B431 sensor with NO_2 as an interferent. The coefficient 'f' accounts for the cross-interference from NO_2, and our empirical data, derived from a substantial number of tests, indicates that 'f' typically falls within the range of 0.8 to 1.2.

$$Conc(NO_2, NO, CO) = a \times V_{DIFF} + b\,, \quad (1)$$

$$Conc(O_3) = a \times V_{DIFF} + b - f \times Conc(NO_2)\,, \quad (2)$$

https://doi.org/10.5194/amt-2024-130
Preprint. Discussion started: 15 August 2024

Atmospheric Measurement Techniques Discussions

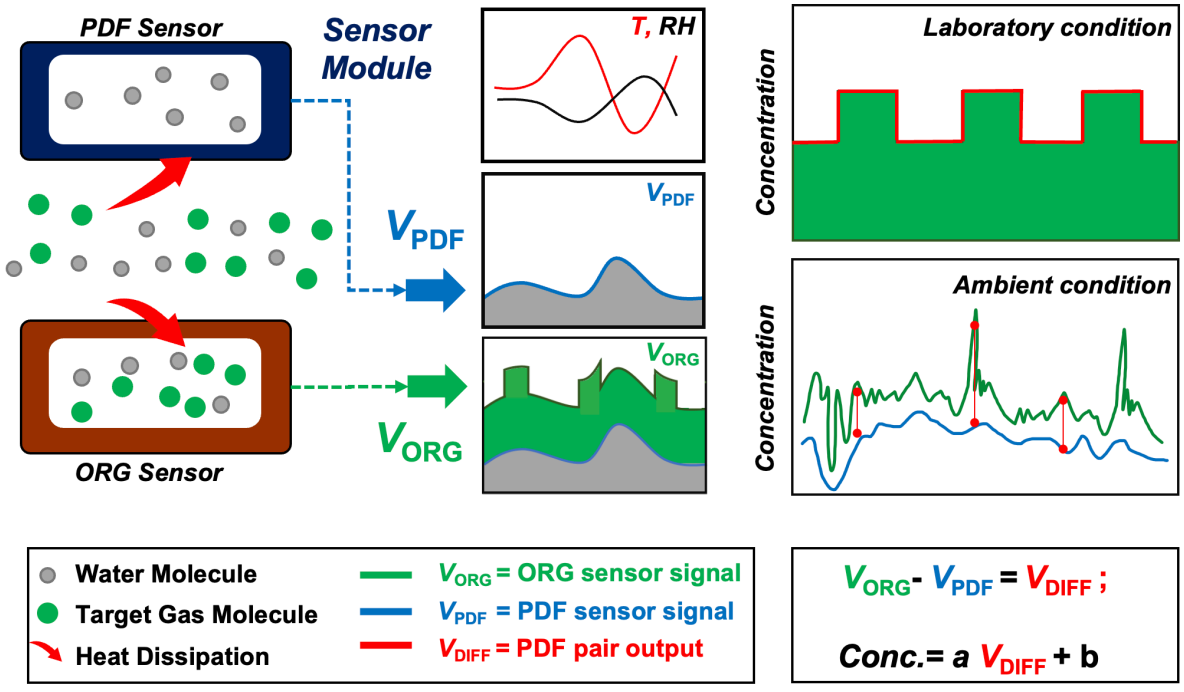


Figure 2. A conceptual diagram of the PDF enabled MAS sensor device.

160

2.3 Impact analysis of three crucial factors on calibration conditions

This study specifically focuses on conducting field tests to identify optimal calibration conditions by examining three primary factors that influence sensor calibration performance: (a) calibration period duration; (b) concentration variation range; and (c) time averaging pre-processing.

165 **Calibration Period Optimization**

Calibration is typically conducted within a specific timeframe, constrained by time and resource availability. Standard protocols involve calibrating sensors over durations ranging from a few days to several weeks prior to their utilization in field monitoring applications. The calibration's effectiveness largely depends on this timeframe, referred to as the calibration period. The calibration period test in this study uses subsets of the full co-location period to generate a range of hypothetical calibration

170 period. We investigated calibration period scenarios ranging from 1 to 15 days. In each scenario, 500 samples were randomly selected using the numpy.random.choice() function in Python to simulate real-world sensor calibration practices and ensure the randomness and independence of sample selection. Sample sizes of 250, 500, and 1000 were tested, results stabilized with 500 samples, indicating minimal impact from decreasing or increasing the sample size further. The 500 randomly selected calibration periods were illustrated in Figure S1 in the supplementary materials, which shows the start times for these periods

175 for NO, with the approach also applied to NO_2, CO, and O_3 sensors.

https://doi.org/10.5194/amt-2024-130
Preprint. Discussion started: 15 August 2024


Atmospheric Measurement Techniques Discussions

These calibration samples were used as the training set for each hypothetical calibration period in the calibration model to evaluate the range of potential R^2 and RMSE when applied in the sensor validation periods. Firstly, these samples were standardized to hourly data to facilitate consistent comparisons across various MAS units. The calibration coefficients (slope and intercept) of these samples were calculated as per Eq. (1) or Eq. (2). Subsequently, these coefficients were validated using the following month's data by comparing the hourly calibrated sensor data and hourly reference data, with superior validation performance suggesting an optimal calibration period. This evaluation was not limited to the calibration period's immediate outcome; it also included a comparison of R^2 and RMSE metrics against the hourly data validation set from the subsequent month. This dual-phase evaluation underscores that the calibration's true merit is better judged during the post-calibration validation phase, adhering to the standard practice of a bounded calibration period followed by an extended validation phase.

Concentration Range Analysis

We propose the hypothesis that users can strategically select a co-location period to minimize the calibration duration, recognizing that the calibration period is not the sole factor to consider when optimizing instrument co-location for calibration purposes. A critical aspect is to evaluate the representativeness of environmental conditions during the calibration period in relation to those observed during the long-term evaluation periods. Since the influence of temperature and RH on sensor signals has been eliminated, concentration emerges as the key factor that accurately reflects environmental conditions. To analyze how the range of pollutant concentrations during the calibration period affected the sensor validation performance, we compared the validation R^2 and RMSE outcomes with the same calibration period length but varied concentration ranges.

Firstly, we segmented the samples into distinct categories based on their concentration ranges while maintaining a constant calibration period. We employed the 5^{th} to 95^{th} percentile of the pollutant concentration in each category to define each range. This approach mitigates the impact of sporadic peak values, ensuring they do not disproportionately affect the overall concentration range assessment. Subsequently, the effectiveness of calibration across these ranges was systematically evaluated by comparing R^2 and RMSE metrics during the validation periods in the subsequent month. This strategy enabled a thorough examination of how the concentration range impacts calibration accuracy, providing insights into the optimal range needed for precise sensor calibration.

Time Averaging Evaluation

We also evaluated the influence of time averaging on calibration efficacy to identify the optimal data resolution for the best calibration outcomes. Given that reference analyzers and sensors can provide data at granular levels, down to minutes or seconds, pre-calibration data processing plays a crucial role in the accuracy of calibration.

In this time averaging analysis, we compared the calibration performance of data averaged over different time intervals, from minutes to hours. After processing the calibration data set with varied time averaging intervals, the resulting calibration coefficients were evaluated against the data from the following month's validation set. For example, for a sample with calibration period of 1 day, sensor and reference data were averaged over 1/3/5/7/9/11/30/60/120/180 minutes and used to determine the sensor coefficients for each time averaging interval. Following that, these coefficients were independently applied to the following one-month validation period with hourly data, to determine the R^2 and RMSE under each time

https://doi.org/10.5194/amt-2024-130
Preprint. Discussion started: 15 August 2024


Atmospheric Measurement Techniques Discussions

averaging intervals. The ideal time averaging interval was determined based on the highest R^2 and lowest RMSE values obtained in this validation phase, pinpointing the most effective time resolution for calibration.

3 Results and discussion

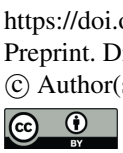
3.1 MAS sensor performance against temperature and RH variability

Before initiating the long-term co-location campaign, four MAS units equipped with NO_2, NO, CO, and O_3 sensors were tested in Hong Kong, demonstrating the dynamic baseline tracking method's ability to enhance performance against varying temperatures and RH. During the 15-day pre-test in the summer (June 1-15, 2021), temperatures varied between 28 ˚C and 42 ˚C, with RH levels from 45% to 87%. Figure 3(b)-(e) depicts the calibration readings contrasting with and without the PDF application. For NO_2, the sensors with the PDF module showed stronger performance, with a high R^2 (0.95-0.99) and low RMSE (0.94-1.73), compared to the lower R^2 (0.44-0.57) and higher RMSE (5.08-5.80) for the sensors without the PDF module. For NO and O_3, the sensors with the PDF module also demonstrated stronger performance compared to the sensors without the PDF module. Specifically, the sensors with the PDF module had strong and consistent R^2 (0.97-0.98 for both NO and O_3) and low RMSEs (1.63-1.79 for NO, 1.02-1.11 for O_3), while the sensors without the PDF module had weaker R^2 (0.73-0.83 for NO, 0.47-0.60 for O_3) and higher RMSEs (4.25-5.37 for NO, 4.14-4.70 for O_3). For CO, the sensors exhibited comparable performance, with R^2 around 0.93-0.94 and RMSE values between 16.70-19.00, regardless of the PDF module. These significant discrepancies, especially for NO, NO_2, and O_3, highlight the importance of the dynamic baseline tracking method in improving the accuracy and reliability of measurements, notably under low concentration conditions influenced by temperature and RH.

Additionally, laboratory tests in environmental chambers assessed the MAS NO sensor (Figure S2), exposing it to broad temperature (0°C to 30°C) and RH (10% to 90%) ranges. Despite these fluctuations, MAS sensors maintained consistent and stable readings after applying the dynamic baseline tracking method, as shown in Figure S2(b), with concentration steps from 50 to 300 ppb. The outcomes from both field and laboratory tests confirm that the dynamic baseline tracking method effectively neutralizes temperature and RH effects, primarily for NO_2, NO, and O_3 sensors, achieving desired performance while focusing primarily on concentration factors for subsequent analysis. Similar pre-tests were also conducted with the MAS units in Macau and Shanghai to assess the effectiveness of the dynamic baseline tracking method.

Upon completion of the pre-tests, the long-term field co-location campaigns were initiated. The dynamic baseline tracking method was first evaluated in this study to prove its effectiveness in long-term field tests. The performance of MAS1, particularly for NO and NO_2, throughout the campaign, was depicted in Figures S3 and S4. It should be noted that a single fixed calibration coefficient was used throughout the entire campaign duration. This fixed coefficient enabled the calibrated sensor data to consistently perform well throughout the co-location campaign. The absolute error (sensor - reference) generally stayed within ± 5ppb, and the relative error (absolute error/reference) was primarily under 15%, indicating effective mitigation of temperature and RH impacts on the sensor's output, even during extended field conditions over a year. Importantly, the

https://doi.org/10.5194/amt-2024-130
Preprint. Discussion started: 15 August 2024



Atmospheric Measurement Techniques Discussions

long-term analysis in Figures S3 and S4 showed that selecting suitable calibration coefficients can ensure the sensors' stability and accuracy over prolonged periods. However, dedicating several months or even up to a year for calibration is not feasible in standard practice. Therefore, our main goal is to determine the optimal coefficients from short-term calibration periods to enhance long-term validation performance.

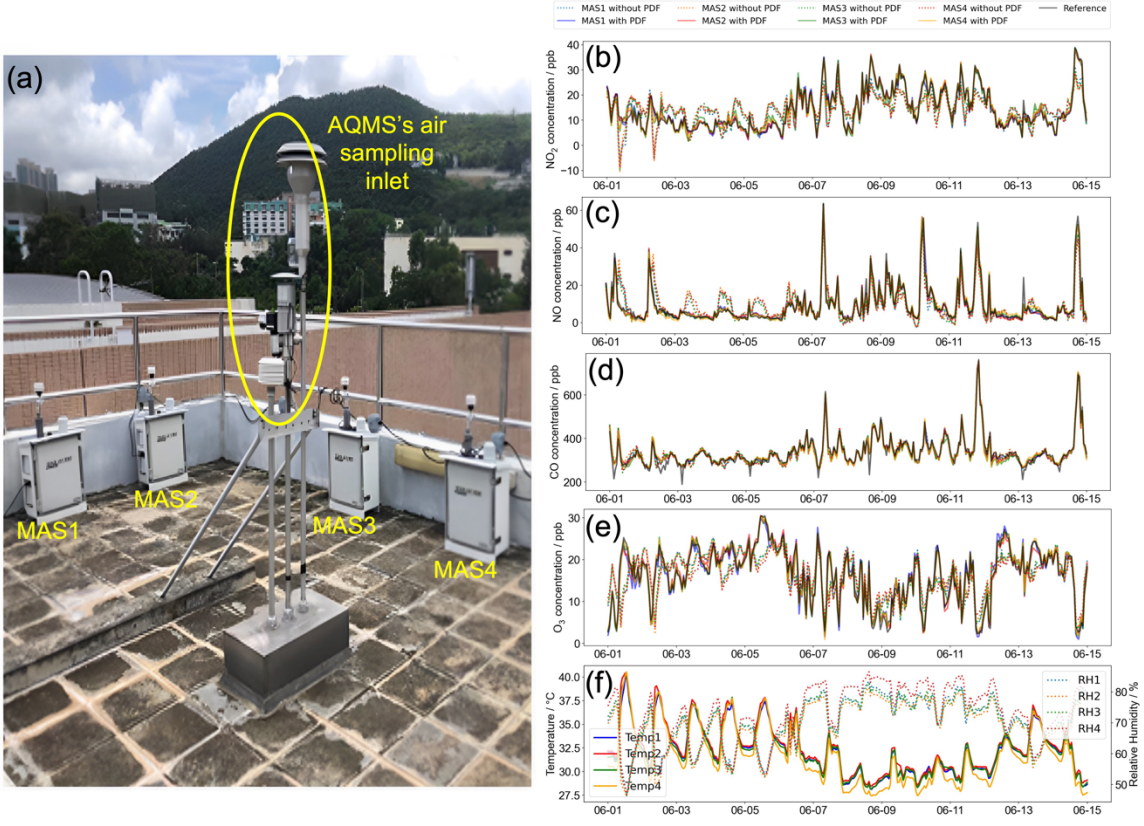


Figure 3. (a) Setup and (b-e) NO_2, NO, CO, and O_3 long-term field data comparison of four MAS units with the AQMS in Hong Kong in 2019. (f) shows the temperature and RH measured inside the four MAS gas sensor modules.

3.2 Impact of calibration period on sensor calibration

As detailed in Section 2.3, we used 500 randomly selected samples for each calibration period, and this process generated 500 sets of calibration slopes and R^2 / RMSE values from the validation period. Figure S5 displays the median and the 25th to 75th percentile range of these R^2 / RMSE results across all eight MAS units with NO_2 and NO sensors and all six units with CO and O_3 sensors. Figure 4 extracts the 25th to 75th percentile of each MASs results and combines them into a boxplot, making the trend across the calibration period more apparent. An increase in the median of R^2 (e.g. for NO, R^2 improved from 0.83 to 0.95

https://doi.org/10.5194/amt-2024-130
Preprint. Discussion started: 15 August 2024


Atmospheric Measurement Techniques Discussions

as the calibration period went from 1 to 15 days) coupled with a reduction in the median of RMSE (e.g. for NO, RMSE decreased from 3.71 to 2.12 over the same calibration period) shown in Figure 4 indicate improved validation performance. The narrowing of the 25th to 75th percentile range across calibration periods (e.g. for NO, R^2 range tightened from 0.66-0.96 to 0.90-0.98 as the calibration period went from 1 to 15 days) further supports this, with a tightening of validation performance towards a steadier state and reduced chance of abnormal calibration.

In Figure 4, the most notable enhancements in validation performance were observed within the initial 1 to 3 days. Beyond this period, the rate of improvement was found to be less clear, with the median R^2 increasing by less than 0.02 and the median RMSE decreasing by less than 0.1 (but less than 1 for CO) for further increases in the calibration period. For NO_2, NO, and O_3, the upward trend in validation R^2 and the downward trend in RMSE were observed, plateauing after 5 days. CO sensors in most MAS units reach stable R^2 after 7 days. This suggests lengthening the calibration period beyond 5 days for NO_2, NO, O_3 or 7 days for CO does not markedly benefit sensor data performance. If the sensor users can strategically select the co-location period to minimize the calibration duration, a period of 5–7 days is identified as most effective for minimizing errors in calibration coefficient and avoiding notably low validation R^2 values.

A noteworthy observation in Figure S5 is that the NO_2 and NO sensors in MAS7 and MAS8 of Shanghai campaign showed consistent performance over all calibration periods, likely due to the high pollutant concentrations in the Shanghai port area. Despite the short calibration duration of 1–3 days, the extensive concentration range assessed contributed to more precise calibration coefficients and improved validation performance, as will be discussed in next section.

https://doi.org/10.5194/amt-2024-130
Preprint. Discussion started: 15 August 2024

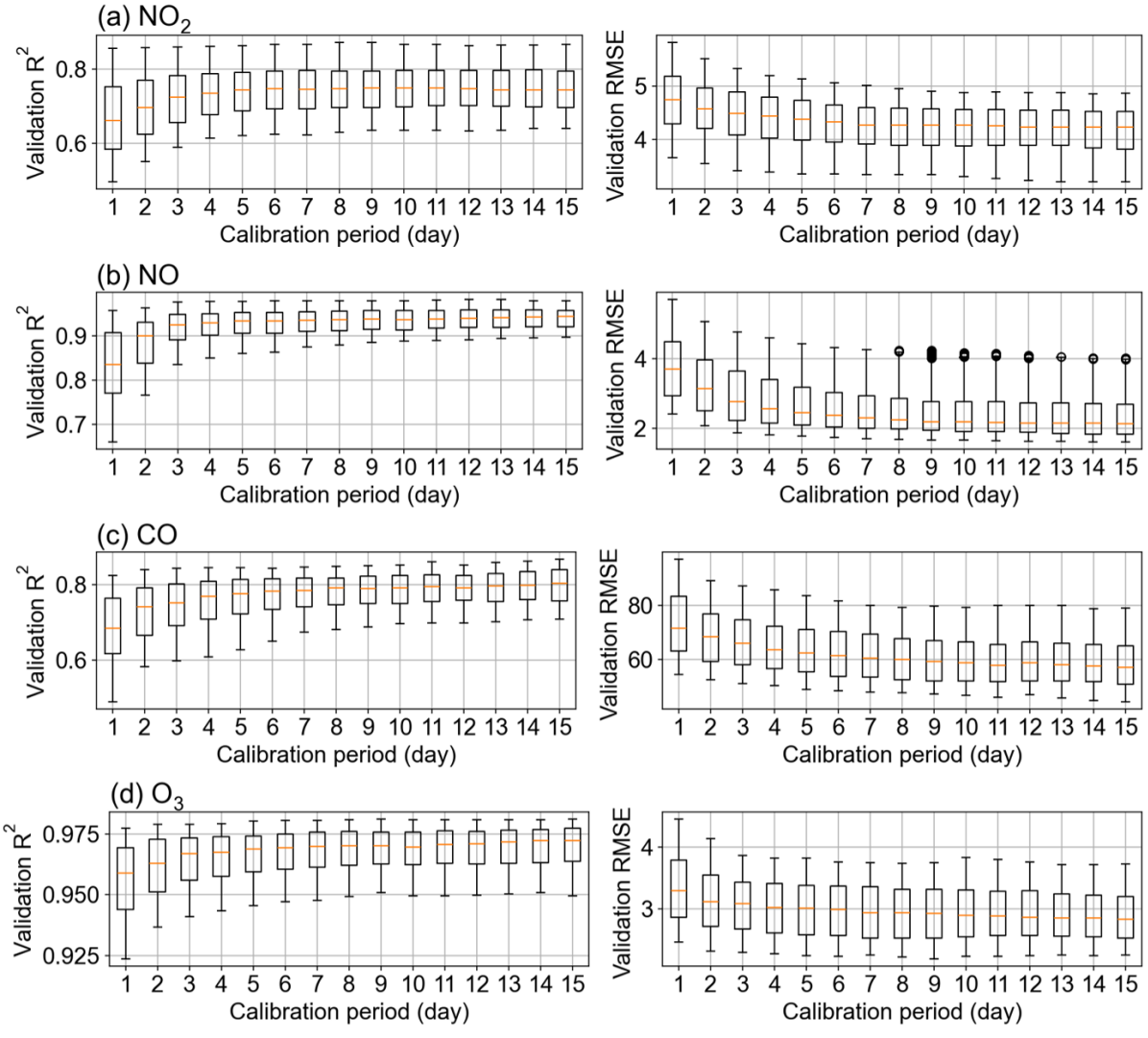


Figure 4. The range of the validation R^2 and RMSE for a given calibration period for all MAS units consists of (a) NO_2, (b) NO, (c) CO, and (d) O_3 sensors. The vertical error bar is the 25%–75% distribution of R^2 and RMSE under different calibration periods.

3.3 Impact of concentration range on sensor calibration

Another critical aspect is the impact of the concentration range experienced by the sensors during calibration periods. Figure S5 shows that MAS7 and MAS8 in the Shanghai campaign could achieve accurate and reliable calibration for NO and NO_2 within just a day, given their exposure to environments with significant concentration variability. Consequently, our second test examined the effect of the concentration range.

https://doi.org/10.5194/amt-2024-130
Preprint. Discussion started: 15 August 2024


Atmospheric Measurement Techniques Discussions

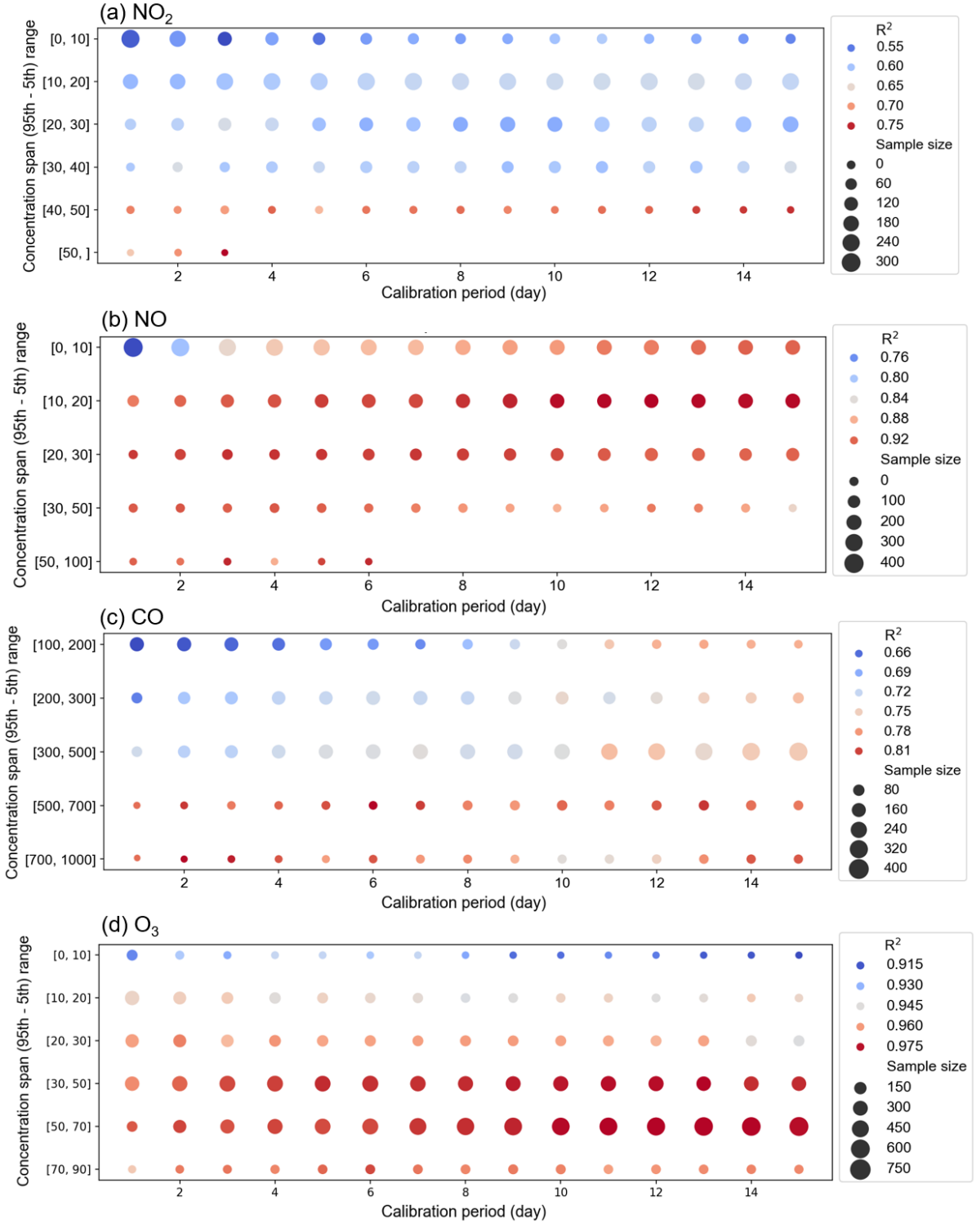
Samples in Figure 4 were grouped based on different concentration ranges, and the results were shown in Figure 5 and Figure S6 to explore the relationship between calibration period length, concentration range, and sensor validation performance, categorizing the MAS units accordingly. For NO_2 and NO sensors of MAS7 and MAS8, a separate analysis was essential due to their higher pollutant concentrations compared to other units, as detailed in Table 1. Therefore, MASs 1-6 were evaluated together in Figure 5 under a lower concentration range, with 90% of NO_2 and NO ranges falling below 40 ppb and 50 ppb, respectively. In contrast, MAS7 and MAS8 were assessed in Figure S6 under higher concentration ranges, where 90% of the readings for both gases exceeded these thresholds.

Figure 5 illustrates the calibration conditions at lower concentrations typical of environments like Hong Kong and Macau. The red zone of Figure 5, indicating higher R^2 values, is primarily concentrated in areas with wider concentration ranges. Specifically, when examining the performance of NO_2 sensors, the lowest R^2 value of 0.55 was recorded in the 0-10 ppb range, while the highest R^2 value of 0.75 was recorded in the >50 ppb range. When the calibration period is held constant, an increase in the concentration range boosts the validation R^2 from 0.55 to 0.75 with a notable turning point at 40 ppb. However, extending the calibration period without increasing the concentration range doesn't obviously improve the validation R^2. NO CO and O_3 also displayed patterns similar to NO_2, with R^2 improvements linked to wider concentration ranges. For all gases, the highest R^2 values were predominantly observed in the broadest concentration ranges. Therefore, achieving higher validation R^2 values above the median, such as $R^2 > 0.65$ for NO_2, $R^2 > 0.84$ for NO, $R^2 > 0.75$ for CO, and $R^2 > 0.95$ for O_3, requires significant concentration ranges, notably more than 40 ppb for NO_2 and 10 ppb for NO, 500 ppb for CO, and 20 ppb for O_3. Reaching these ranges allows the calibration coefficients to stabilize and align closely with those derived from year-long calibration results.

Additionally, the differences in the concentration range thresholds suitable for the different gas sensors may be attributed to the distribution characteristics of the gas pollutants in the surrounding environment. Notably, as determined in the just-obtained results, the NO concentration range of 10 ppb is the lowest, possibly due to the prevalence of high ambient NO concentrations frequently appearing in the form of peaks. Consequently, when employing the 5^{th} to 95^{th} percentile as the criteria for concentration range, the NO range is observed to be the lowest among the gases. Moreover, the higher concentration range analysis in Figure S6 shows that increasing the concentration range beyond 40 ppb for NO_2 and 50 ppb for NO does not improve validation R^2 values, further indicating a threshold in the concentration range beyond which no additional sensor performance benefits are observed. Overall, this underscores the inadequacy of merely extending the calibration duration, and it is crucial to ensure an adequate concentration range during the calibration period. However, beyond a certain concentration range threshold, further increases in the calibration range do not lead to additional improvements in the calibration results.

https://doi.org/10.5194/amt-2024-130
Preprint. Discussion started: 15 August 2024

Atmospheric Measurement Techniques Discussions



https://doi.org/10.5194/amt-2024-130
Preprint. Discussion started: 15 August 2024


Atmospheric Measurement Techniques Discussions

315 **Figure 5. (a) NO_2, (b) NO, (c) CO, and (d) O_3 bubble plot of median R^2 of MAS units 1–6 (as located in the low-concentration regions (Hong Kong and Macau)) and two factors: calibration period and concentration range. The size of the bubbles represents the number of samples. The color represents the median R^2 values in corresponding categories. Red represents the higher R^2 value, while blue represents the lower R^2 value.**

320 **3.4 Impact of time averaging on sensor calibration**

Another factor influencing calibration is the time averaging of the raw data, particularly for high-frequency measurements, taken at intervals of a minute or seconds. Performing temporal averaging is critical before formulating the calibration equation. As indicated in Table 1, the reference data from Hong Kong provides a one-minute temporal resolution. Thus, for calibrating sensors MAS1 - 4, identifying the optimal time averaging is crucial, as it enhances the accuracy of calibration coefficients and
325 guarantees a substantial data volume for a reliable calibration process.

Using MAS1 as an illustrative case, Figure 6 shows the calibration sample processing for its NO_2 sensors, with different colors denoting time averages ranging from one minute to three hours. This indicates the sensor and reference data for each calibration sample underwent time averaging across intervals of 1/3/5/7/9/11/30/60/120/180 minute(s). Subsequent calibration and validation led to the determination of the calibration slope, R^2 of the validation set, and RMSE for these time-averaged intervals.
330 Extending this process to 500 samples per calibration period, we derived their median values, as depicted in Figure 6. Analysis of Figure 6's vertical axis reveals that, for a one-day calibration period, R^2 values improved post hourly (R^2 = 0.68) and 5-minute averaging (R^2 = 0.66) compared to the baseline 1-minute data (R^2 = 0.59), with a corresponding reduction in RMSE. For periods exceeding a day, median R^2 values exhibited a modest rise from 0.64-0.66 for 1-minute data to 0.68-0.70 for hourly data, suggesting the shorter the calibration period, the more pronounced the benefit of longer time averaging. Hence, calibrating
335 with minute-level data over short periods of 1-3 days may lead to suboptimal validation performance. Similar trends were observed for NO, CO, and O_3, as shown in Figures S7-S9, with MAS2, MAS3, and MAS4 in Hong Kong mirroring the findings from MAS1.

The results indicate that data averaged over an hour are more suitable for calibration than minute-level data. As depicted in Figure 6 and Figures S7–S9, a critical juncture is identified at the 5-minute mark (highlighted by a green line with a star). After
340 this point, the improvements in validation R^2 and RMSE become substantially less obvious. Thus, for data originally recorded at 1-minute intervals, applying a time averaging of 5 minutes or longer boosts the performance of the validation set, aligning the calibration coefficient more closely with the optimal one. The enhanced performance of hourly over minute-level data across various calibration periods warrants further investigation in next section to understand the underlying factors.

https://doi.org/10.5194/amt-2024-130
Preprint. Discussion started: 15 August 2024

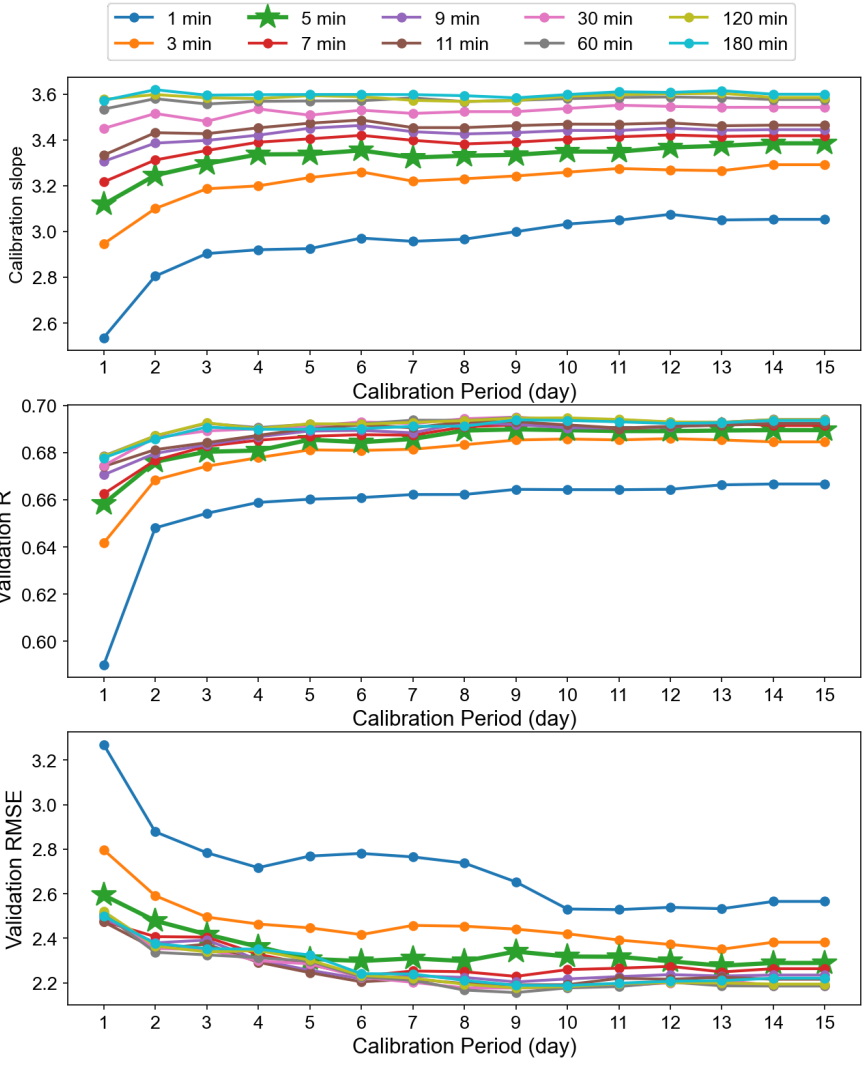


Figure 6. The calibration slope median, the R^2 median of the validation set, and the RMSE median of the validation set under different time averaging for a given calibration period for MAS1's NO_2 sensor.

3.5 Potential causes of sensor calibration coefficient variation

We selected a sample from the MAS1 NO_2 sensor with a one-day calibration period to analyze the benefits of hourly over minute-level data averaging. Regression analysis between sensor and reference data was performed for both 1-minute and 1-hour averages. Initially, data fitting during the calibration period was assessed. The time series plot in Figure 7(a) shows that both minute and hourly averaged data closely align with the reference. However, obvious differences emerged when computing the calibration equations separately for each time frame. The calibration slope for minute-level data (a = 2.70) was substantially

https://doi.org/10.5194/amt-2024-130
Preprint. Discussion started: 15 August 2024


Atmospheric Measurement Techniques Discussions

lower than that for hourly data (a = 3.91), corroborating the trends noted in Figure 6. This discrepancy is evident in Figure 7(b), where the regression curves for minute-level and hourly data diverge. The orange line for hourly data intersects more closely with the dense cluster of orange dots representing minute-level data, unlike the minute-level data's blue fitting line, which misses this dense area. In the validation phase, applying the distinct calibration coefficients derived for minute and hour averages to the next month's dataset also highlighted clear differences. Figure 7(c) and (d) illustrate that minute-level calibration coefficients (blue line) resulted in less consistent sensor data with the reference (R^2 = 0.72, RMSE = 5.60) than the hourly data (R^2 = 0.93, RMSE = 2.74), especially at lower concentrations.

The discrepancy between the two sets of calibration coefficients is further illustrated in the data distribution plots in Figure S10, where sensor and reference data distributions for varying time averaging lengths are compared. As the time averaging interval increases, the sensor data distribution more closely mirrors the reference data distribution. This observation supports the notion that time averaging can refines the accuracy of the calibration by aligning data distributions, leading to more precise calibration outcomes. This pattern consistently appeared across various samples, MAS units, and gases, as described in section 3.4, demonstrating the superior calibration accuracy achieved with longer averaging periods.

Furthermore, we investigated the potential factors for the observed pattern by analyzing the residual term in sensor calibration model from the mathematical perspective. One plausible explanation is that the existence of influential factors has not been incorporated in the calibration model. As a result, the predictive capability of the calibration model may be compromised as it fails to accurately capture the relationship between sensor and reference data. The detailed analysis of this pattern is provided in Text S1 of the Supplementary Material.

https://doi.org/10.5194/amt-2024-130
Preprint. Discussion started: 15 August 2024

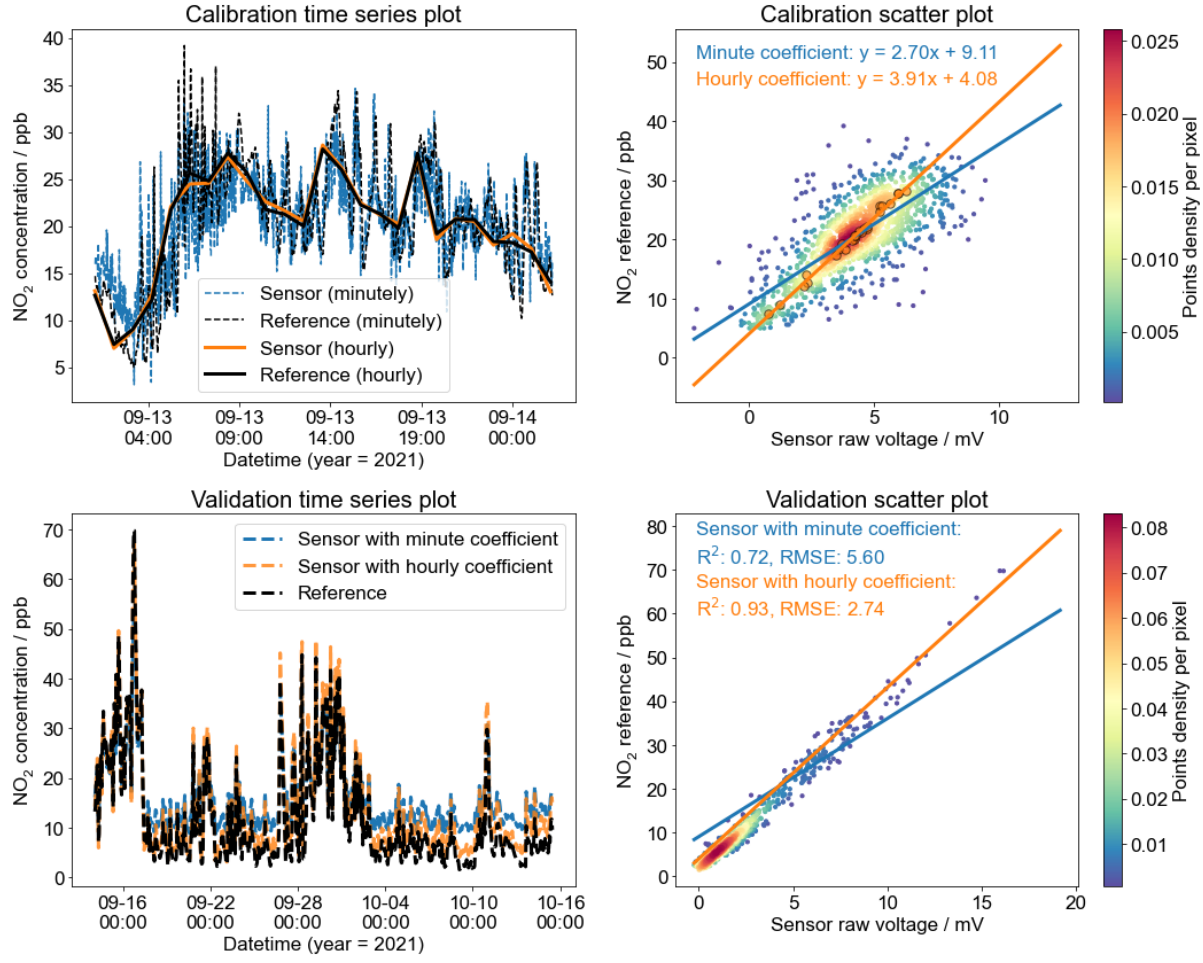


Figure 7. One of the calibration samples for MAS1's NO_2 sensor with a calibration period of 1 day. (a) is the time series and (b) is the scatter plot of minute-level and hourly data for the NO_2 sensor and reference during the calibration period. (c) is the time series and (d) is the scatter plot of minute-level and hourly data for the NO_2 sensor and reference during the validation period. The color bars in (b) and (d) represent the sample size in each region.

4. Conclusions

This study aimed to identify and analyze three critical factors influencing the sensor calibration performance of PDF based electrochemical NO_2, NO, CO, and O_3 sensors: calibration period, concentration range, and time averaging. By co-locating eight MAS units with reference analyzers in three cities over a period of up to 22 months, a comprehensive framework for sensor calibration was established. The study utilized a dynamic baseline tracking method, enhancing the consistency between

https://doi.org/10.5194/amt-2024-130
Preprint. Discussion started: 15 August 2024


Atmospheric Measurement Techniques Discussions

MAS sensor data and reference measurements. This method effectively countered the impacts of temperature and RH, focusing on pollutant concentration as the primary factor for MAS performance assessment.

In the calibration period analysis, equations were derived from 500 randomly selected samples for each period ranging from 1 to 15 days, with subsequent evaluation against the validation data. Initial improvements in validation performance were notable within the first 1 to 3 days of the calibration period, stabilizing around 5 to 7 days. This pattern suggests that extending the calibration period beyond 7 days yields negligible benefits, hence a 5–7 days calibration period is advocated to reduce calibration coefficient errors.

The concentration range assessment indicated that broader ranges enhance the validation R^2 values across all gas sensors. This finding emphasizes the necessity of establishing a concentration range threshold to facilitate effective calibration. Optimal ranges were determined as over 40 ppb for NO_2, 10 ppb for NO, 500 ppb for CO, and 20 ppb for O_3, with these thresholds ensuring reliable calibration coefficients and minimizing uncertainty in the results.

Time averaging's impact on calibration was significant, with improved coefficients and validation performance as averaging intervals increased. Notably, a one-day calibration period showed the most substantial improvement, with hourly and 5-minute averages providing higher R^2 values than one-minute intervals. A 5-minute threshold emerged as critical, advocating for a minimum of 5-minute averaging to enhance calibration accuracy and align coefficients with the optimal standard.

This study offers comprehensive insights into calibrating electrochemical gas sensors, highlighting the calibration period, concentration range, and time averaging's roles. Recommended practices for optimal calibration include: (1) a calibration period of 5–7 days using hourly data, (2) a concentration variation range (5^{th} to 95^{th} percentile range) exceeding 40 ppb for NO_2, 10 ppb for NO, 500 ppb for CO, and 20 ppb for O_3, and (3) a time averaging of 5 minutes or longer, preferably utilizing hourly data. The findings highlight the importance of balancing these factors to achieve optimal calibration outcomes, while extending certain calibration aspects beyond recommended thresholds may not yield additional benefits.

Acknowledging the limitations of this study, which focused exclusively on our MAS sensor technology with its active flow gas sampler, it should be noted that the specific calibration protocol described may not be directly applicable to studies involving different sensor types, commercial sensor packages from various manufacturers, or different air sampling methods using passive samplers. Optimal calibration conditions may vary depending on the sensor's specific features and the calibration methods employed. Future research endeavors should aim to diversify sensor types and increase the number of test sensors, thereby enhancing the generalizability and practicality of the findings. Nonetheless, the primary objective of this study is to provide methodological insights that can serve as a valuable reference for calibrating various sensor types. The developed dynamic baseline tracking method, along with the determined optimal calibration period, concentration range thresholds, and time averaging period, can inform and guide future research and calibration efforts for a wide range of sensors used in air quality monitoring. By establishing a foundation for standardized calibration approaches, this study contributes to advancing sensor technologies and promoting the generation of reliable and comparable air quality data across diverse monitoring networks.

https://doi.org/10.5194/amt-2024-130
Preprint. Discussion started: 15 August 2024

Atmospheric Measurement Techniques Discussions

CRediT authorship contribution statement

Han Mei: Writing – original draft, Visualization, Methodology, Data curation, Conceptualization. **Peng Wei:** Writing – review & editing, Validation, Methodology, Conceptualization. **Meisam Ahmadi Ghadikolaei:** Writing – review & editing, Investigation. **Nirmal Kumar Gali:** Writing – review & editing, Visualization, Investigation. **Ya Wang:** Software, Methodology. **Zhi Ning:** Writing – review & editing, Validation, Supervision, Methodology, Conceptualization.

Declaration of competing interest

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

Code/Data availability

Code and data will be made available on request.

Acknowledgements

The authors acknowledge the financial support received from the Research Grants Council of Hong Kong through the General Research Fund (16212022) and also acknowledge the support received from the Environmental Protection Department, HKSAR.

References

Ariyaratne, R., Elangasinghe, M. A., Zamora, M. L., Karunaratne, D. G. G. P., Manipura, A., Jinadasa, K. B. S. N., & Abayalath, K. H. N. (2023). Understanding the effect of temperature and relative humidity on sensor sensitivities in field environments and improving the calibration models of multiple electrochemical carbon monoxide (CO) sensors in a tropical environment. *Sensors and Actuators B: Chemical*, *390*, 133935. https://doi.org/10.1016/j.snb.2023.133935

Bisignano, A., Carotenuto, F., Zaldei, A., & Giovannini, L. (2022). Field calibration of a low-cost sensors network to assess traffic-related air pollution along the Brenner highway. *Atmospheric Environment*, *275*, 119008. https://doi.org/10.1016/j.atmosenv.2022.119008

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*, *99*, 293–302. https://doi.org/10.1016/j.envint.2016.12.007

https://doi.org/10.5194/amt-2024-130
Preprint. Discussion started: 15 August 2024

Cross, E. S., Williams, L. R., Lewis, D. K., Magoon, G. R., Onasch, T. B., Kaminsky, M. L., Worsnop, D. R., & Jayne, J. T. (2017). Use of electrochemical sensors for measurement of air pollution: Correcting interference response and validating measurements. *Atmospheric Measurement Techniques*, *10*(9), 3575–3588. https://doi.org/10.5194/amt-10-3575-2017

Datta, A., Saha, A., Zamora, M., Buehler, C., Hao, L., Xiong, F., Gentner, D., & Koehler, K. (2020). Statistical field calibration of a low-cost PM2.5 monitoring network in Baltimore. *ATMOSPHERIC ENVIRONMENT*, *242*. https://doi.org/10.1016/j.atmosenv.2020.117761

Gao, M., Cao, J., & Seto, E. (2015). A distributed network of low-cost continuous reading sensors to measure spatiotemporal variations of PM2.5 in Xi'an, China. *ENVIRONMENTAL POLLUTION*, *199*, 56–65. https://doi.org/10.1016/j.envpol.2015.01.013

Han, P., Mei, H., Liu, D., Zeng, N., Tang, X., Wang, Y., & Pan, Y. (2021). Calibrations of Low-Cost Air Pollution Monitoring Sensors for CO, NO_2, O_3, and SO_2. *SENSORS*, *21*(1), 256. https://doi.org/10.3390/s21010256

Holstius, D. M., Pillarisetti, A., Smith, K. R., & Seto, E. (2014). Field calibrations of a low-cost aerosol sensor at a regulatory monitoring site in California. *Atmospheric Measurement Techniques*, *7*(4), 1121–1131. https://doi.org/10.5194/amt-7-1121-2014

Kim, J., Shusterman, A., Lieschke, K., Newman, C., & Cohen, R. (2018). The BErkeley Atmospheric CO2 Observation Network: Field calibration and evaluation of low-cost air quality sensors. *ATMOSPHERIC MEASUREMENT TECHNIQUES*, *11*(4), 1937–1946. https://doi.org/10.5194/amt-11-1937-2018

Levy Zamora, M., Buehler, C., Datta, A., Gentner, D. R., & Koehler, K. (2023). Identifying optimal co-location calibration periods for low-cost sensors. *Atmospheric Measurement Techniques*, *16*(1), 169–179. https://doi.org/10.5194/amt-16-169-2023

Li, J., Hauryliuk, A., Malings, C., Eilenberg, S. R., Subramanian, R., & Presto, A. A. (2021). Characterizing the Aging of Alphasense NO2 Sensors in Long-Term Field Deployments. *ACS Sensors*, *6*(8), 2952–2959. https://doi.org/10.1021/acssensors.1c00729

Mukherjee, A., Brown, S., McCarthy, M., Pavlovic, N., Stanton, L., Snyder, J., D'Andrea, S., & Hafner, H. (2019). Measuring Spatial and Temporal PM2.5 Variations in Sacramento, California, Communities Using a Network of Low-Cost Sensors. *SENSORS*, *19*(21). https://doi.org/10.3390/s19214701

Okorn, K., & Hannigan, M. (2021). Improving Air Pollutant Metal Oxide Sensor Quantification Practices through: An Exploration of Sensor Signal Normalization, Multi-Sensor and Universal Calibration Model Generation, and Physical Factors Such as Co-Location Duration and Sensor Age. *Atmosphere*, *12*(5), Article 5. https://doi.org/10.3390/atmos12050645

Papapostolou, V., Zhang, H., Feenstra, B. J., & Polidori, A. (2017). Development of an environmental chamber for evaluating the performance of low-cost air quality sensors under controlled conditions. *Atmospheric Environment*, *171*, 82–90. https://doi.org/10.1016/j.atmosenv.2017.10.003

Pinto, J., Dibb, J., Lee, B., Rappenglück, B., Wood, E., Levy, M., Zhang, R., Lefer, B., Ren, X., Stutz, J., Tsai, C., Ackermann, L., Golovko, J., Herndon, S., Oakes, M., Meng, Q., Munger, J., Zahniser, M., & Zheng, J. (2014). Intercomparison of field

measurements of nitrous acid (HONO) during the SHARP campaign. *JOURNAL OF GEOPHYSICAL RESEARCH-ATMOSPHERES*, *119*(9), 5583–5601. https://doi.org/10.1002/2013JD020287

Si, M., Xiong, Y., Du, S., & Du, K. (2020). Evaluation and calibration of a low-cost particle sensor in ambient conditions using machine-learning methods. *Atmospheric Measurement Techniques*, *13*(4), 1693–1707. https://doi.org/10.5194/amt-13-1693-2020

Sousan, S., Koehler, K., Hallett, L., & Peters, T. M. (2016). Evaluation of the Alphasense optical particle counter (OPC-N2) and the Grimm portable aerosol spectrometer (PAS-1.108). *Aerosol Science and Technology*, *50*(12), 1352–1365. https://doi.org/10.1080/02786826.2016.1232859

Spinelle, L., Gerboles, M., Villani, M. G., Aleixandre, M., & Bonavitacola, F. (2015). Field calibration of a cluster of low-cost available sensors for air quality monitoring. Part A: Ozone and nitrogen dioxide. *Sensors and Actuators B: Chemical*, *215*, 249–257. https://doi.org/10.1016/j.snb.2015.03.031

Spinelle, L., Gerboles, M., Villani, M. G., Aleixandre, M., & Bonavitacola, F. (2017). Field calibration of a cluster of low-cost commercially available sensors for air quality monitoring. Part B: NO, CO and CO2. *Sensors and Actuators B: Chemical*, *238*, 706–715. https://doi.org/10.1016/j.snb.2016.07.036

Sun, L., Westerdahl, D., & Ning, Z. (2017). Development and Evaluation of A Novel and Cost-Effective Approach for Low-Cost NO2 Sensor Drift Correction. *Sensors*, *17*(8), 1916. https://doi.org/10.3390/s17081916

Topalovic, D., Davidovic, M., Jovanovic, M., Bartonova, A., Ristovski, Z., & Jovasevic-Stojanovic, M. (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*, 640–658. https://doi.org/10.1016/j.atmosenv.2019.06.028

Wei, P., Sun, L., Anand, A., Zhang, Q., Huixin, Z., Deng, Z., Wang, Y., & Ning, Z. (2020). Development and evaluation of a robust temperature sensitive algorithm for long term NO2 gas sensor network data correction. *Atmospheric Environment*, *230*, 117509. https://doi.org/10.1016/j.atmosenv.2020.117509

Zimmerman, N., Presto, A. A., Kumar, S. P. N., Gu, J., Hauryliuk, A., Robinson, E. S., Robinson, A. L., & R. Subramanian. (2018). A machine learning calibration model using random forests to improve sensor performance for lower-cost air quality monitoring. *Atmospheric Measurement Techniques*, *11*(1), 291–313. https://doi.org/10.5194/amt-11-291-2018