https://doi.org/10.5194/egusphere-2025-6203
Preprint. Discussion started: 26 January 2026


# Enhancing Low-Cost $PM_{2.5}$ Sensor Reliability Through Multi-Model Calibration Against a Beta Attenuation Monitor

Galang Adira Prayoga[1], Puji Lestari[2], Emir Husni[1]

[1]School of Electrical Engineering and Informatics, Institut Teknologi Bandung, Bandung, 40132, Indonesia
[2]Faculty of Civil and Environmental Engineering, Institut Teknologi Bandung, Bandung, 40132, Indonesia

*Correspondence to*: Puji Lestari (puji_L@itb.ac.id)

**Abstract.** Accurate particulate matter ($PM_{2.5}$) monitoring using low-cost sensors requires careful consideration of meteorological influences and calibration against reference instruments. This study evaluates the performance of a low-cost optical sensor through an outdoor co-location experiment with a Beta Attenuation Monitor (BAM 1022). Raw measurements showed strong temporal agreement but substantial overestimation, particularly under high relative humidity, which induced hygroscopic particle growth and amplified light-scattering responses. Correlation and regression analyses confirmed humidity as the dominant environmental factor affecting low-cost sensor bias, while temperature exhibited only minor influence. To address these limitations, multiple calibration models (including Linear Regression, Random Forest, Gradient Boosting, Support Vector Regression, and an Adaptive-blend ensemble) were developed and assessed. Nonlinear and ensemble-based models significantly improved accuracy, reducing MAE from 17.40 µg/m³ (uncalibrated) to 5.85 µg/m³ after calibration. These findings demonstrate the necessity of environmental compensation and model-based correction for reliable low-cost $PM_{2.5}$ monitoring and support their integration into high-resolution air quality networks.

***Keywords***—*Particulate Matter (PM); Air Temperature; Relative Humidity; Meteorological Factors; Statistical Analysis; Air Quality Monitoring*

## I. INTORDUCTION

Air pollution has emerged as one of the most pressing environmental and public health challenges of the 21st century, driven by rapid urbanization, industrial expansion, and the escalating impacts of climate change (Edwards et al., 2025; Schneider et al., 2020). Among various air pollutants, fine particulate matter, particularly $PM_{2.5}$ and $PM_{10}$, poses the most severe risk to human health and environmental sustainability (Chacón-mateos et al., 2025; Orellano et al., 2020). These microscopic particles originate from both natural sources, such as dust and biomass burning, and anthropogenic activities, including vehicular emissions, industrial combustion, and construction processes (Orellano et al., 2020). Due to their small aerodynamic diameter, particulate matter can penetrate deep into the respiratory system and even enter the bloodstream, leading to a range of adverse health effects such as respiratory infections, cardiovascular diseases, and increased premature mortality (Pouri et al., 2024). Beyond human health, elevated concentrations of particulate matter also contribute to

https://doi.org/10.5194/egusphere-2025-6203
Preprint. Discussion started: 26 January 2026


atmospheric visibility degradation, alter radiative balance, and influence local and regional climate dynamics, thereby underscoring the urgent need for continuous monitoring and comprehensive understanding of PM behavior under varying environmental conditions (Orellano et al., 2020; Pouri et al., 2024).

However, air pollution extends far beyond its health implications; it also exerts profound environmental and ecological impacts. Elevated levels of atmospheric pollutants contribute to phenomena such as acid rain, reduced visibility, ecosystem degradation, and climate alteration through the modification of radiative and cloud-forming processes (Bolan et al., 2024). Although these broader environmental consequences are not the primary focus of this study, they reinforce the necessity of understanding the mechanisms governing particulate matter dynamics. In particular, environmental parameters such as ambient air temperature and relative humidity play a critical role in influencing the formation, transformation, and dispersion of $PM_{2.5}$ and $PM_{10}$ particles (Jayaratne et al., 2018; J. Wang & Ogawa, 2015; Zender-Świercz et al., 2024). Investigating how these meteorological factors affect particulate matter concentrations is essential for developing more accurate air quality models, improving sensor calibration, and formulating effective strategies to safeguard public health and environmental quality (Mahajan & Helbing, 2025; Venkata et al., 2024).

In addition to its health and environmental implications, air pollution imposes substantial economic burdens on society. It contributes to decreased agricultural productivity, accelerates the deterioration of infrastructure, and increases public health expenditures due to pollution-related illnesses (Lanzi et al., 2020). These combined effects translate into significant financial losses at both local and national levels, underscoring the broad socio-economic relevance of air quality management. Consequently, investigating the influence of meteorological factors on air pollutant behavior is not only of scientific interest but also of practical importance. A deeper understanding of how temperature and humidity affect particulate matter concentrations can support more effective urban planning, environmental policy development, and adaptive response strategies to mitigate air pollution in cities and regions worldwide (Galiszewska et al., 2024).

Numerous studies worldwide have examined the influence of meteorological parameters such as particularly air temperature, wind speed, and relative humidity, on the concentration of particulate matter ($PM_{2.5}$ and $PM_{10}$) (Mahajan & Helbing, 2025; P. Wang et al., 2021; Zender-Świercz et al., 2024). These studies consistently highlight that humidity and temperature play a critical role in shaping the physical behavior of airborne particles, affecting their size distribution, hygroscopic growth, and atmospheric residence time. As summarized in Table 1, relative humidity often leads to overestimation in optical sensors due to light scattering by water-absorbing particles, while temperature influences both particle dispersion and sensor response. Although extensive research has been conducted to characterize these meteorological effects, most previous works have focused on ambient PM variability rather than on how these factors specifically influence measurement accuracy across different sensor technologies.

The key research gap lies in understanding how temperature and humidity differentially affect low-cost optical particle sensors compared with reference-grade instruments such as the Beta Attenuation Monitor (BAM). Low-cost sensors are

https://doi.org/10.5194/egusphere-2025-6203
Preprint. Discussion started: 26 January 2026


widely adopted for dense air quality monitoring networks due to their affordability and portability, yet their performance under varying environmental conditions remains inconsistent. Differences in sensor design, optical principles, and lack of environmental compensation models often lead to deviations from reference measurements, particularly under high humidity or temperature fluctuations. Therefore, this study aims to systematically analyze and quantify the influence of ambient temperature and relative humidity on particulate matter readings obtained from low-cost sensors relative to BAM reference instruments. The outcome of this analysis will provide insights necessary for developing more robust calibration and environmental compensation frameworks, ultimately improving the reliability of low-cost sensor data for air quality assessment and public health applications.

**Tabel 1. Summary of the existing literature on the impact of meteorological parameters and air pollution by particulate matter.**

| Reference | Focus | Meteorological Parameters | Findings | Impact |
|---|---|---|---|---|
| Evaluation and Calibration of a Low-Cost Particle Sensor in Ambient Condition Using ML (2020) (Si et al., 2020) | Sensor calibration using machine learning under varying T and RH | Temperature, Relative Humidity | ML models (e.g., RF, ANN) significantly improved PM2.5 estimates by including RH and T; humidity strongly affects scattering response. | Demonstrated need for environmental compensation in low-cost sensors. |
| Improved PM2.5 Concentration Estimates from Low-Cost Sensors Using Calibration Models Categorized by RH (2021) (Hua et al., 2021) | Categorical calibration model by RH levels | Relative Humidity | Separate calibration models by RH improved accuracy; PM2.5 overestimation at high RH confirmed. | Introduced RH-based segmentation as an effective compensation strategy. |
| The Effect of Temperature and Humidity of Air on the Concentration of Particulate Matter (2024) (Zender-Świercz et al., 2024) | Correlation study between meteorological variables and PM | Temperature, Humidity | Temperature negatively correlated with PM concentration, humidity showed positive correlation especially at RH > 70%. | Revealed nonlinear RH influence; emphasized local climatic variability. |
| The Influence of Humidity on the Performance of a Low-Cost Air Particle Mass Sensor (2018) (Jayaratne et al., 2018) | Humidity influence on sensor readings | Relative Humidity | Sensor readings increase exponentially with RH; signal distortion due to hygroscopic particle growth. | Quantified humidity artifacts; supports RH correction necessity. |
| Determining the Correlation Between PM10 and Meteorological Factors (2022) (Kirešová & Guzan, 2022) | Statistical analysis of PM10â€“weather relationships | Temperature, Humidity, Wind Speed, Precipitation, Pressure | Negative correlation with wind speed and temperature; positive correlation with RH; rainfall reduces PM10. | Validated classical meteorological influence patterns on PM. |
| Diurnal and Daily Variation of PM2.5 and Its Multiple Wavelet Coherence with Meteorological Variables (2024) (Cholianawati et al., 2024) | Wavelet-based temporal analysis of PM2.5 vs meteorological factors | Temperature, Humidity, Wind Speed, Pressure | Strong coherence between PM2.5 and RH at multi-day scales; temperature influence varies by season. | Showed complex, scale-dependent meteorological influence using wavelet methods. |
| Effect of an Aerosol Dryer on Ambient PM Measurement with SDS011 Low-Cost Sensor (2023) (Nothhelfer et al., 2023) | Mitigating humidity effects using aerosol dryer | Relative Humidity | Dryer reduces RH-related overestimation; calibrated readings align with reference monitors. | Demonstrated practical mitigation strategy for RH interference. |
| Effect of Relative Humidity on the Performance of Five Cost-Effective PM Sensors (2021) (P. Wang et al., 2021) | Comparative RH sensitivity of multiple sensors | Relative Humidity | All sensors show overresponse above 70% RH; degree varies by sensor type. | Established inter-sensor variability and the need for RH-dependent correction. |
| Effects of Meteorological Conditions on PM2.5 Concentrations in Nagasaki, Japan (2015) (J. Wang & Ogawa, 2015) | Long-term observation of PM2.5 and meteorology | Temperature, Humidity, Wind Speed, Pressure | Wind speed and precipitation reduce PM2.5; RH and temperature show complex seasonal patterns. | Highlighted regional and seasonal dependence of meteorological effects. |
| Impact Analysis of Temperature and Humidity Condition on Electrochemical Sensor Response in Ambient Air (2018) (Wei et al., 2018) | Impact of T and RH on gas and PM sensors | Temperature, Humidity | Temperature and RH significantly alter baseline and sensitivity; compensation models improve stability. | Validated environmental correction across multiple sensor types. |

https://doi.org/10.5194/egusphere-2025-6203
Preprint. Discussion started: 26 January 2026


## II. MATERIAL AND METHODS

### 2.1 Sensor Particulate Matter Low-Cost vs Reference BAM

In this study, two types of particulate matter (PM) measurement instruments were utilized: the low-cost optical sensor Winsen ZH03A and the reference-grade Beta Attenuation Monitor (BAM 1022, Met One Instruments Inc.). The Winsen ZH03A operates based on the principle of laser light scattering, where airborne particles passing through a laser beam cause scattering proportional to their size and concentration (Dong et al., 2025). This compact sensor is widely used in low-cost air quality monitoring networks due to its affordability (approximately USD 16), portability, and digital output compatibility with microcontrollers. It measures particles as small as 0.3 µm within a concentration range of 0–1000 µg $m^{-3}$, operates at 5 V DC with less than 120 mA current draw, and performs reliably within ambient conditions of –10 °C to +50 °C and 0–85 % relative humidity (Hapidin et al., 2019). However, because it relies on optical scattering, its readings are highly sensitive to environmental factors such as humidity, which can cause hygroscopic growth of particles and lead to signal overestimation under high RH conditions (Mahajan & Helbing, 2025; Zender-Świercz et al., 2024).

In contrast, the BAM 1022 functions as a federal equivalent method (FEM) instrument that measures particulate concentration using the beta-ray attenuation principle, providing traceable and reference-grade data (Shukla & Aggarwal, 2022). The device draws a continuous airflow through a filter tape, where particles are collected and subsequently irradiated by a beta source. The degree of beta-particle attenuation is proportional to the mass of particles on the filter, yielding precise and humidity-compensated PM measurements. The BAM 1022 can measure particulate concentrations over a broad range (-15 to 10 000 µg $m^{-3}$ for $PM_{2.5}$), with integrated temperature control to mitigate moisture effects and maintain stability in diverse environmental conditions. Unlike low-cost sensors, BAM instruments require regular maintenance and calibration but provide high-accuracy, regulatory-compliant data for air quality assessments.

Figure 1 presents a side-by-side comparison between the reference-grade particulate matter sampler BAM 1022 and the Low-Cost Sensor (LCS) AQMS used in this study. The comparison highlights differences in physical design, airflow mechanisms, and environmental shielding for temperature and humidity measurements.

The BAM 1022 (Figure 1(a)) is a U.S. EPA-approved Federal Equivalent Method (FEM) instrument for continuous $PM_{2.5}$ and $PM_{10}$monitoring. It operates based on the beta attenuation principle, in which air is drawn through a filter tape and the particle mass is quantified by measuring the attenuation of beta radiation. The BAM unit has a vertical configuration with a total height of approximately 139.7 cm, a footprint width of 55.9 cm, and a depth of 45.1 cm. The core components include the PM inlet head, virtual impactor/size-selective cyclone (VSCC), condensation jar, nozzle block, beta source, and the automatic filter tape transport mechanism. Figure 1(b) illustrates the internal airflow mechanism of BAM. Ambient air enters through the PM inlet, passes through size-selective separation (e.g., VSCC for $PM_{2.5}$), and is then drawn across the filter tape by an internal pump. After measurement, the processed air exits through an outlet pathway. This controlled, isokinetic airflow ensures accurate mass concentration measurement under varying environmental conditions. The BAM uses an external multi-

https://doi.org/10.5194/egusphere-2025-6203
Preprint. Discussion started: 26 January 2026


plate radiation shield (Figure 1(c)) to house the temperature and relative humidity sensors. The shield minimizes solar heating effects, ensuring that the environmental parameters used for RH correction and QC checks remain accurate.

The LCS AQMS system (Figure 1(d)) is a compact, solar-powered air quality monitoring unit integrating low-cost particulate matter sensors. This system is designed for high-density networks with significantly lower deployment costs than reference instruments. The LCS enclosure has a height of 45 cm, a width of 29 cm, and a depth of 27 cm, making it considerably more compact than the BAM. The system includes a microcontroller, PM sensor module, gas sensors, lithium battery, and charge controller. A solar panel is mounted externally for autonomous outdoor operation. Figure 1(e) shows the airflow pathway inside the LCS. Ambient air enters through a dedicated air inlet channel, passes across the optical particulate matter sensor and exits via a designated outlet. The airflow relies on natural convection or a micro-fan, depending on the design, allowing continuous sensing with minimal power consumption. As with BAM, the LCS includes a radiation shield (Figure 1(f)) to protect temperature and humidity sensors from direct solar radiation. The shield consists of stacked plates that allow adequate airflow while minimizing thermal bias, ensuring accurate compensation of LCS PM signals for environmental effects.

Table 2 summarizes the technical specifications of the low-cost $PM_{2.5}$ sensors used in this study (GP2Y, ZH03, and SDS011) and compares them with the U.S. EPA Federal Equivalent Method (FEM) reference instrument BAM 1022. The comparison highlights key differences in detection principles, measurement capability, operational limits, and practical considerations relevant to field deployment and calibration. The three LCS devices rely on optical light-scattering, either using IR-LED (GP2Y) or laser diodes (ZH03 and SDS011). Laser-based sensors generally provide higher sensitivity and lower minimum detectable particle size (~0.3 µm). In contrast, the BAM 1022 utilizes beta attenuation, a reference-grade gravimetric surrogate method that measures the mass collected on a filter tape by detecting attenuation of beta radiation. Unlike optical sensors, the BAM does not specify a minimum optical particle size threshold because it directly measures mass.

The concentration range and detection capability of low-cost PM sensors (LCSs) are notably limited compared to reference instruments, with ZH03 operating between 0–1,000 µg $m^{-3}$ and SDS011 between 0.0–999.9 µg $m^{-3}$, while the GP2Y does not output concentration values and provides only proportional analog signals. In contrast, the BAM 1022 offers a substantially wider dynamic range (−15 to 10,000 µg $m^{-3}$), enabling accurate mass quantification under both clean-air and high-pollution conditions. From an electrical and operational standpoint, LCS devices function at low voltage (5 ± 0.1 V) and low current, making them compatible with battery-powered and IoT-based monitoring systems, whereas the BAM requires a higher AC power supply (100–240 VAC), limiting its use to stationary installations. Their environmental operating limits also differ, as LCSs support broader temperature ranges (−10 to 50°C) with humidity tolerance up to 70–85%, while the BAM operates within 30 to 50°C but accommodates higher humidity up to 90% RH under non-condensing conditions within its regulated measurement chamber. In terms of signal output, the GP2Y relies solely on analog voltage, whereas ZH03 and SDS011 use digital communication for improved stability; the BAM provides optically isolated analog outputs (0–1, 0–2.5, 0–5 VDC) coupled with high-resolution internal logging. Physical size, lifetime, and cost further differentiate these devices: LCSs are

https://doi.org/10.5194/egusphere-2025-6203
Preprint. Discussion started: 26 January 2026


compact (46–71 mm wide and 17–73 mm high), with lifetimes ranging from ~3 years (ZH03) to ~8,000 hours for SDS011, while the BAM is a large cabinet-style instrument (559 × 457 × 1397 mm) with filter-tape–dependent operational lifetimes exceeding 2.6 years at low sampling frequency. Cost differences are substantial, with LCSs priced between ~$8 and ~$25.9, compared to the BAM's cost of approximately $5,500, underscoring the practicality of LCS units for dense monitoring networks and community-scale deployments, while the BAM remains reserved for regulatory-grade air quality monitoring.

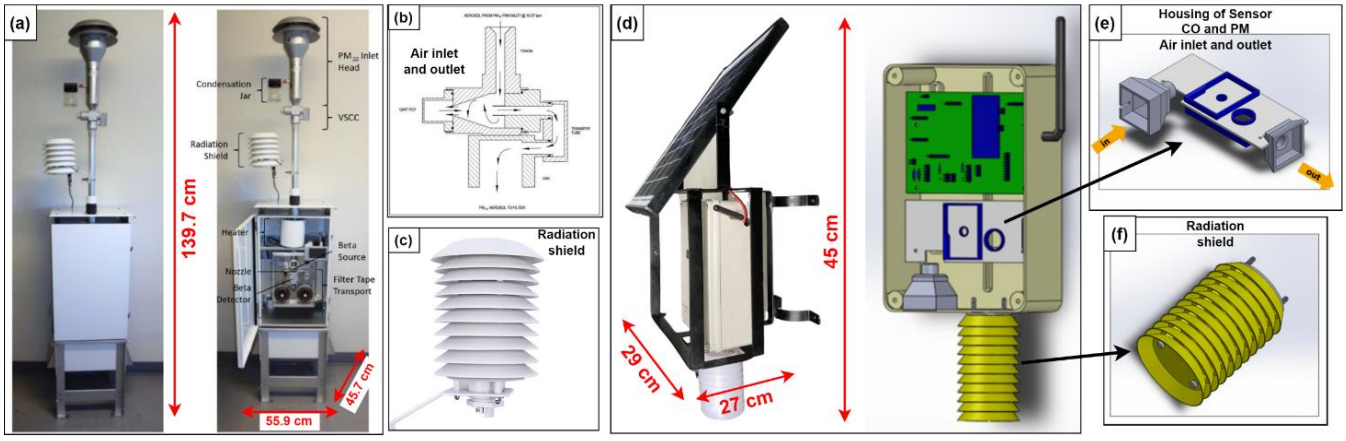


**Figure 1: Comparison (a) reference sensor BAM 1022 device with (b) inlet and outlet air mechanism in BAM, and (c) radiation shield temperature and humidity measurement of BAM, vs (d) Low-Cost Sensor (LCS) AQMS with (e) inlet and outlet air mechanism in LCS AQMS, and (f) radiation shield temperature and humidity of LCS AQMS.**

**Tabel 2. Low-cost sensor PM2.5 vs reference sensor BAM 1022 specification**

| Abbreviation | GP2Y | ZH03 (Winsen, 2016) | SDS0 | BAM (Instruments, 2018) |
|---|---|---|---|---|
| Model | Sharp GP2Y1010-AU0F | Winsen ZH03A | Novafitness SDS | Met One Instruments BAM 1022 |
| Light source | IR-LED | Laser | Laser | Beta attenuation (beta-ray attenuation across filter tape) |
| Minimum particle size ($\mu m$) | N/A | 0.3 | 0.3 | not typically specified as optical threshold for BAM |
| Concentration range ($\mu g m^{-3}$) | N/A | 0 to 1000 | 0-999.9 | -15 to 10.000 |
| Rated voltage ($V$) | $5 \pm 0.5$ | $5 \pm 0.1$ | $4.7 - 5.3$ | 100-240 VAC 50/60 Hz universal input; 12 VDC, 8.5A. |
| Rated current ($mA$) | < 20 | < 120 | $70 \pm 10$ | - |
| Dimension WxHxD ($mm^3$) | 46 x 30 x 17.6 | 50 x 32.4 x 21 | 71 x 70 x 23 | 559 x 457 x 1397 |
| Operating temperature (°C) | -10 to + 65 | -10 to + 50 | -20 to +50 | -30 to +50 |
| Operating humidity (%) | N/A | 0 to 85 | 0 – 70 | 0 to 90, non-condensing. |
| Output signal | Analog | Digital | Digital | Analog output: Two channels; optically isolated; 0-1, 0-2.5, 0-5 VDC. |
| Lifetime | N/A | 3 years | 8000 hours | NA (Data logger memory 2.6 years @ 1 record/hr.; 15.6 days @ 1 record/min) |
| Cost (USD) | ~$8 | ~$16 | ~$25.90 | ~$ 5.500 (online platform) (not publicly disclosed) |

https://doi.org/10.5194/egusphere-2025-6203
Preprint. Discussion started: 26 January 2026


## 2.2 Architecture System LCS AQMS vs BAM AQMS

The low-cost air quality monitoring unit developed in this study, referred to as the AQUHIs system (Prayoga et al., 2025), was designed to provide autonomous and reliable measurement of particulate matter ($PM_{2.5}$) alongside meteorological parameters such as temperature and relative humidity. Figure 2 compares the overall system architectures of the Low-Cost Sensor Air Quality Monitoring System (LCS AQMS) and the reference-grade BAM 1022. The diagrams illustrate the differences in power supply, sensing modules, air-handling mechanisms, signal acquisition, and data-processing pathways between the two types of monitoring systems.

Figure 2(a) shows a compact and energy-efficient architecture built around an ESP32 microcontroller, which integrates wireless communication modules (Wi-Fi, GPRS, LoRaWAN) and interfaces with multiple sensors through ADC, I²C, SPI, and TWI communication buses. The system draws power from a 10 W photovoltaic panel, regulated through MPPT and BMS circuits, and stored in a rechargeable lithium battery, enabling autonomous operation in outdoor environments. The particulate matter sensor (Winsen ZH03) samples air via a simple inlet and micro-pump pathway, while temperature and humidity are measured using an SHT31 waterproof sensor housed in a radiation shield. A voltage regulation module ensures stable power delivery to the sensors and microcontrollers. The data acquisition pathway is digital, enabling real-time storage (MicroSD), processing, and wireless transmission. Electrical supply, data flow, and air flow pathways are highlighted to show the tightly integrated, low-power design suitable for distributed sensor networks.

Figure 2(b) depicts the more complex architecture of the BAM 1022, which relies on precision hardware components and regulated environmental control to achieve reference-grade measurements. The system is powered by an external 100–230 VAC power supply unit, reflecting its stationary installation requirements. Air is drawn through a mass-flow–controlled inlet, passing through a size-selective $PM_{2.5}/PM_{10}$ cyclone and then through a glass-fiber filter tape that advances hourly. A vacuum pump and flow controller maintain strict volumetric flow conditions, while an inlet heater and in-situ temperature control minimize humidity-induced measurement errors. Core measurement is achieved through a beta attenuation subsystem, where a Carbon-14 beta source irradiates the filter tape, and a beta detector quantifies particle mass accumulation using continuous attenuation readings. Environmental parameters are monitored by integrated temperature, humidity, and pressure sensors. A high-resolution MCU controller with multi-channel ADC (16–24 bit) processes beta pulses, regulates the tape stepper motor, and stores data locally on SD/flash memory. Communication interfaces include UART, USB, and wireless modems (Wi-Fi/4G/LoRaWAN/NB-IoT).

Figure 3 illustrates the end-to-end data processing pipelines for the Low-Cost Sensor Air Quality Monitoring System (LCS AQMS) and the reference BAM 1022. The comparison highlights fundamental differences in data acquisition, preprocessing, validation, communication pathways, and backend integration between low-cost optical sensing systems and regulatory-grade beta attenuation monitors.

https://doi.org/10.5194/egusphere-2025-6203
Preprint. Discussion started: 26 January 2026


Figure 3(a) shows that the LCS AQMS begins its data pipeline at the air sampling stage, where $PM_{2.5}$/$PM_{10}$, temperature, and humidity are recorded by the sensing layer. Raw optical scattering signals are converted into particulate matter values, followed by onboard preprocessing such as smoothing, averaging, and timestamping. The processed data are then assembled into compact transmission packets (≤ 250 bytes) and subjected to reliability checks including RSSI, SNR, and survival rate evaluation.

The system dynamically selects a communication pathway depending on network availability. If a GPRS network is accessible, the device transmits data through a GPRS module to the gateway server; otherwise, the system switches to LoRaWAN transmission (SF10, BW125 kHz, 14 dBm) through the nearest LoRa gateway. Regardless of the chosen network, data are ultimately delivered to a cloud server database and API, where they are archived, visualized, and analyzed on dashboards or analytics platforms. This lightweight, adaptive data flow is optimized for intermittent connectivity, low-power operation, and integration into distributed sensing networks.

Figure 3(b) illustrates the more complex and controlled data pipeline of the BAM 1022. Air is drawn through the inlet and deposited onto a filter tape, where mass accumulation is quantified via the beta attenuation measurement process. Signals from the beta detector are converted into raw PM mass readings and then undergo the same preprocessing steps smoothing, averaging, and timestamping as seen in the LCS pipeline. The BAM continuously validates its measurements by performing internal diagnostic checks; if anomalies or errors are detected, data are flagged and maintenance alerts are logged and transmitted.

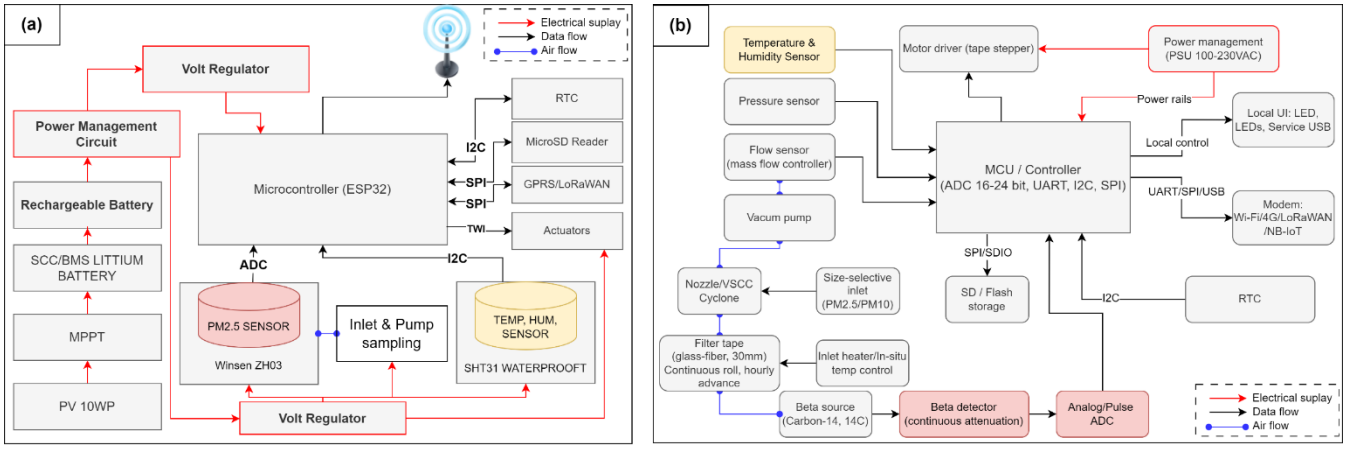


**Figure 2. The overall architecture of the system (a) LCS AQMS vs (b) Reference BAM 1022**

After data validation, measurements are formatted into structured outputs (JSON, CSV, Modbus), followed by application of security and protection protocols such as TLS encryption or token-based authentication. The BAM supports multiple communication interfaces, including Wi-Fi, 3G/4G/NB-IoT, and other LPWA options, enabling transmission to backend

https://doi.org/10.5194/egusphere-2025-6203
Preprint. Discussion started: 26 January 2026


servers. If network outages occur, the BAM stores data locally with ACK & retry mechanisms to ensure no data are lost. Telemetry and maintenance logs are also transmitted to support long-term reliability and regulatory compliance.

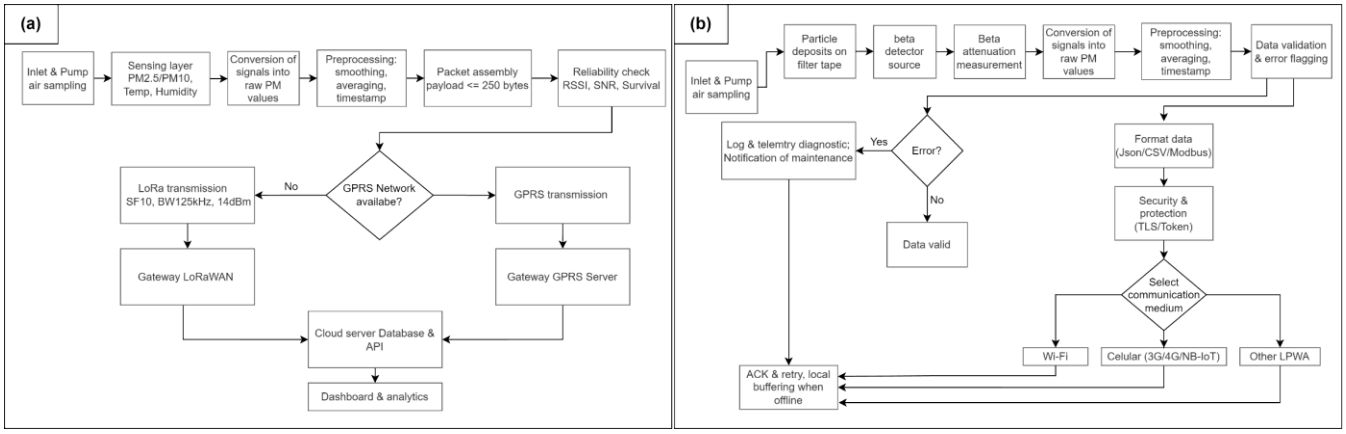


**Figure 3. Data flow of (a) LCS AQMS vs (b) Reference BAM**

### 2.3 Co-location Setup and Data Collection

The overall co-location and data collection method shown in Figure 4. The overall workflow for the co-location experiment and data collection process was used to evaluate the performance of the LCS AQMS against the reference Beta Attenuation Mass (BAM) monitor. The methodology consists of three main stages: (1) system deployment and network registration, (2) co-location with a regulatory-grade reference instrument, and (3) acquisition, preprocessing, and statistical analysis of the sensor data.

In the first stage (Figure 4, Step 1), the LCS AQMS device is designed, assembled, and installed at the monitoring site. Each LCS unit is registered on the communication network either LoRaWAN or GPRS, depending on signal availability. Once active, the device transmits measurement packets (including $PM_{2.5}$, temperature, humidity, VOC, and battery voltage) at intervals of 15–20 minutes. The packets travel through nearby LoRa/GPRS gateways and are forwarded to the IoT platform. Raw sensor data are then stored in the AQMS cloud database, forming the primary dataset for preprocessing.

Figure 4, Step 2 illustrates the co-location setup in which the LCS AQMS is placed in close proximity to a BAM reference station to ensure that both systems measure the same ambient air conditions. This side-by-side configuration enables direct comparison and calibration between the LCS and the reference instrument. The LCS transmits data every few minutes, while the BAM generates validated measurements on an hourly cycle, based on beta attenuation across the filter tape. Both data streams are stored in their respective databases and periodically synchronized to ensure temporal alignment during analysis.

https://doi.org/10.5194/egusphere-2025-6203
Preprint. Discussion started: 26 January 2026


In the final stage (Figure 4, Step 3), data from both the LCS and BAM databases undergo a unified data preprocessing pipeline. This includes timestamp matching, averaging, smoothing, outlier removal, and integrity checks. After preprocessing, paired datasets from both systems are used in the statistical analysis module, where variables such as $PM_{2.5}$ concentration, temperature, humidity are compared. Time-series plots and statistical metrics (e.g., RMSE, bias, correlation) are generated to quantify differences and evaluate sensor performance under real environmental conditions.

The LCS AQMS instruments positioned approximately 2 meters above ground level to ensure they sampled the same ambient air and shared identical meteorological exposure. The AQUHIs unit was mounted on a rigid aluminum frame, and all measurements were synchronized using timestamps from its integrated Real-Time Clock (RTC). Prior to analysis, the datasets were preprocessed through unit conversion, timestamp alignment, outlier removal, and temporal averaging to match the BAM's hourly measurement resolution, enabling direct comparison between the two systems. The co-location campaign spanned multiple days and captured a wide range of temperature and humidity conditions that could influence sensor responses, and the resulting harmonized dataset provided the basis for statistical and regression analyses aimed at quantifying environmental effects and evaluating the agreement between the low-cost sensor and the reference BAM measurements.

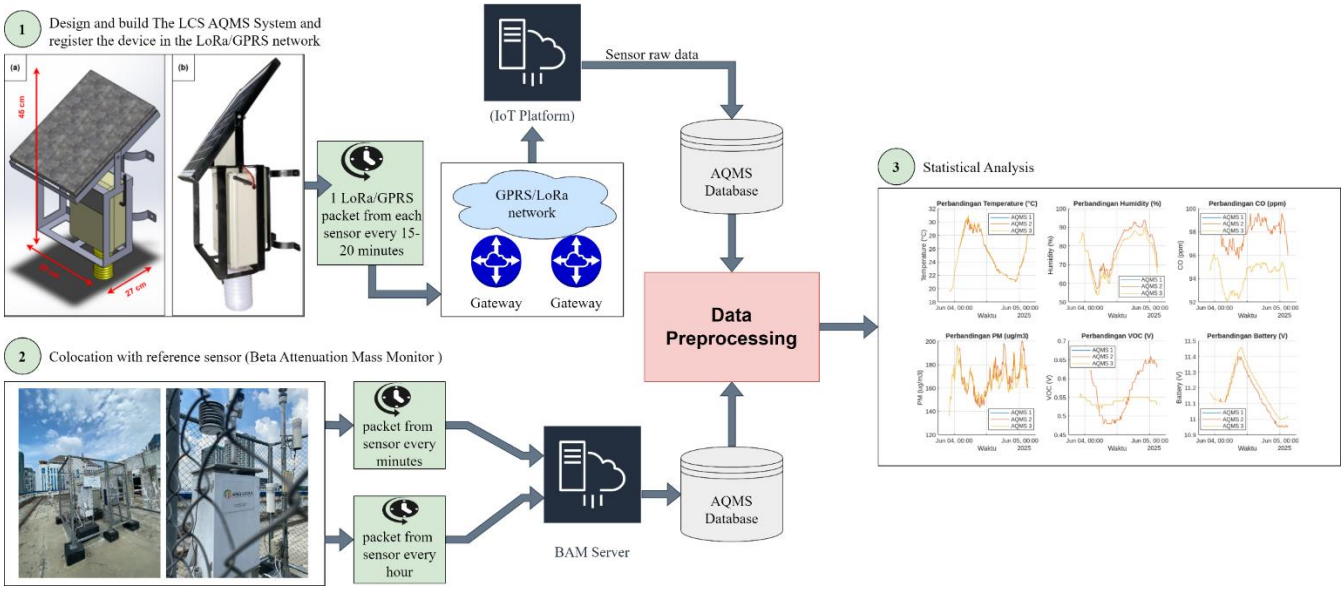


**Figure 4. The overall co-location and data collection method**

https://doi.org/10.5194/egusphere-2025-6203
Preprint. Discussion started: 26 January 2026


### 2.4 Data Preprocessing and calibration models

Figure 5 illustrates the workflow used to develop, train, and evaluate calibration functions for improving the accuracy of the low-cost particulate matter sensor. The process comprises three main stages: data preprocessing, regression modeling, and independent validation using subsequent co-location datasets.

In the first stage, the raw data collected from both the AQUHIs system and the BAM 1022 reference instrument underwent a structured preprocessing workflow prior to statistical analysis. All measurements transmitted from the AQUHIs units via the GPRS/LoRaWAN network were first aggregated and archived within the AQMS database. The dataset was then inspected to identify and remove incomplete or corrupted records arising from packet loss, sensor start-up instability, or power interruptions.

Following this initial screening, outliers were identified and removed to ensure consistency between the two monitoring systems. Outlier detection employed a Z-score–based approach (DeVore, 2017) in which both independent and dependent variables were standardised according to Eq. (1):

$$Z = \frac{X-\mu}{\sigma}, \quad (1)$$

where $X$ is the measured value, $\mu$is the mean, and $\sigma$is the standard deviation. Data points exceeding $| Z |> 3$ were classified as potential outliers and excluded from subsequent modelling. Although the raw dataset did not contain severe anomalies such as negative readings or flatlines, several isolated concentration spikes were observed. While these may reflect true short-term pollution events, their rarity can bias regression estimates; therefore, they were flagged and removed during the cleaning process. For variables with different physical units (e.g., temperature, humidity, particulate concentration), Z-score normalisation was applied to ensure a standardised scale for model fitting.

To enable direct comparison with the BAM 1022 reference instrument, AQUHIs data, originally transmitted at 15–20 minute intervals were resampled to hourly averages. Time alignment between the two datasets was achieved through timestamp matching using each device's Real-Time Clock (RTC), ensuring that paired observations reflected identical atmospheric conditions. Hourly averaging also mitigates small spatiotemporal discrepancies that may arise even under close colocation.

In the second stage, the calibration of low-cost $PM_{2.5}$ sensors is a critical step for improving data reliability, particularly when such sensors are deployed alongside reference-grade instruments such as BAM. In this work, we implemented and evaluated several calibration approaches, ranging from traditional statistical techniques to contemporary machine-learning models. Each method addresses different types of sensor errors offset bias, scale distortion, temporal drift, and nonlinear sensitivity to environmental factors (e.g., temperature and humidity).

https://doi.org/10.5194/egusphere-2025-6203
Preprint. Discussion started: 26 January 2026


The uncalibrated output represents the raw measurement produced by the low-cost sensor without any correction. This serves as the baseline against which all calibrated models are compared. Raw $PM_{2.5}$ readings typically exhibit several types of errors, including systematic bias, nonlinear response across concentration ranges, sensitivity to relative humidity, and temporal instability. Although uncalibrated data can capture general pollution trends, the magnitude of error relative to reference instruments is often substantial, making calibration essential for quantitative applications.

The first calibration approach applies a linear affine correction to remove systematic offset and scale errors (Aix et al., 2023). A simple linear regression is fitted between the raw low-cost sensor reading $PM_{raw}$ and the reference measurement $PM_{ref}$

$$PM_{ref} = \alpha + \beta PM_{raw}. \quad (2)$$

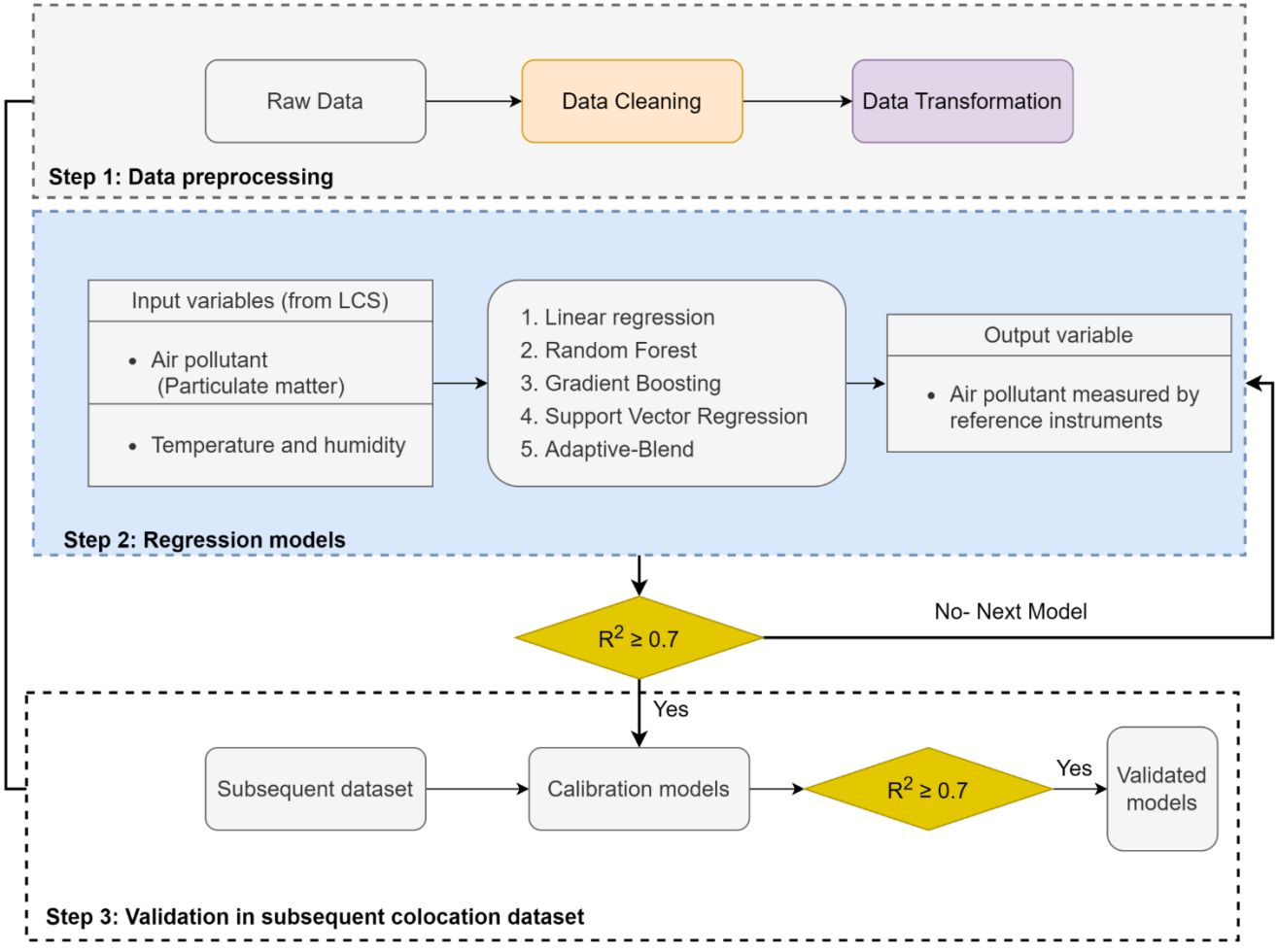


**Figure 5. Flow chart of developing and evaluating calibration functions**

https://doi.org/10.5194/egusphere-2025-6203
Preprint. Discussion started: 26 January 2026


From this model, a corrected sensor signal (sensor_corr) is derived through an offset–scale adjustment:

$$\mathrm{PM}_{\mathrm{corr}} = \beta(\mathrm{PM}_{\mathrm{raw}} - a_s),\ a_s = -\frac{\alpha}{\beta} \tag{3}$$

In this formulation, $a_s$ represents the estimated sensor offset (i.e., the raw reading at which the reference concentration is predicted to be zero), while $\beta$ adjusts the sensitivity of the sensor. This transformation forces the corrected signal to have a unity slope with respect to the reference, thereby removing consistent additive and multiplicative biases. Linear calibration is computationally efficient, interpretable, and provides substantial improvement when sensor errors are predominantly linear; however, it cannot capture nonlinear dynamics or complex environmental influences.

To address nonlinear relationships, ensemble learning methods were employed. Random Forest (RF) constructs an ensemble of decision trees and aggregates their predictions to model nonlinear interactions (Ma et al., 2025) between the sensor output and environmental variables such as temperature, humidity, diurnal cycle, and the linearly corrected signal $PM_{\mathrm{corr}}$. Because RF is non-parametric and robust to noisy data, it effectively captures complex yet stable patterns in sensor behaviour. However, tree-based methods partition the feature space rather than learning smooth functions, which can limit performance when extrapolating beyond the training range.

In addition to RF, Gradient Boosting (GB) was employed to further enhance the model's ability to capture subtle, structured error patterns (Si et al., 2020). Unlike RF, which builds trees independently and aggregates their outputs, Gradient Boosting constructs trees sequentially, where each new tree is trained to correct the residual errors of the previous ensemble. This boosting strategy allows GB to approximate complex nonlinear relationships with greater precision, particularly in scenarios where sensor errors exhibit compound interactions. For example, when the influence of humidity varies across different $PM_{2.5}$ concentration ranges or when temporal effects interact with environmental conditions. Gradient Boosting models are typically more sensitive to fine-scale patterns in the data and can achieve superior predictive accuracy relative to RF, although they require careful tuning to avoid overfitting. In the context of low-cost $PM_{2.5}$ sensor calibration, GB demonstrated strong capability in capturing nonlinear biases and environmental dependencies, making it a powerful component of the calibration pipeline.

Support Vector Regression (SVR) offers a kernel-based alternative that models nonlinear relationships through a smooth functional mapping. SVR uses an $\varepsilon$-insensitive loss function to tolerate measurement noise while identifying a regression function with optimal complexity–accuracy tradeoff. This makes SVR suitable for capturing progressive nonlinear distortions in sensor response (Mai et al., 2025). SVR can achieve strong performance but requires careful tuning of kernel parameters (e.g., $C$, $\varepsilon$, and $\gamma$) and may be computationally expensive for large datasets.

https://doi.org/10.5194/egusphere-2025-6203
Preprint. Discussion started: 26 January 2026


To further enhance robustness, an Adaptive Calibration Model was developed by blending the physics-based linear correction and the nonlinear machine-learning prediction (Mahajan & Helbing, 2025). After training a base model (Gradient Boosting or LightGBM), we compute the blended output as:

$$PM_{blend} = \omega \cdot PM_{corr} + (1 - \omega) \cdot PM_{model} \quad (4)$$

here $PM_{model}$ is the machine-learning prediction and $\omega \in [0,1]$ is the blending coefficient. A grid search over $\omega$ is performed using a separate validation set to minimize the mean absolute error (MAE). This procedure prevents overreliance on either the linear correction (which may be too rigid) or the machine-learning prediction (which may be unstable in noisy or data-sparse regions). The adaptive approach consistently achieved the best overall performance, balancing accuracy, stability, and resilience against environmental variability.

In the third stage, model development followed a principled strategy that prioritized interpretability while aiming to meet the U.S. EPA performance criterion of $R^2 \geq 0.7$. The analysis progressed from simple univariate linear regression to multivariable regression and, where necessary, to nonlinear machine-learning models. After calibration, statistical analyses evaluated the agreement between the low-cost ZH03 sensor and the BAM reference across different environmental regimes. Performance metrics including the coefficient of determination ($R^2$), root mean square error (RMSE), mean absolute error (MAE), and mean bias error (MBE), quantified both accuracy and systematic deviation. Additional correlation analysis assessed the influence of temperature and relative humidity on sensor performance.

Results were further interpreted through visual diagnostic tools such as scatter plots, time-series overlays, Bland–Altman plots, and residual distributions, which helped elucidate temporal behaviour and environmental dependencies. All analyses were conducted using Python 3.11 and MATLAB R2023a, leveraging scientific libraries such as NumPy, Pandas, and Matplotlib. This integrated analytical framework captured both the temporal and environmental characteristics of the ZH03 sensor, providing a robust foundation for calibration model development and subsequent environmental compensation modeling.

## III. RESULT DISCUSSION

### 3.1 Data raw LCS ZH03 and Reference BAM

Table 3 presents the raw data distribution of $PM_{2.5}$ measurements recorded by the low-cost sensor (LCS ZH03) and the reference BAM 1022 across varying humidity and temperature ranges during the co-location campaign. The statistical summary includes the first and third quartiles (Q1 and Q3), minimum, median, and maximum concentrations, allowing an initial evaluation of the agreement and divergence between the two instruments before calibration.

https://doi.org/10.5194/egusphere-2025-6203
Preprint. Discussion started: 26 January 2026


Overall, the BAM reference measurements exhibit more constrained variability across environmental conditions, reflecting the stability of the beta-attenuation method. In contrast, the LCS ZH03 displays a broader dynamic range in the upper quantiles, particularly under elevated humidity levels. For instance, within the 85–95% RH range, the LCS recorded maximum $PM_{2.5}$ values up to 209.2 µg/m³, nearly double the BAM maximum of 122.2 µg/m³. This pattern is consistent with known limitations of optical particle counters, where hygroscopic particle growth and humidity-driven scattering lead to overestimation of particulate concentrations at high relative humidity.

At moderate humidity levels (55–75%), both instruments show more comparable medians—e.g., at 65–75% RH, the BAM median is 37.2 µg/m³, whereas the LCS median slightly overestimates at 47.9 µg/m³. This indicates that the LCS retains reasonable sensitivity and linearity under mid-range conditions but deviates more substantially as humidity approaches saturation. Temperature stratification exhibits a similar trend. At lower temperature ranges (22–25°C), both instruments demonstrate lower variability and closer agreement, with BAM and LCS medians of 28.4 µg/m³ and 39.9 µg/m³, respectively. However, at temperatures above 31°C, the LCS again shows greater spread, with maximum values nearly doubling the BAM readings. This reflects the temperature sensitivity of the ZH03's laser scattering chamber, which can influence signal amplitude and particle count processing.

Collectively, the distribution patterns in Table 3 confirm two essential characteristics. First, BAM produces stable and less environmentally sensitive measurements, serving as a consistent reference across all conditions. Second, LCS ZH03 is highly influenced by humidity and, to a lesser extent, temperature, resulting in higher measurement dispersion and systematic overestimation during unfavorable conditions. These findings reinforce the necessity of applying environmental compensation and calibration models to adjust LCS readings prior to interpretation. The raw data behavior observed here forms the empirical foundation for the regression-based correction functions developed in the subsequent stages of this research.

**Tabel 3. Data Raw distribution of LCS ZH03 and Reference BAM 1022**

| Environemnt variable | Value (HR in % and Temp in C) | PM2.5 BAM reference | | | | | PM2.5 LCS reference | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Q1 | Q3 | Min | Median | Max | Q1 | Q3 | Min | Median | Max |
| Humidity rate (%) | 35–45 | 28.6 | 34.3 | 27.8 | 33.7 | 39.8 | 31.2 | 41 | 24.2 | 33.2 | 56.2 |
| | 45–55 | 31.5 | 47.8 | 17.9 | 37.7 | 76.2 | 29 | 49.8 | 11.5 | 40 | 102 |
| | 55–65 | 29.4 | 51.9 | 10.1 | 40 | 94.6 | 27 | 59.5 | 4.8 | 42.7 | 137.8 |
| | 65–75 | 27.8 | 46.8 | 11.8 | 37.2 | **136.9** | 31.8 | 68.6 | 5.2 | 47.9 | **209.2** |
| | 75–85 | 29.9 | 61.5 | 10.7 | 44.3 | **122.4** | 42.9 | 100.5 | 0.8 | 68.8 | **250** |
| | **85–95** | 29.2 | 55.7 | 3.6 | 45.2 | **122.2** | 32.4 | 90.1 | 0.5 | 60.8 | **162.5** |
| Temperature (C) | 22–25 | 17.9 | 45.3 | 3.6 | 28.4 | 83 | 21.4 | 71.6 | 2.7 | 39.9 | 127.2 |
| | 25–28 | 33.1 | 62 | 8.2 | 46.4 | 136.9 | 45.5 | 100.5 | 0.5 | 71.2 | 250 |
| | 28–31 | 26.5 | 45.3 | 11.8 | 35.6 | 108.6 | 28.1 | 66.1 | 0.8 | 44.2 | 179.5 |
| | 31–34 | 31.8 | 52 | 10.1 | 40.3 | 99.6 | 30 | 58.8 | 4.8 | 44 | 137.8 |
| | 34–37 | 27.2 | 50.7 | 21.8 | 35.1 | 77.3 | 29 | 45.1 | 13.2 | 35.5 | 104.2 |

https://doi.org/10.5194/egusphere-2025-6203
Preprint. Discussion started: 26 January 2026


Figure 6 compares the raw $PM_{2.5}$ data from the low-cost sensor (LCS ZH03) and the reference BAM 1022 during the June–August 2025 co-location period. The daily boxplots in Figure 6(a) show that the ZH03 consistently reports higher $PM_{2.5}$ values than the BAM, with larger medians, wider spreads, and more outliers, particularly on humid or highly variable days. This pattern reflects the sensitivity of optical sensors to humidity-driven particle growth and enhanced light scattering. In contrast, the BAM exhibits more stable and constrained distributions. The hourly time-series in Figure 6(b) demonstrates that both instruments track similar temporal patterns, but the ZH03 displays amplified peaks, often reaching 150–250 µg/m³. Figure 6 indicates strong temporal agreement but systematic overestimation and higher variability in the ZH03 raw measurements.

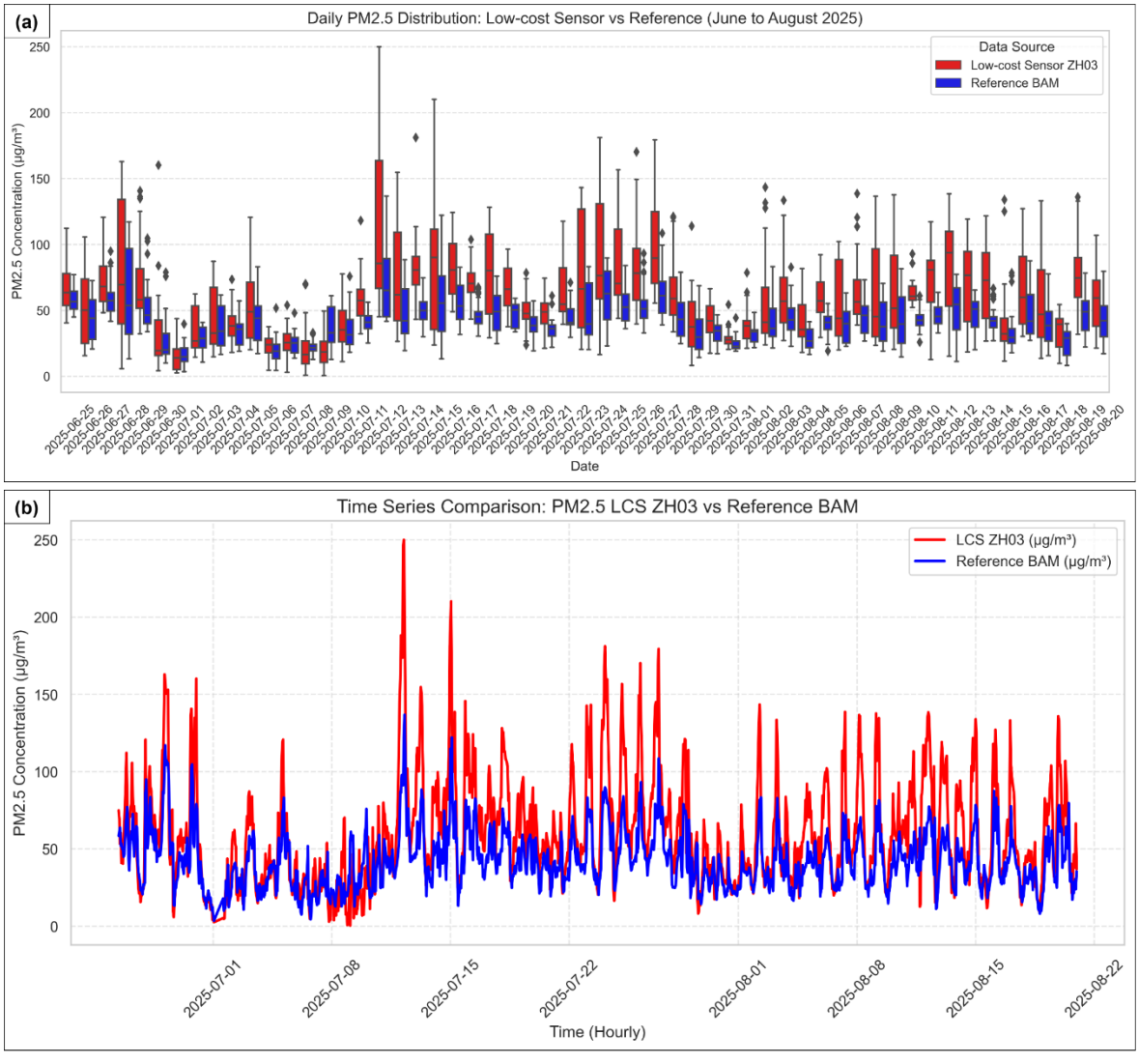


**Figure 6. Time series data raw comparison of PM2.5 between LCS and BAM**

https://doi.org/10.5194/egusphere-2025-6203
Preprint. Discussion started: 26 January 2026


Figure 7 presents detailed time-series comparisons of raw $PM_{2.5}$ measurements from the low-cost ZH03 sensor and the BAM 1022 reference instrument, alongside ambient temperature and relative humidity for selected days during the co-location campaign. These four representative days illustrate different atmospheric scenarios, such as high pollution events, extreme humidity conditions, and large humidity variability, to show how environmental factors influence the raw readings of the LCS.

Figure 7(a) illustrates the day with the highest LCS $PM_{2.5}$ peak, reaching values above 250 µg/m³. During the early morning hours (07–12 UTC+7), humidity exceeded 80%, coinciding with the rapid escalation in $PM_{2.5}$ reported by the ZH03. In contrast, the BAM measurement rose more gradually and peaked significantly lower, indicating that the LCS over-response was not driven solely by real aerosol loading but was also amplified by humidity-induced particle scattering. As humidity decreased after noon, the LCS signal dropped sharply, aligning more closely with the BAM values. This pattern highlights the strong dependency of optical sensors on rapidly changing microclimatic conditions.

Figure 7(b) shows another high-PM day where the ZH03 again exhibited elevated readings relative to the BAM, particularly during the early morning hours when humidity increased steadily above 85%. Although both devices captured the same peak time window, the LCS consistently overestimated by 40–60 µg/m³. The temperature remained relatively stable throughout the day, reinforcing the interpretation that humidity, rather than temperature, was the primary driver of the divergence.

In contrast, Figure 7(c) presents a day with the highest recorded humidity levels (>90%). Under these conditions, the mismatch between the two sensors becomes even more pronounced. The LCS signal fluctuated widely and produced several high peaks not observed in the BAM dataset. The optical chamber of the ZH03 is particularly susceptible to hygroscopic growth of fine particles and microdroplet formation, both of which enhance light scattering and produce artificial inflation of the $PM_{2.5}$ values. Despite temperature remaining stable around 29–31°C, the humidity-driven anomalies dominated the LCS behavior.

Lastly, Figure 7(d) represents the day with the largest humidity range (max–min = 43.5%). This strong variability caused rapid shifts in the LCS measurement profile, with $PM_{2.5}$ spikes occurring in parallel with humidity increases. The BAM signal remained stable and unaffected by these fluctuations. This subplot clearly illustrates the sensitivity of LCS optical detection to humidity dynamics rather than to actual changes in particulate loading.

Overall, Figure 7 demonstrates that while the LCS ZH03 can capture temporal patterns similar to the BAM, its raw $PM_{2.5}$ readings are strongly influenced by humidity magnitude and variability. These findings reinforce the critical need for environmental compensation and calibration models to correct the sensor's overresponse under humid conditions.

https://doi.org/10.5194/egusphere-2025-6203
Preprint. Discussion started: 26 January 2026


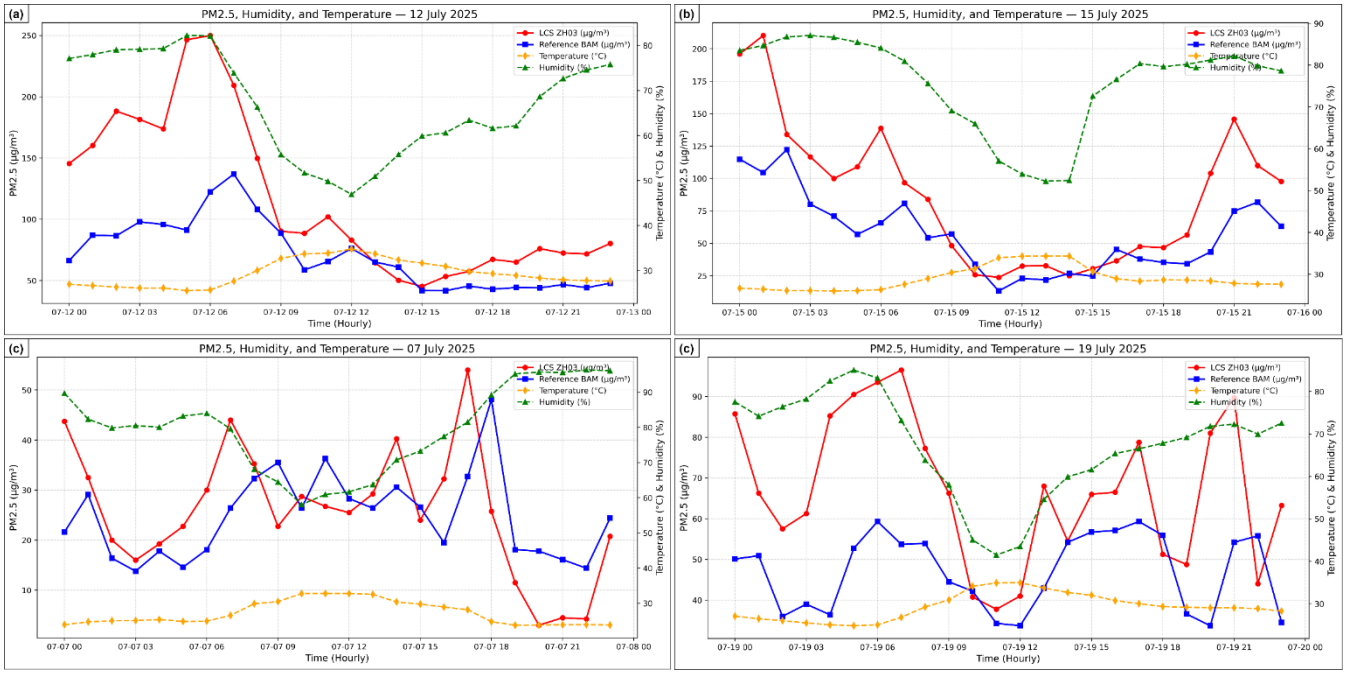


**Figure 7. Data raw of PM2.5 LCS, PM2.5 Reference, Ambient air temperature, and ambient air humidity rate (a) and (b) plot as highest PM2.5 LCS value and (c) plot as highest humidity rate above 90% along measurement period and (d) plot as highest humidity rate range (max-min=43.5)**

## 3.2 Raw data Precission between LCS and reference sensor

Figure 8 presents the precision analysis between the low-cost sensor (LCS ZH03) and the reference BAM 1022 based on hourly raw $PM_{2.5}$ measurements. The figure includes (a) a linear regression comparison and (b) a Bland–Altman analysis, providing a comprehensive assessment of agreement, proportional bias, and measurement variability prior to calibration.

The scatter plot in Figure 8(a) displays the relationship between LCS ZH03 and BAM $PM_{2.5}$ concentrations along with the best-fit regression line, $y = 0.44x + 16.57$. The slope less than 1 indicates that the LCS overestimates $PM_{2.5}$ at low-to-moderate concentrations but underestimates relative to the BAM at higher values. This nonlinear behavior reflects the combined influence of humidity-driven optical scattering and sensor saturation at elevated PM levels. The coefficient of determination $R^2 = 0.676$ demonstrates moderate correlation, showing that the LCS captures major temporal trends but with limited precision. Error metrics reinforce this: RMSE = 18.01 µg/m³, MAE = 8.01 µg/m³, and MBE ≈ 0 indicate substantial spread but minimal overall bias when averaged across the dataset. The relatively high dispersion around the regression line highlights the need for calibration to correct environmental sensitivity and improve predictive accuracy.

The Bland–Altman plot in Figure 8(b) evaluates the measurement differences (LCS – BAM) against their average. The mean bias is +17.24 µg/m³, confirming that the LCS systematically overestimates $PM_{2.5}$ relative to the BAM. The limits of agreement (LoA), ranging from −27.30 to +61.77 µg/m³, indicate wide variability and substantial disagreement at the

https://doi.org/10.5194/egusphere-2025-6203
Preprint. Discussion started: 26 January 2026


individual-measurement level. The upward trend in differences with increasing concentration suggests proportional bias, where overestimation becomes more pronounced under higher pollution levels or elevated humidity conditions. Most data points fall within the LoA range, but the spread increases at higher means, characteristic of optical sensor response amplification during periods of high particle scattering efficiency or hygroscopic aerosol growth. This aligns with earlier findings showing humidity as a major driver of LCS signal inflation.

Together, the regression and Bland–Altman analyses confirm that while the LCS ZH03 follows the general $PM_{2.5}$ trend measured by the BAM, it exhibits systematic overestimation, high variability, and proportional bias in its raw form. These results support the necessity of applying calibration models to correct the sensor's environmental sensitivity and improve reliability for quantitative air-quality monitoring.

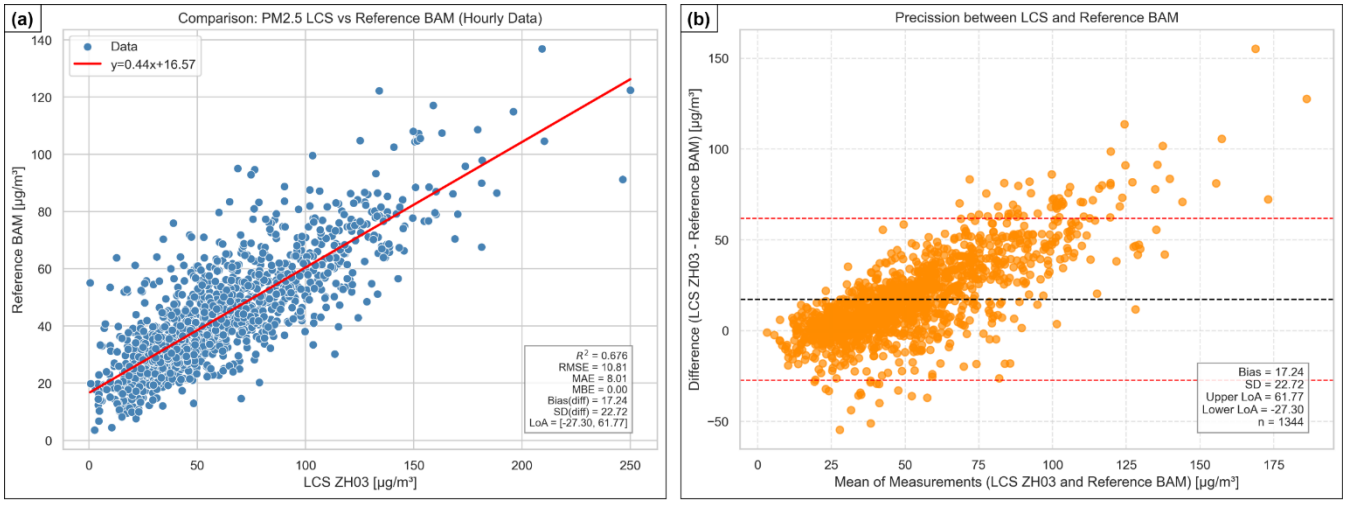


**Figure 8. Data raw PM2.5 LCS vs BAM Precision analysis with linear regression**

### 3.3 Particulate matter correlation to environment variable

Figure 9 presents a correlation analysis between $PM_{2.5}$ concentrations and key environmental variables (temperature and relative humidity) using both linear regression and Pearson's correlation coefficients. The results provide insight into how environmental conditions influence raw $PM_{2.5}$ measurements from the low-cost sensor (LCS ZH03) and the reference BAM 1022.

The scatter plot in Figure 9(a) shows a weak positive relationship between BAM $PM_{2.5}$ and relative humidity, with an extremely low coefficient of determination ($R^2 = 0.007$). This indicates that humidity has minimal impact on the reference measurements, highlighting the robustness of the BAM's beta attenuation method, which is designed to minimize moisture

https://doi.org/10.5194/egusphere-2025-6203
Preprint. Discussion started: 26 January 2026


interference. The fitted regression line is nearly flat, emphasizing that changes in humidity do not significantly affect BAM readings.

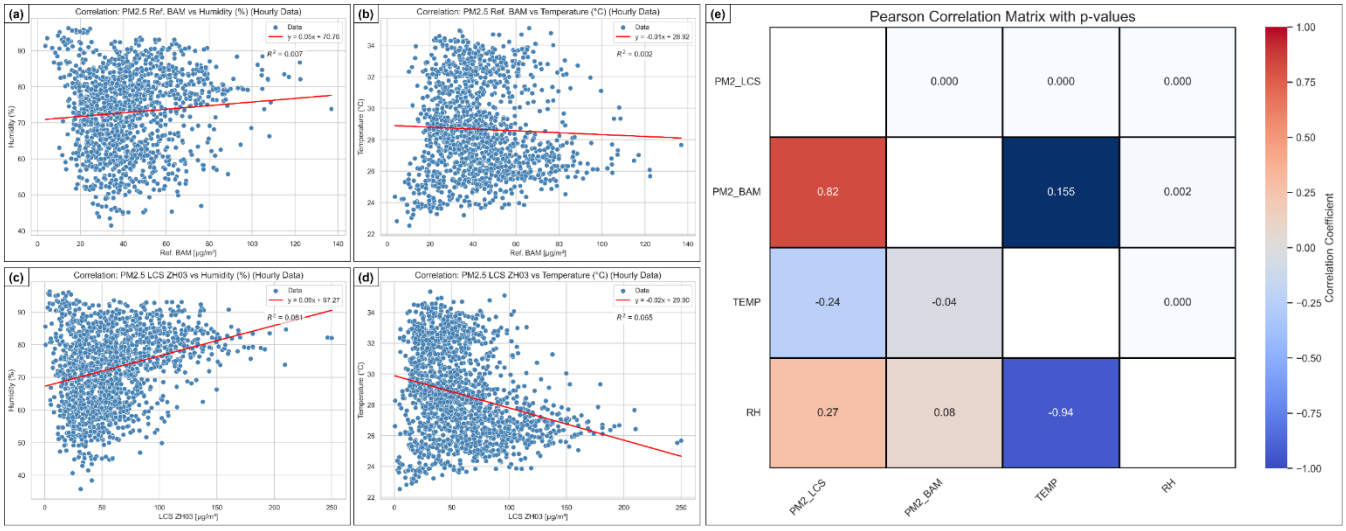


**Figure 9. Correlation analysis of PM2.5 with temperature and humidity use linear regression (Lin-Reg) and Pearson correlation (a) Lin-Reg PM2.5 BAM reference with humidity, (b) Lin-Reg PM2.5 BAM reference with temperature, (c) Lin-Reg PM2.5 LCS with humidity, (d) Lin-Reg PM2.5 LCS with temperature, and (e) Pearson correlation matrix PM2.5 LCS, PM2.5 BAM Reference, Humidity, and Temperature.**

Similarly, Figure 9(b) demonstrates a negligible relationship between BAM $PM_{2.5}$ and temperature ($R^2$ = 0.002). The slight negative slope suggests a weak inverse trend, but the correlation is statistically insignificant. These results confirm that the BAM maintains stable performance across typical outdoor temperature fluctuations during the study period.

Figure 9(c) reveals a stronger association between LCS $PM_{2.5}$ readings and humidity, with a higher (but still modest) coefficient ($R^2$ = 0.061). The positive slope indicates that $PM_{2.5}$ values reported by the ZH03 increase with rising humidity. This aligns with known limitations of optical sensors, where hygroscopic particle growth and increased light scattering contribute to an overestimation of $PM_{2.5}$ under humid conditions.

The relationship between LCS $PM_{2.5}$ and temperature shown in Figure 9(d) is also weak, with $R^2$ = 0.045, though slightly stronger than the BAM relationship. The negative slope suggests that higher temperatures may reduce the LCS $PM_{2.5}$ signal, potentially due to reduced relative humidity or thermal effects on the sensor's optical chamber. Nevertheless, temperature remains a secondary factor compared to humidity.

The Pearson correlation matrix in Figure 9(e) summarizes linear associations across all variables. $PM_{2.5}$ from the LCS strongly correlates with BAM $PM_{2.5}$ ($r = 0.82$, $p < 0.01$), reflecting good temporal agreement despite known biases. However, LCS $PM_{2.5}$ also exhibits moderate positive correlation with humidity ($r = 0.27$), while BAM $PM_{2.5}$ shows negligible correlation ($r = 0.08$). Temperature is weakly and negatively correlated with both $PM_{2.5}$ measurements, with the strongest association observed between temperature and humidity ($r = -0.94$), confirming their expected inverse

https://doi.org/10.5194/egusphere-2025-6203
Preprint. Discussion started: 26 January 2026


meteorological relationship. Overall, Figure 9 demonstrates that humidity is a major source of bias in LCS measurements, whereas BAM readings remain largely unaffected by environmental conditions. These correlations justify the need for humidity-aware calibration models to enhance LCS data accuracy.

### 3.4 Calibration models

Figure 10 presents the accuracy improvement of the low-cost $PM_{2.5}$ sensor (LCS ZH03) after applying various calibration models, using regression plots comparing model-adjusted predictions to the BAM 1022 reference measurements. Six calibration conditions are shown: uncalibrated raw data and five calibrated models (Linear Regression, Random Forest, Gradient Boosting, Support Vector Regression (SVR), and an Adaptive-blend ensemble). Each subplot illustrates predicted versus reference $PM_{2.5}$ on the test dataset along with key statistical indicators (MAE, RMSE, $R^2$).

The uncalibrated LCS shows poor agreement with the reference, characterized by wide scattering around the 1:1 line and substantial overestimation at higher concentrations. Performance metrics (MAE = 17.40, RMSE = 22.59, $R^2$ = −0.970) confirm low accuracy and negative predictive power, reflecting the strong influence of humidity and nonlinear bias present in the raw data. Applying a simple Linear Regression model improves performance notably relative to raw data. The scatter becomes more aligned, and prediction errors decrease (MAE = 8.32, RMSE = 10.77, $R^2$ = 0.552). However, the model remains limited by its inability to capture nonlinear interactions between $PM_{2.5}$, humidity, and temperature, resulting in moderate residual bias.

The Random Forest model provides a substantial improvement, capturing nonlinear relationships more effectively. Predictions align closely with the reference values, with significantly reduced error (MAE = 6.21, RMSE = 7.99) and strong explanatory power ($R^2$ = 0.754). The Random Forest handles humidity-driven variability far better than linear methods. Then, Gradient Boosting performs similarly to Random Forest but with slightly better accuracy (MAE = 5.93, RMSE = 7.76, $R^2$ = 0.768). The model produces a tighter clustering around the 1:1 line, suggesting superior capability in addressing complex, nonlinear error structures inherent in optical PM sensor data. Then, the SVR model also demonstrates strong predictive ability (MAE = 6.04, RMSE = 7.91, $R^2$ = 0.758). The regression plot shows compact dispersion along the ideal line, indicating that SVR effectively captures nonlinear patterns. Performance is comparable to Random Forest and slightly below Gradient Boosting. Lastly, The Adaptive-blend model, an ensemble combining strengths of multiple algorithms—achieves the highest performance among all models (MAE = 5.85, RMSE = 7.69, $R^2$ = 0.771). The regression plot reveals excellent alignment with the reference, with minimal scatter and low systematic error. The model successfully compensates for humidity effects and nonlinear LCS deviations.

In Figure 10 demonstrates that calibration substantially enhances the accuracy of low-cost $PM_{2.5}$ measurements. While simple linear correction provides moderate improvement, nonlinear machine-learning models (particularly Gradient Boosting, Random Forest, SVR, and the Adaptive-blend ensemble) significantly reduce measurement bias and variability.

https://doi.org/10.5194/egusphere-2025-6203
Preprint. Discussion started: 26 January 2026


The Adaptive-blend model performs best overall, offering a robust and reliable correction for LCS ZH03 under varying environmental conditions.

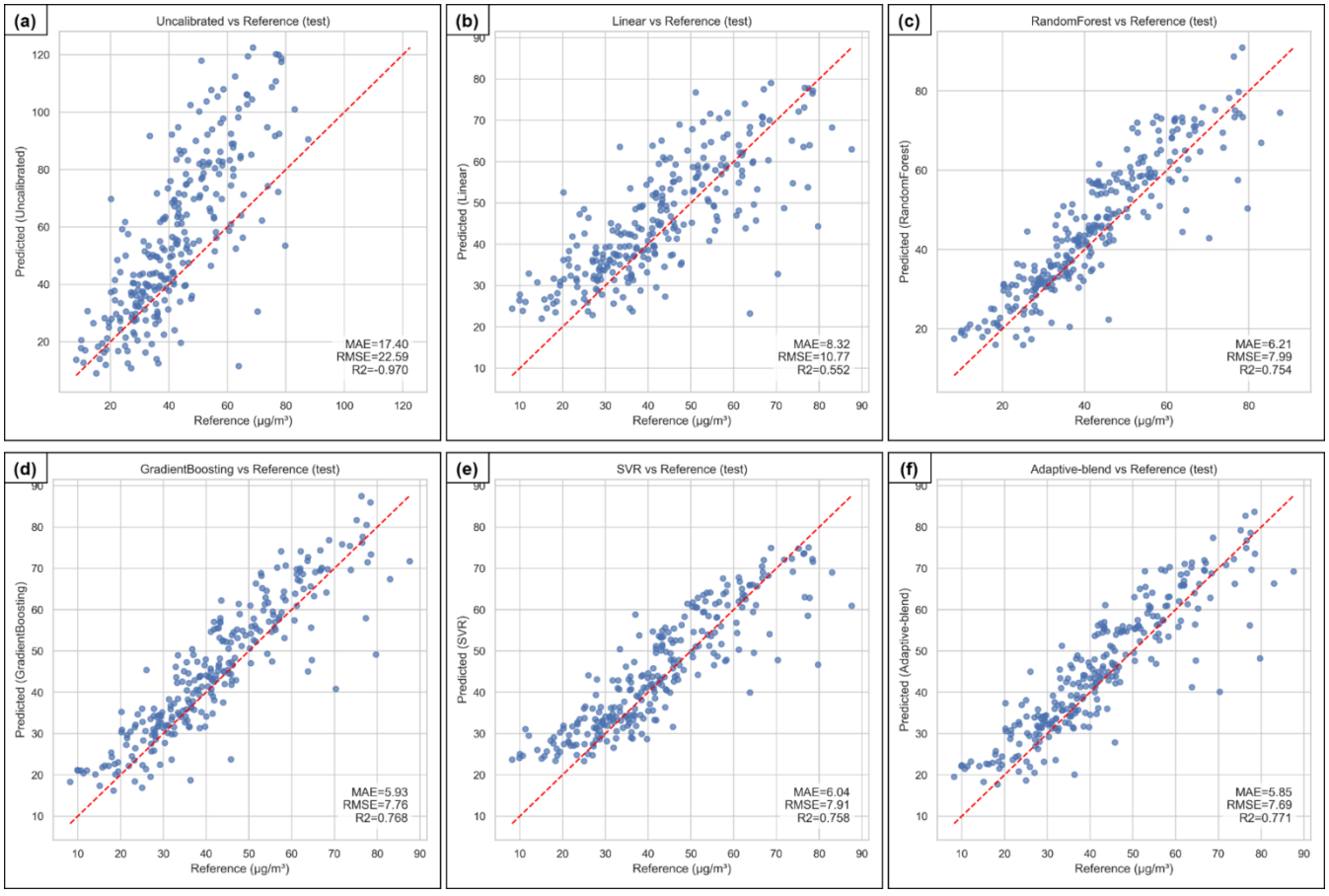


**Figure 10. Accuracy analysis of LCS after calibrated with various model vs BAM reference (a) regression plot of uncalibrated LCS, (b) regression plot of linear regression calibrated LCS, (c) regression plot of Random Forest calibrated LCS, (d) regression plot of Gradient Boosting calibrated LCS, (e) regression plot of Support Vector Regression calibrated LCD, and (f) regression plot of Adaptive-blend calibrated LCS.**

Figure 11 presents Bland–Altman plots comparing the agreement between the low-cost $PM_{2.5}$ sensor (LCS ZH03) and the BAM 1022 after applying various calibration models. Each subplot evaluates the mean difference (bias), spread of differences (standard deviation), and percentage of points outside the limits of agreement (LoA), providing insight into systematic error and precision improvement achieved by each model.

The uncalibrated sensor (Figure 11(a)) exhibits the largest positive bias (+13.52 µg/m³) and substantial scatter (SD = 18.13), with 6.71% of data points falling outside the LoA. The large spread and upward skew indicate significant overestimation and poor agreement with the reference, consistent with humidity-induced optical scattering effects observed earlier. Then, Applying Linear Regression (Figure 11(b) reduces the bias to +3.22 µg/m³ and narrows variability (SD =

https://doi.org/10.5194/egusphere-2025-6203
Preprint. Discussion started: 26 January 2026


10.30). The percentage of outliers decreases to 4.9%, reflecting moderate improvement. However, the residual positive bias indicates that linear correction is insufficient to fully compensate for the nonlinear environmental effects influencing the LCS.

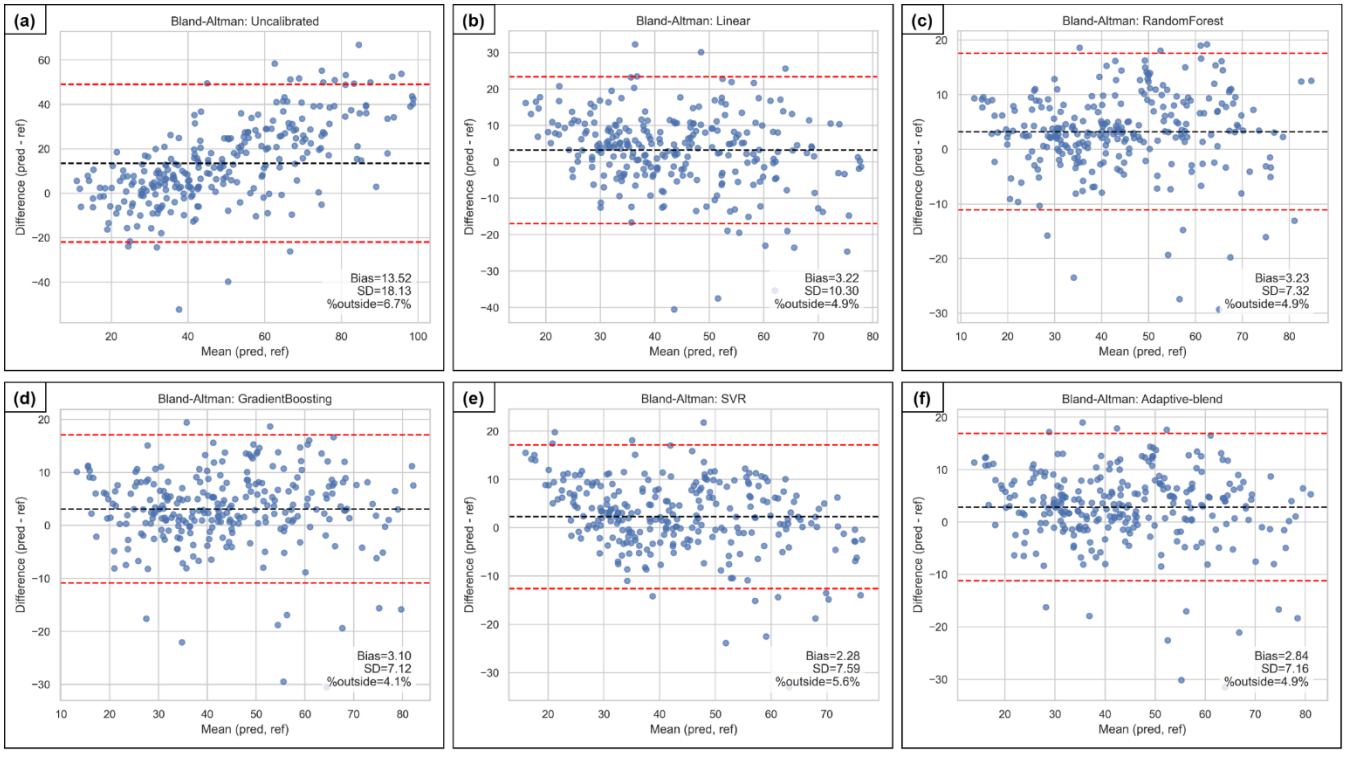


**Figure 11. Bland-Altman Plot for LCS after calibrated with various model vs BAM reference (a) regression plot of uncalibrated LCS, (b) regression plot of linear regression calibrated LCS, (c) regression plot of Random Forest calibrated LCS, (d) regression plot of Gradient Boosting calibrated LCS, (e) regression plot of Support Vector Regression calibrated LCD, and (f) regression plot of Adaptive-blend calibrated LCS**

The Random Forest (Figure 11(c)) model further reduces systematic error, yielding a bias of +3.23 µg/m³ and SD = 7.32, demonstrating tighter clustering around zero difference. The proportion of points outside the LoA drops to 4.9%, illustrating improved precision and effective handling of nonlinear relationships. Then, Gradient Boosting performs (Figure 11(d)) similarly to Random Forest but with slightly lower spread (SD = 7.12) and reduced bias (+3.10 µg/m³). Only 4.1% of data fall outside the LoA, indicating strong stability and reliable correction across the full $PM_{2.5}$ range. Then, The SVR model (Figure 11(e)) produces the lowest bias (+2.28 µg/m³) and relatively small variability (SD = 7.59). However, 5.6% of measurements lie outside the LoA, showing good but slightly less stable performance compared to ensemble-based models. SVR effectively corrects nonlinear deviations but shows modest sensitivity to extreme values. Lastly, The Adaptive-blend model (Figure 11(f)) achieves one of the best agreements, with bias +2.84 µg/m³ and SD = 7.16, and 4.9% outside the LoA.

https://doi.org/10.5194/egusphere-2025-6203
Preprint. Discussion started: 26 January 2026


The tight distribution and minimal systematic deviation indicate that combining multiple algorithms yields robust calibration performance, balancing bias reduction and variability control.

Figure 11 demonstrates that all calibration models significantly improve agreement between the LCS and BAM compared to the uncalibrated state. Ensemble-based methods (Gradient Boosting, Random Forest, and the Adaptive-blend) provide the most consistent performance, achieving low bias, reduced standard deviation, and fewer outliers. These models effectively correct the nonlinear, humidity-sensitive behavior of the LCS, supporting their suitability for field deployment in variable environmental conditions.

Figure 12 presents a comparative analysis of the Mean Absolute Error (MAE) achieved by each calibration method, highlighting the substantial improvement obtained through model-based correction relative to the uncalibrated low-cost sensor (LCS). The uncalibrated $PM_{2.5}$ measurements exhibit the highest MAE at 17.40 µg/m³, reflecting the significant raw bias and environmental sensitivity of the ZH03 sensor. Applying a simple Linear Regression model reduces the error to 8.32 µg/m³, demonstrating that even a basic correction strategy substantially enhances accuracy. However, more advanced machine-learning methods outperform the linear approach. Random Forest and Support Vector Regression yield MAE values of 6.21 µg/m³ and 6.04 µg/m³, respectively, while Gradient Boosting achieves a slightly lower error of 5.93 µg/m³. The Adaptive-blend model provides the best overall performance, achieving the lowest MAE of 5.85 µg/m³, indicating its superior ability to capture nonlinear patterns and environmental interactions affecting LCS measurements. In Figure 12 confirms that machine-learning–based calibration models significantly improve the reliability of low-cost $PM_{2.5}$ sensors, with ensemble and hybrid approaches yielding the most accurate predictions relative to the reference BAM instrument.

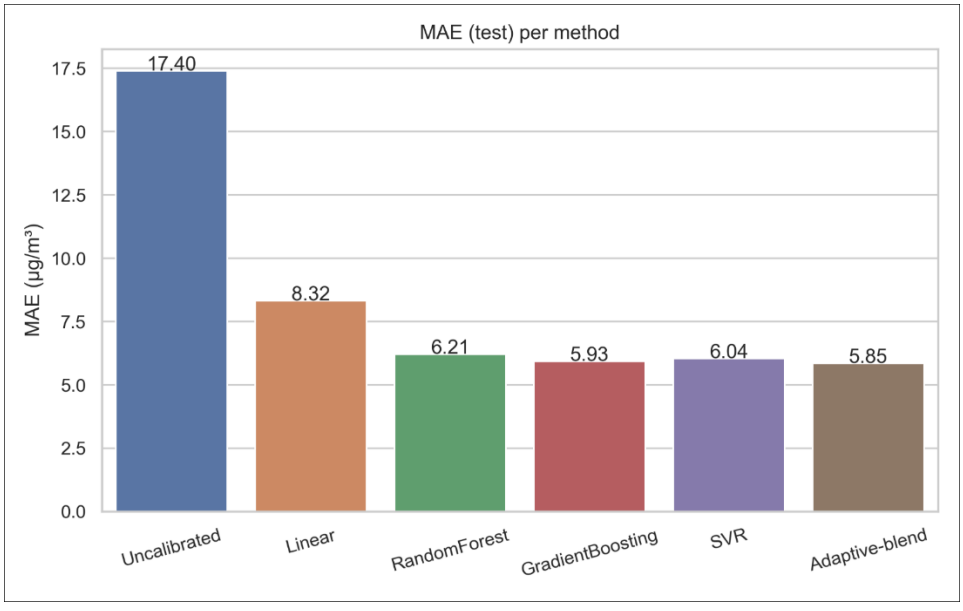


**Figure 12. Comparative analysis of calibration methods showing the MAE across different approaches.**

https://doi.org/10.5194/egusphere-2025-6203
Preprint. Discussion started: 26 January 2026


Figure 13 illustrates the time-series comparison of $PM_{2.5}$ measurements from the low-cost sensor (LCS ZH03) before and after calibration against the BAM 1022 reference instrument, demonstrating the substantial improvement achieved through model-based correction. The raw LCS signal (red dashed line) shows pronounced overestimation and unstable fluctuations, particularly during peak pollution periods, where readings exceed 140–160 µg/m³ despite the reference remaining closer to 80–100 µg/m³. After applying calibration models, the corrected outputs align much more closely with the reference time series, significantly reducing both amplitude bias and short-term variability. All calibrated models (Linear Regression, Random Forest, Gradient Boosting, Support Vector Regression, and the Adaptive-blend ensemble) successfully track the diurnal $PM_{2.5}$ pattern, capturing both rising and declining trends with improved fidelity. Among them, ensemble and nonlinear models show the closest agreement, producing smooth trajectories that closely follow the BAM measurements across the entire period. This convergence demonstrates that calibration effectively mitigates humidity-induced signal inflation and nonlinear sensor distortions, enabling the LCS to provide reliable and representative $PM_{2.5}$ estimates under real-world environmental conditions.

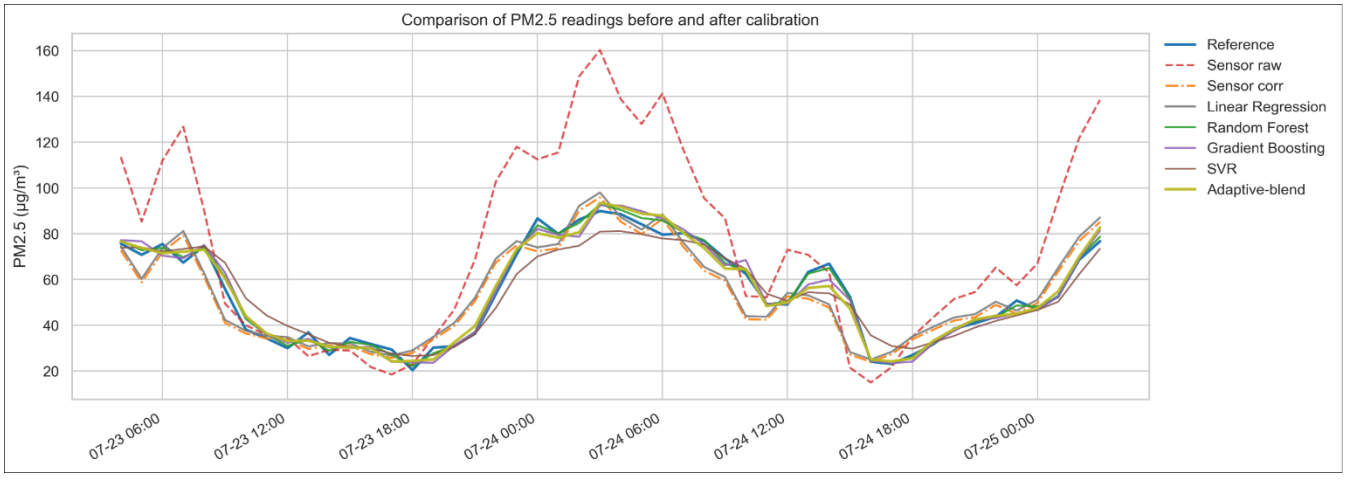


**Figure 13. Post calibration time series for LCS compared to the BAM reference**

## IV. CONCLUSION AND FUTUTRE WORKS

This study systematically evaluated the performance of a low-cost particulate matter sensor (Winsen ZH03) through an extensive co-location experiment with a reference-grade Beta Attenuation Monitor (BAM 1022) under real outdoor environmental conditions. The analysis of raw data revealed that although the ZH03 sensor successfully captured temporal $PM_{2.5}$ trends, it exhibited substantial overestimation and increased variability, particularly during periods of elevated humidity. Environmental correlation analyses confirmed humidity as the dominant driver of measurement bias, whereas temperature

https://doi.org/10.5194/egusphere-2025-6203
Preprint. Discussion started: 26 January 2026


exerted only a minor influence. These findings underscore the need for robust calibration strategies when deploying low-cost sensors in ambient air-quality monitoring applications.

A comprehensive calibration framework was developed, incorporating multiple regression approaches (including Linear Regression, Random Forest, Gradient Boosting, Support Vector Regression, and an Adaptive-blend ensemble) to correct raw LCS readings. Comparative performance assessment demonstrated that while linear correction reduced baseline bias, nonlinear machine-learning models provided markedly superior improvements. Ensemble-based models, particularly Gradient Boosting and the Adaptive-blend approach, consistently delivered the lowest prediction errors (MAE ≈ 5.8–5.9 μg/m³) and strongest agreement with the BAM reference, as evidenced by higher $R^2$ values and tight Bland–Altman limits of agreement. These models effectively mitigated humidity-induced signal inflation and nonlinear distortions inherent to optical scattering sensors.

Post-calibration time-series comparisons further validated the enhanced performance, showing that calibrated LCS outputs closely followed BAM measurements across varied pollution episodes and meteorological conditions. The improved alignment confirms that applying appropriate calibration models transforms low-cost sensors from trend-only indicators into quantitatively reliable instruments suitable for finer-scale air-quality monitoring.

This research demonstrates that low-cost $PM_{2.5}$ sensors (when properly corrected for environmental influences) can serve as effective complements to regulatory monitoring networks, enabling high-resolution spatial coverage at significantly lower cost. The proposed calibration methodology provides a practical pathway for improving data quality and supports broader adoption of low-cost sensing technologies in air-quality applications, particularly in regions with limited monitoring infrastructure. Future work may extend this framework by incorporating additional pollutants, real-time adaptive calibration, or physics-informed machine-learning models to further enhance robustness and scalability

**Code, data, or code and data availability**

The code and data supporting this study are available upon request.

**Author contributions**

GAP conceptualized the study, developed device, experiment, analysis, and write the article. PL Conceptualize the study, analysis of the result and discussion, acquired the funding, and review. EH supervised the work and review.

**Competing interests**

The contact author has declared that none of the authors has any competing interests.

https://doi.org/10.5194/egusphere-2025-6203
Preprint. Discussion started: 26 January 2026


**Acknowledgements**

This research was partially supported by the *Program Magister Doktor Sarjana Unggul* (PKPI-PMDSU) at Institut Teknologi Bandung. The authors also gratefully acknowledge additional support from the Clean Air Catalyst project led by the World Resources Institute (WRI).

**References**

Aix, M., Schmitz, S., & Bicout, D. J. (2023). Calibration methodology of low-cost sensors for high-quality monitoring of fi ne particulate matter. *Science of the Total Environment*, *889*(February). https://doi.org/10.1016/j.scitotenv.2023.164063

Bolan, S., Padhye, L. P., Jasemizad, T., Govarthanan, M., Karmegam, N., Wijesekara, H., Amarasiri, D., Hou, D., Zhou, P., Kumar, B., Balasubramanian, R., & Wang, H. (2024). Science of the Total Environment Impacts of climate change on the fate of contaminants through extreme weather events. *Science of the Total Environment*, *909*(August 2023). https://doi.org/10.1016/j.scitotenv.2023.168388

Chacón-mateos, M., García-salamero, H., Laquai, B., & Vogt, U. (2025). Calibration and performance evaluation of PM 2 . 5 and NO 2 air quality sensors for environmental epidemiology. *Atmospheric Measurement Techniques*, *2*, 4061–4085.

Cholianawati, N., Sinatra, T., Nugroho, G. A., Permadi, D. A., Indrawati, A., Halimurrahman, Kallista, M., Romadhon, M. S., Ma'ruf, I. F., Yudhatama, D., Madethen, T. A. P., & Awaludin, A. (2024). Diurnal and Daily Variations of PM2.5 and its Multiple-Wavelet Coherence with Meteorological Variables in Indonesia. *Aerosol and Air Quality Research*, *24*(3), 1–18. https://doi.org/10.4209/aaqr.230158

DeVore, G. R. (2017). Computing the Z Score and Centiles for Cross-sectional Analysis: A Practical Approach. *Journal of Ultrasound in Medicine*, *36*(3), 459–473. https://doi.org/https://doi.org/10.7863/ultra.16.03025

Dong, L., Han, Y., Hu, M., Zhang, Y., & Zhou, Q. (2025). Fast Atmospheric Aerosol Size and Shape Imaging Instrument : Design , Calibration , and Intelligent Interaction. *IEEE Transactions on Instrumentation and Measurement*, *74*, 1–17. https://doi.org/10.1109/TIM.2025.3551849

Edwards, T. D., Wong, Y. K., Jeong, C., Wang, J. M., Su, Y., & Evans, G. J. (2025). Comparison of methods for resolving the contributions of local emissions to measured concentrations. *Atmospheric Measurement Techniques*, 2201–2240.

Galiszewska, B., Telejko, M., & Starzomska, M. (2024). The effect of temperature and humidity of air on the concentration of. *Atmospheric Research*, *312*(May). https://doi.org/10.1016/j.atmosres.2024.107733

Hapidin, D. A., Saputra, C., Maulana, D. S., Munir, M. M., & Khairurrijal, K. (2019). Aerosol Chamber Characterization for Commercial Particulate Matter ( PM ) Sensor Evaluation. *Aerosol and Air Quality Research*, 181–194. https://doi.org/10.4209/aaqr.2017.12.0611

Hua, J., Zhang, Y., de Foy, B., Mei, X., Shang, J., Zhang, Y., Sulaymon, I. D., & Zhou, D. (2021). Improved PM2.5 concentration estimates from low-cost sensors using calibration models categorized by relative humidity. *Aerosol*

https://doi.org/10.5194/egusphere-2025-6203
Preprint. Discussion started: 26 January 2026


*Science and Technology*, *55*(5), 600–613. https://doi.org/10.1080/02786826.2021.1873911

Instruments, M. O. (2018). BAM 1022 Real-Time Portable Beta Attenuation Mass Monitor. *Metone.Com*.

Jayaratne, R., Liu, X., Thai, P., Dunbabin, M., & Morawska, L. (2018). The influence of humidity on the performance of a low-cost air particle mass sensor and the effect of atmospheric fog. *Atmospheric Measurement Techniques*, *11*(8), 4883–4890. https://doi.org/10.5194/amt-11-4883-2018

Kirešová, S., & Guzan, M. (2022). Determining the Correlation between Particulate Matter PM10 and Meteorological Factors. *Eng*, *3*(3), 343–363. https://doi.org/10.3390/eng3030025

Lanzi, E., Dellink, R., & Chateau, J. (2020). The sectoral and regional economic consequences of outdoor air pollution to 2060. *Energy Economics*, *71*(2018), 89–113. https://doi.org/10.1016/j.eneco.2018.01.014

Ma, X., Fan, Y., Wang, Y., Wang, X., Tan, Z., Li, D., Gao, J., Zhang, L., Xu, Y., Liu, X., Cai, S., Ma, Y., & Huang, Y. (2025). A Machine Learning-Based Calibration Framework for Low-Cost PM 2 . 5 Sensors Integrating Meteorological Predictors. *Chemosensors*, 1–21.

Mahajan, S., & Helbing, D. (2025). Dynamic calibration of low-cost PM 2 . 5 sensors using trust-based consensus mechanisms. *Climate and Atmospheric Science*.

Mai, C., Wang, Z., Chen, L., Huang, Y., Li, M., Shirazi, A., Altaee, A., & Zhou, J. L. (2025). Field-based calibration and operation of low-cost sensors for particulate matter by linear and nonlinear methods. *Atmospheric Pollution Research*, *16*(12), 102676. https://doi.org/10.1016/j.apr.2025.102676

Nothhelfer, M., Sperber, O., Todea, A. M., Schunke, B., Romazanowa, O., Schumacher, S., Bathen, D., & Asbach, C. (2023). Effect of an Aerosol Dryer on Ambient PM Measurements with SDS011 Low Cost Sensors during a Two-year Period in Duisburg, Germany. *Aerosol and Air Quality Research*, *23*(10). https://doi.org/10.4209/aaqr.230080

Orellano, P., Reynoso, J., Quaranta, N., Bardach, A., Ciapponi, A., Investigaciones, C. De, Nicolás, S., Tecnológica, U., Conicet, N., & Nicolás, S. (2020). Short-term exposure to particulate matter ( PM 10 and PM 2 . 5 ), nitrogen dioxide ( NO 2 ), and ozone ( O 3 ) and all-cause and cause-specific mortality : Systematic review and meta -analysis. *Environment International*, *142*(June), 105876. https://doi.org/10.1016/j.envint.2020.105876

Pouri, N., Karimi, B., Kolivand, A., & Hamed, S. (2024). Science of the Total Environment Ambient dust pollution with all-cause , cardiovascular and respiratory mortality : A systematic review and meta-analysis. *Science of the Total Environment*, *912*(August 2023), 168945. https://doi.org/10.1016/j.scitotenv.2023.168945

Prayoga, G. A., Husni, E., & Darmakusuma, R. (2025). A Wireless Sensor Network LCS-AQMS: Design, Instrumentation, Calibration Modelling, and Data Analysis. *IEEE Access*, *13*(August), 144362–144384. https://doi.org/10.1109/ACCESS.2025.3598855

Schneider, A., Borsdorff, T., Brugh, J. De, Aemisegger, F., Feist, D. G., Kivi, R., Hase, F., Schneider, M., & Landgraf, J. (2020). First data set of H 2 O / HDO columns from the Tropospheric Monitoring Instrument ( TROPOMI ). *Atmospheric Measurement Techniques*, 85–100.

Shukla, K., & Aggarwal, S. G. (2022). A Technical Overview on Beta-Attenuation Method for the Monitoring of Particulate

https://doi.org/10.5194/egusphere-2025-6203
Preprint. Discussion started: 26 January 2026


Matter in Ambient Air. *Aerosol and Air Quality Research*, *22*(12), 1–21.

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

Venkata, N., Rani, S., Abimbola, I., Ahmed, T., Anton, I., Ibrahim, Q., Banerjee, A., Tiwari, A., Gharbia, S., & Riaz, K. (2024). Challenges and Opportunities in Calibrating Low-Cost Environmental Sensors. *Sensors (Switzerland)*.

Wang, J., & Ogawa, S. (2015). Effects of meteorological conditions on PM2.5 concentrations in Nagasaki, Japan. *International Journal of Environmental Research and Public Health*, *12*(8), 9089–9101. https://doi.org/10.3390/ijerph120809089

Wang, P., Xu, F., Gui, H., Wang, H., & Chen, D. R. (2021). Effect of relative humidity on the performance of five cost-effective PM sensors. *Aerosol Science and Technology*, *55*(8), 957–974. https://doi.org/10.1080/02786826.2021.1910136

Wei, P., Ning, Z., Ye, S., Sun, L., Yang, F., Wong, K. C., Westerdahl, D., & Louie, P. K. K. (2018). Impact analysis of temperature and humidity conditions on electrochemical sensor response in ambient air quality monitoring. *Sensors (Switzerland)*, *18*(2). https://doi.org/10.3390/s18020059

Winsen. (2016). Laser Dust Module Manual ZH03/ZH03A/ZH03B. *Zhengzhou Winsen Electronics Technology Co. Ltd.*, 1–8.

Zender-Świercz, E., Galiszewska, B., Telejko, M., & Starzomska, M. (2024). The effect of temperature and humidity of air on the concentration of particulate matter - PM2.5 and PM10. *Atmospheric Research*, *312*(October). https://doi.org/10.1016/j.atmosres.2024.107733