ECOLOGIA BALKANICA

2025, Vol. 17, Issue 2 December 2025 pp. 166-174

Performance evaluation and cross-validation of low-cost particulate matter sensors for environmental research

Atanas Terziyski, Nikolay Kochev*, Stoyan Tenev, Svetlanna Georgieva

University of Plovdiv "Paisii Hilendarski", Faculty of Chemistry, Department of Analytical Chemistry and Computer Chemistry, 24 Tsar Assen Str., 4000, Plovdiv, BULGARIA

*Corresponding author: nick@uni-plovdiv.net

Abstract. Particulate monitoring data plays a vital role in supporting analysis, policymaking, and citizen initiatives, especially in areas related to ecology, air quality, public health, and overall quality of life. However, traditional regulatory monitoring systems are expensive and have drawbacks: they do not provide real-time data and cover only a limited number of official locations. As an alternative, low-cost monitoring devices are increasingly being used, but concerns remain regarding their accuracy and the reliability of the data they produce. As part of efforts to validate low-cost monitoring approaches, this work presents the design, implementation, and ongoing development of a low-cost, sensor-based air quality monitoring system dedicated to monitoring fine particulate matter (PM) and other atmospheric indicators within the METER.AC network. The system integrates multiple devices with various sensors, such as Honeywell HPMA115S0, Sensirion SEN55, and related SEN series, along with GPS modules for precise timestamp and geolocation. Data acquisition is synchronized and automated via UNIX shell scripts, which extract, convert, and process measurement data into structured CSV files containing parameters like PM1, PM2.5, PM10, temperature, humidity, etc. Fluke 985 Particle Counter is a high-quality professional device used as an independent benchmark for cross-validation, providing particle counts for sizes: 0.3, 0.5, 1.0, 2.0, 5.0, and 10.0 µm. Measurement data are collected at 10-minute intervals and uploaded to a public visualization platform, where interactive graphs and summaries are generated. Also, this study aims to approximate the mass of airborne particulate matter based on particle size distribution derived from the FLUKE device output. On the basis of particle counts, the underlying particle size distribution is estimated with regression models.

Key words: fine particles, monitoring network, meter.ac, optical particle counter, MLRA.

Introduction

Measuring particulate matter (PM) concentrations is crucial for air quality monitoring, environmental research, and public health. There are several alternative methods, each with its own principles, advantages, and limitations, summarized in Table 1. In gravimetric (filter-based) methods, particles are collected on a filter over a known air volume, and the filter is weighed before and after sampling to determine the mass of PM. This approach is the most accurate and widely used for regulatory purposes, research, or long-term monitoring. Also, it allows chemical analysis of the collected particles but does not provide real-time data. Additionally, it is labor-intensive with low temporal resolution (e.g., 24-hour averages) and requires precise weighing in controlled humidity and temperature. Some gravimetric methods, such as the Tapered Element Oscillating Microbalance (TEOM), provide near real-time measurements but still depend on filter replacement and operational care (Lee et al., 2005). Optical Particle Counters (OPCs) and photometers measure PM by detecting how particles scatter, absorb, or extinguish light from a source such as a laser diode (Feng et al., 2025). These real-time instruments provide size-specific particle counts and approximate mass concentrations after calibration against gravimetric reference instruments. OPCs are commonly used for portable and stationary air quality monitoring and allow continuous data collection. Beta-Attenuation Monitors (BAM) use beta radiation (Shukla et al., 2022). The particles collected on a filter attenuate the radiation, which correlates with the particle mass. This approach allows continuous, automated, and reliable data collection with high temporal resolution; however, the equipment is expensive and requires safety precautions due to the radioactive source.

Table 1. Comparison of common methods for particulate matter (PM) measurement and applications.

Method	Principle	Data Type	Accuracy	Application
Gravimetric	Filter sampling and weighing	Mass concentration	Highest	Regulatory, health research
Tapered Element Oscillating Microbalance (TEOM)	Real-time mass via oscillating filter	Continuous mass	High	Near real-time regulatory
Optical Particle Counter (OPC)	Light scattering by particles	Particle count & size	Moderate to high with calibration	Real-time monitoring, research
Beta Attenuation Monitor (BAM)	Beta radiation attenuation	Continuous mass	High	Regulatory compliance
Low-cost Optical Sensors	Light scattering in compact sensors	Real-time concentration	Moderate, variable	Community, personal, mobile monitoring

Low-cost PM sensors have become widely used for community and personal air quality monitoring due to their affordability and ease of deployment. Employing optical detection principles, these sensors can be stationary, mobile (vehicle- or drone-mounted), or portable (handheld). While generally less accurate than reference-grade instruments, they provide valuable high-resolution spatial and temporal data, supporting both research and public awareness, which is exactly the case of METER.AC network (Terziyski et al., 2020). Their performance depends on proper calibration and environmental factors, including temperature and humidity.

We present the design and implementation of a low-cost, sensor-based air quality monitoring system for fine particles, together with other atmospheric parameters within the METER.AC network. We describe the system integration of various sensors together with Fluke 985 OPC, as well as the methodologies developed for data management and analysis. This study also seeks to estimate the mass of airborne particulate matter by analyzing the particle size distribution obtained from the FLUKE device data. Based on OPC channels, regression models are applied to infer the underlying particle size distribution as well as to establish direct links to the mass concentration.

Materials and methods

The METER.AC initiative utilizes open-source tools, remote data acquisition, and CLI-based automation for scalable monitoring in diverse environments by reducing reliance on high-cost equipment. The hardware and software infrastructure of METER.AC was used as a basis for the current study. Additionally, we have set our own experimental setup specially tailored for the purposes of PM measurement devices comparison.

The Fluke 985 Particle Counter is a high-precision professional instrument used as an independent benchmark for cross-validation, providing particle counts across size ranges of 0.3, 0.5, 1.0, 2.0, 5.0, and 10.0 µm. The Honeywell HPMA 115S0 is a compact, laser-based sensor that measures PM1.0, PM2.5, PM4.0, and PM10 using UART output. It is reliable, fast, and suitable for basic indoor air quality monitoring. Also, in this study, we used the Sensirion SEN55, which is a more advanced module that measures PM, VOCs, NO_x, temperature, and humidity. It offers both UART and I²C interfaces and comes pre-calibrated, making it great for detailed air quality studies.

For the purposes of comparative study, a combined device for monitoring particulate matter (PM) was developed, consisting of three types of sensors: a reference-grade Fluke 985 Particle Counter and two low-cost sensors, SEN55 and HPMA 115S0. A GPS module was also integrated to provide georeferencing of the data stream. The block diagram of the device is shown in Fig. 1. It includes three sensors (on the left), a data management module based on a Raspberry Pi (in the center), and a power supply unit (on the right).

Fig. 2 shows the combined instrument in operation at one of the designated measurements locations. Measurements are recorded at 10-minute intervals and automatically uploaded to a public visualization platform that generates interactive graphs and summary statistics.

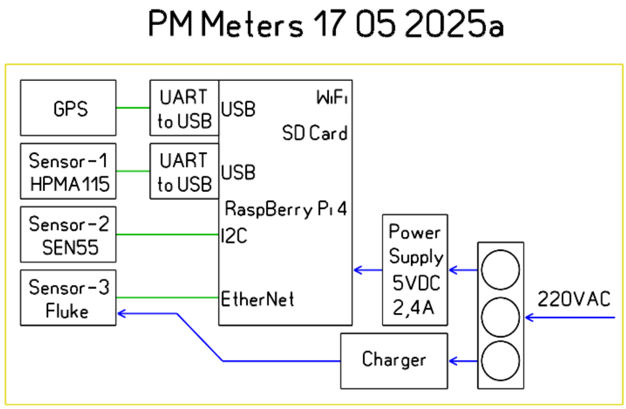

Fig. 1. Block diagram of the combined monitoring device: three sensors (on the left), a data management module based on a Raspberry Pi (in the center), and a power supply unit (on the right).

Fig. 2. View of the assembled monitoring device deployed at the rooftop of the Rectorate building, Plovdiv University (PU).

Once collected and transmitted to the central METER.AC database, the data can be accessed directly through the METER.AC GUI interface (Fig. 3) or downloaded as raw data files for further analysis.

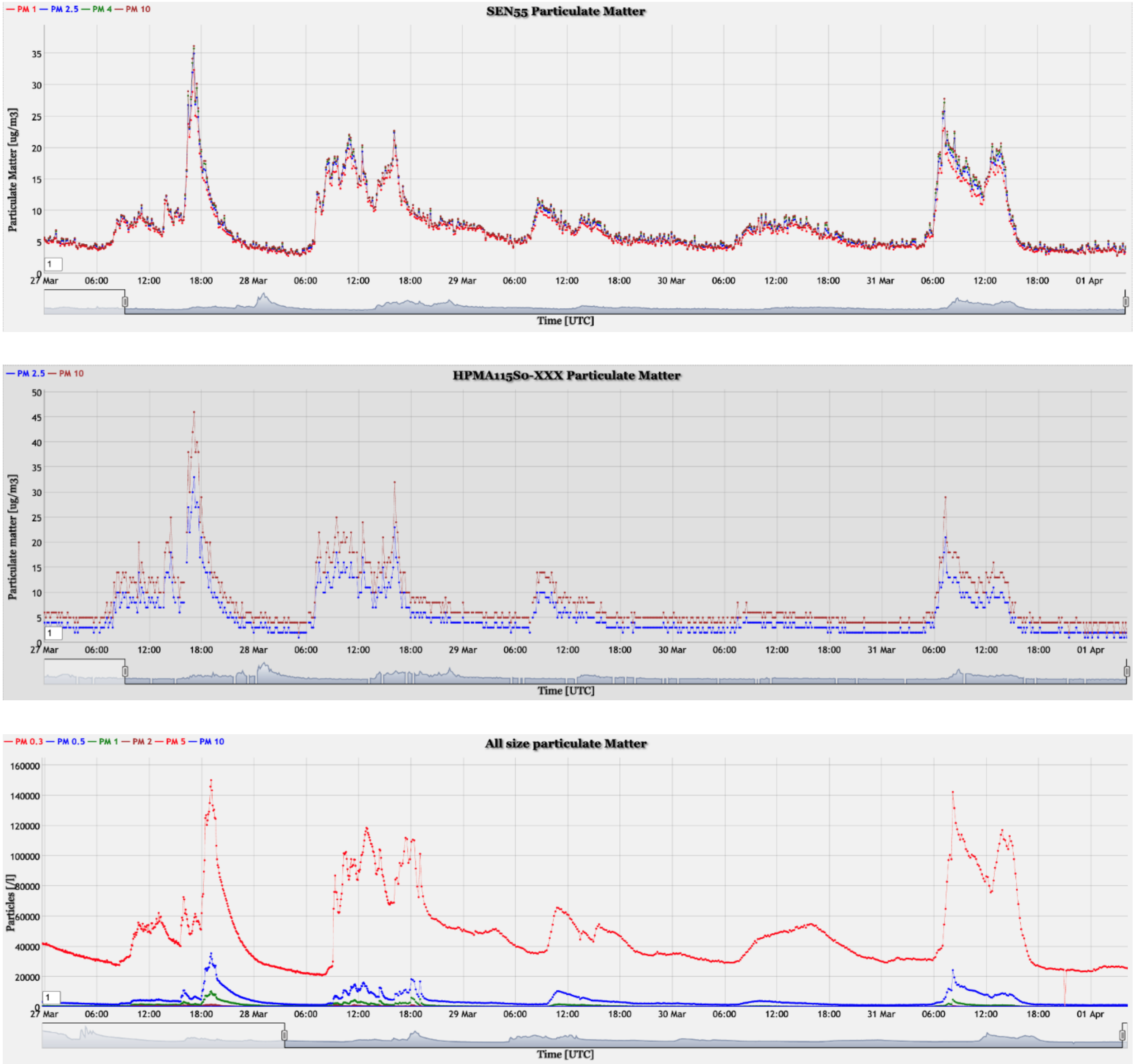

Fig. 3. Data visualization in the METER.AC web interface showing SEN55 four-channel PM mass concentrations (top), HPMA115S0 two-channel PM mass concentrations (middle), and Fluke 985 particle counts for sizes of 0.3, 0.5, 1.0, 2.0, 5.0, and 10.0 µm.

Furthermore, custom Jupyter Notebooks were developed for data extraction, processing, and analysis. The raw measurement data are first downloaded from the METER.AC platform in CSV format. Subsequently, data from the three different devices are synchronized and aggregated using their UTC timestamps to ensure temporal alignment. Regression and correlation analysis were performed to establish linear relationship models between various combinations of monitored parameters: OPC for sizes 0.3, 0.5, 1.0, 2.0, 5.0, 10.0 µm, and PM1, PM2.5, PM4, PM10 (in µg/m^3). The PM2.5 and PM10 values were measured by both SEN55 and HPMA115S0 sensors, with the latter denoted as PM2.5(2) and PM10(2), respectively.

In the second stage of our study, each FLUKE data point (a six-channel OPC vector) underwent particle size distribution analysis through non-linear transformation, followed by an estimation with a regression model. The Jupyter Notebook prepares the data for analysis and performs linear regressions on the transformed data across different power values (p) to identify the best-fitting model. The results of these regressions are saved to an Excel file, and a plot is generated for a specific p-value to visualize the relationship between particle size and the transformed particle counts.

During the reporting period, we conducted two-week measurement campaigns at multiple locations, which are summarized in the table below.

All raw sensor measurements and computed results are available in the supplementary materials hosted on Zenodo: 10.5281/zenodo.17488826.

Table 2. Locations selected for experimental monitoring.

Location	Start Date	End Date
Sarnegor village (outdoor)	15 Apr 2025	30 Apr 2025
PU, roof of the main building (outdoor)	3 May 2025	16 May 2025
PU, ACCC (Dep. of Analytical and Computer Chemistry) foyer	16 May 2025	29 May 2025
PU, OC (Organic Chemistry lab)	30 May 2025	10 June 2025

Results

Direct comparison between the two PM mass-reporting sensors, HPMA115S0 and SEN55, is straightforward, as both devices report mass concentration in the same units. Small discrepancies are observed, resulting in a correlation coefficient of 0.91 and a mean absolute deviation (MAD) of 2.34 μg/m^3 when comparing PM10 data from SEN55 with PM10(2) data from HPMA115S0, measured in the Organic Chemistry Lab at Plovdiv University (PU). This is expected, as the observed MAD falls in the manufacturer's reported uncertainty, which tends to show higher relative deviations at low PM concentrations - a scenario that occurs most frequently, fortunately, from both ecological and health perspectives.

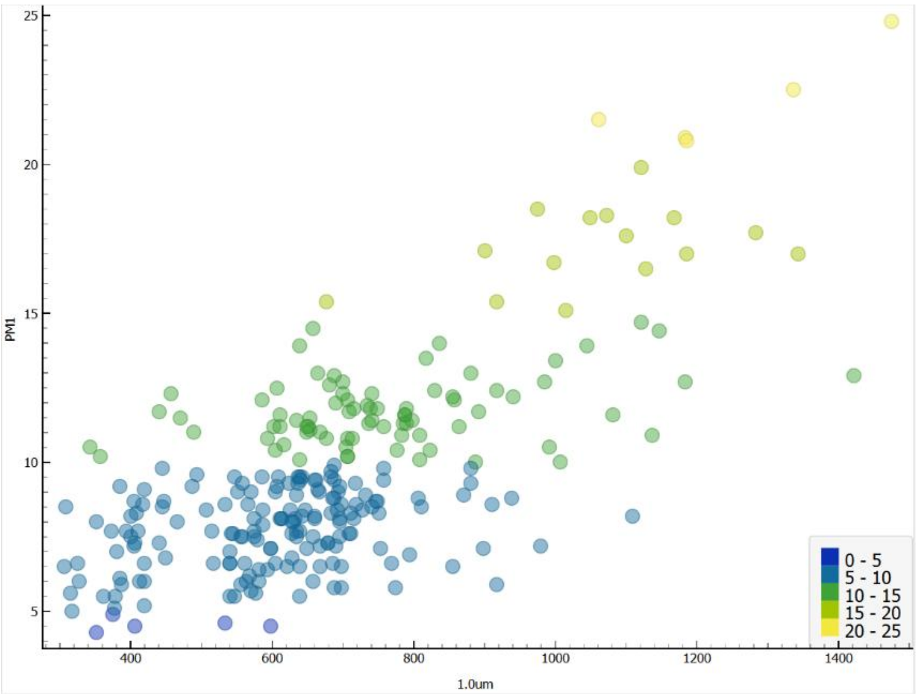

Fig. 4. Scatter plot of MP$_{1.0}$ mass concentration (μg/m^3) measured by the low-cost SEN55 sensor (Y-axis) versus the number of 1.0 μm particles measured by the OPC (X-axis).

A greater challenge lies in establishing a reliable correlation between the FLUKE OPC readings and the mass concentrations measured from SEN55 and HPMA115S0 sensors. Although some concordance in the trends is observed (e.g., comparing the top and middle plots in Fig. 3 with the bottom one), there is little to no significant linear relationship between the measured mass concentrations and the counts from individual FLUKE channels, as illustrated in Fig. 4. For example, the Pearson Moment Correlation Coefficient (PMCC, R^2) is 0.20 for MP1 versus OPC1, and even lower for MP10 versus OPC10 (R^2 = 0.0062). The complete list of all PMCC correlations is presented in the bottom third of Table 3, generally highlighting the absence of direct linear relationships. This suggests that the relationship between the variables is more complex.

Table 3. MLRA statistics (top), PM_x–OPC_y correlations (bottom), and average environmental parameters (temperature, relative humidity, and VOC index; middle).

		Sarnegor		PU, OC		PU, ACCC		PU, roof	
		R^2/Av.	*RMSE [μg/m³]*	*R^2/Av.*	*RMSE [μg/m³]*	*R^2/Av.*	*RMSE [μg/m³]*	*R^2/Av.*	*RMSE [μg/m³]*
MLRA	***PM1, $OPC_{0.3-10}$***	0.544	6.714	0.771	4.213	0.996	0.348	0.977	1.725
MLRA	***PM2.5, $OPC_{0.3-10}$***	0.507	7.715	0.767	4.579	0.996	0.403	0.977	1.872
MLRA	***PM4, $OPC_{0.3-10}$***	0.478	8.313	0.764	4.723	0.994	0.473	0.975	1.962
MLRA	***PM10, $OPC_{0.3-10}$***	0.464	8.598	0.763	4.793	0.993	0.512	0.974	2.014
MLRA	***PM2.5(2), $OPC_{0.3-10}$***	0.545	4.637	0.755	2.617	0.944	0.885	0.918	2.144
MLRA	***PM10(2), $OPC_{0.3-10}$***	0.587	6.415	0.778	3.669	0.945	1.373	0.921	3.061
Average	***T** [°C]*	16.85		25.44		25.16		19.80	
Average	***RH** [%]*	54.25		41.60		39.11		56.95	
Average	***VOC-Index***	33.52		40.18		32.35		35.19	
PMCC	***PM1, OPC_1***	0.20		0.48		0.67		0.46	
PMCC	***PM2.5, OPC_2***	0.13		0.31		0.40		0.26	
PMCC	***PM2.5 (2), OPC_2***	0.17		0.44		0.67		0.62	
PMCC	***PM10, OPC_{10}***	0.00		0.08		0.15		-0.17	
PMCC	***PM10 (2), OPC_{10}***	0.02		0.11		0.37		0.021	

We explored various approaches for Multivariate Linear Regression Analysis (MLRA) to establish relationships between all measured FLUKE particle count channels and the mass concentrations reported by low-cost sensors. The resulting MLRA models for PM_x are expressed as follows:

$$\boldsymbol{PM_x = a^x_{0.3} \times N_{0.3} + a^x_{0.5} \times N_{0.5} + a^x_1 \times N_1 + a^x_2 \times N_2 + a^x_5 \times N_5 + a^x_{10} \times N_{10}} \quad (1)$$

where N_k represents the OPC_k FLUKE channel (measured particle counts for size k), and $\boldsymbol{a^x_k}$ are the corresponding model coefficients.

For example, the MLRA equation for PM_1 at the PU, Organic Chemistry Lab location, is:

$$\boldsymbol{PM_1 = -0.037 \times N_{10} + 0.054 \times N_5 - 0.012 \times N_2 + 0.002 \times N_1 + 6.75 \times 10^{-5} \times N_{0.5} + 0.0001 \times N_{0.3}}$$

RMSE = 4.21 μg/m³ , R^2 = 0.77

The top section of Table 4 shows that the performance of the MLRA models varies across locations. The model for Sarnegor data demonstrates a relatively weak relationship between measured mass concentrations and OPC counts (e.g., $R^2 \approx 0.50$ and RMSE of 6 - 8 μg/m³). In contrast, the PU Organic Chemistry lab results are satisfactory ($R^2 \approx 0.77$, RMSE ≈ 5 μg/m³). The PU Analytical and Computer Chemistry Department (ACCC) location exhibits excellent model performance, with R^2 values ranging from 0.95 to 0.99 and RMSE < 1.5 μg/m³. Similar, though slightly lower, performance is observed for the PU roof location.

Size distribution estimation

Based on the six-channel FLUKE OPC data, we conducted a comprehensive study of particle size distributions. Initially, we explored a linear relationship between $\ln(\ln(N_k))$ and particle size

($size_i$), which yielded a coefficient of determination $R^2 = 0.93$.

Our primary approach for modeling the particle size distribution density (probability density) involves relating $size_i$ value to an appropriate power (p) of N_i (the number of particles of size $size_i$), where p is negative and typically around -0.5.

$$N_i^p = a \times size_i + b \quad (2)$$

Fig. 5 shows the linear regression line obtained from a single FLUKE measurement (a six-point vector) for p = -0.5.

From equation (2), by expressing the quantity proportional to Ni, the particle size distribution function is obtained in the following hyperbolic form (see also Fig. 5, right):

$$Q_i = (a \times size_i + b)^{1/p} \quad (3)$$

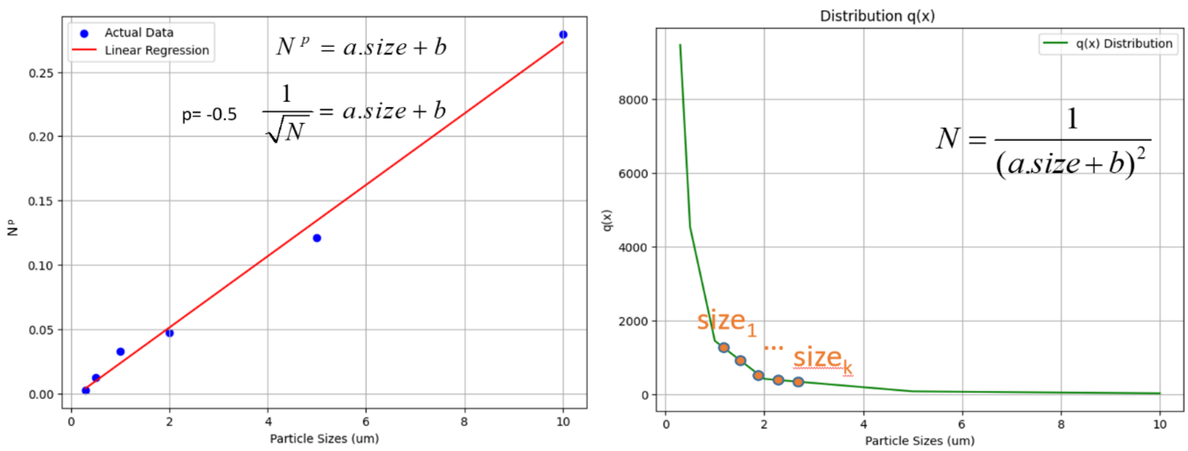

Fig. 5. Linear regression illustrating the relationship between size and $N^{-1/2}$ (left) and resulting particle size density distribution function (right).

We performed an optimization of the parameter p using a grid search. The p-values were varied from -0.05 to -0.95 in steps of 0.05. For each p and each measurement, a regression model, based on equation (2), was fitted. This resulted in approximately 1500 regressions per location and each specific p-value. For every location and p, the average R^2 and RMSE were calculated. Fig. 6 presents the results. Excellent regression performance is observed for p in the interval [-0.5, -0.2], with the optimum at approximately p = -0.25. This corresponds to a linear relationship between particle size and the reciprocal value of the fourth root of the particle count N_i.

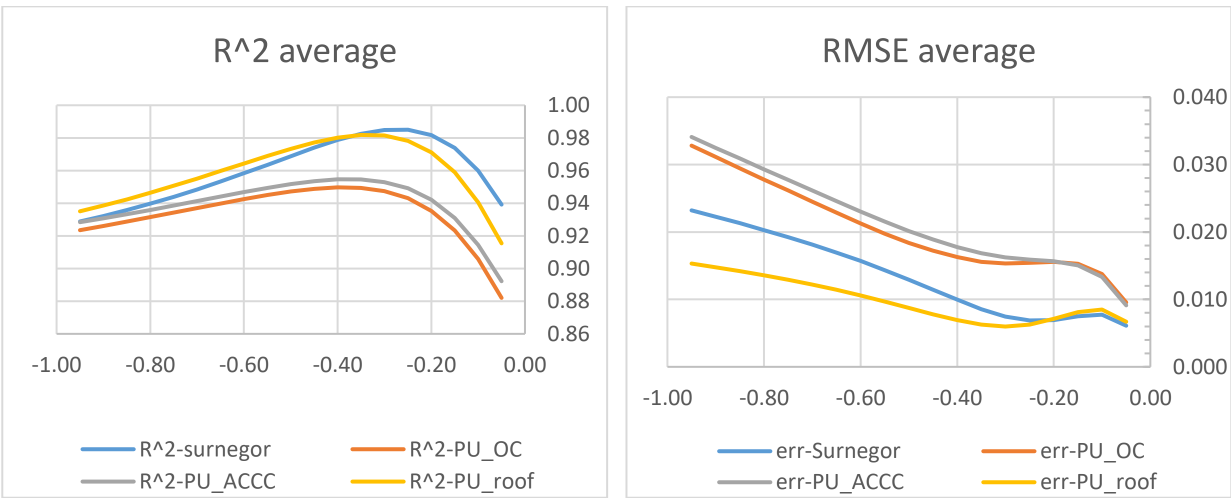

Fig. 6. Average regression coefficient of determination (R^2, top) and RMSE (bottom) as functions of the parameter p, with different colors representing the four experimental locations.

Discussion

Several key factors contribute to the absence of a direct linear relationship between measured mass concentrations (PM_x) and individual FLUKE channels (OPCs):

(i) Variations in particle density and composition: particle mass depends not only on size but also on chemical composition and density. Thus, two particles with the same diameter can have substantially different masses.

(ii) Aerodynamic and optical properties: the FLUKE instrument counts particles based on light scattering, which is affected by particle shape, surface texture, and refractive index rather than by mass directly. As a result, discrepancies arise when correlating optical counts with actual mass concentrations.

(iii) Heterogeneity of size distribution: total particulate mass is often dominated by larger particles (e.g., PM_5 PM_{10}), while particle number concentrations are primarily determined by smaller fractions ($PM_{0.3\text{-}1.0}$). This imbalance between size groups disrupts the expected linear relationship between count and mass.

(iv) Nonlinear behavior at high concentrations: at elevated particle concentrations, optical counting instruments may experience limitations such as coincidence errors, where multiple particles simultaneously pass through the laser sampling volume. This effect can lead to under- or overestimation of particle counts relative to their actual mass.

(v) Influence of environmental conditions: ambient temperature and humidity can cause condensation or evaporation on particle surfaces, altering their effective mass without affecting their numerical concentration. Such effects can further obscure the correlation between count and mass measurements.

(vi) Calibration constraints: FLUKE channels are generally calibrated using standard particles with well-defined optical properties. In contrast, real atmospheric aerosols often differ in composition, morphology, and refractive index, resulting in deviations from the calibration model and reduced measurement accuracy.

Environmental conditions likely contribute to the observed variability in MLRA model accuracy. As shown in the upper section of Table 4, some MLRA models perform poorly, while others exhibit excellent results under seemingly similar conditions. Moreover, differences in experimental locations imply variations in PM sources, formation mechanisms, and consequently in particle chemical composition, morphology, and size distribution. Therefore, a direct MLRA approach should be validated for each specific case, and if successful, applied exclusively to that particular context. The variability in MLRA performance clearly indicates that transferring models between different environmental conditions or locations is not feasible.

Assuming spherical particle geometry and a uniform material density (with an estimated density value ρ), the following mass approximation model, Mass = f(N), can be applied based on a predefined binning of the particle size interval (see Fig. 5, right):

$$\boldsymbol{Mass} = \sum_{i=1}^{bins} \boldsymbol{N P_i \rho} \left(\frac{4}{3}\boldsymbol{\pi size_i^3}\right) \quad (4)$$

where probability values P_i, (based on the corresponding Q_i values) are determined with the parameter p = -0.5. The performance characteristics are quite close to those obtained with the optimal value p = -0.25; however, the analytical form of the function offers additional advantages for further analysis:

$$\boldsymbol{Q_i} = \frac{\mathbf{1}}{(\boldsymbol{a} \times \boldsymbol{size_i} + \boldsymbol{b})^2} \quad (5)$$

$$\boldsymbol{P_i} = \frac{\boldsymbol{Q_i}}{\sum_{k=1}^{bins} \boldsymbol{Q_k}}$$

The density values (ρ) of particulate matter are well documented in the literature and therefore fall outside the scope of the present study. In practice, these values can vary depending on the dominant particle composition and source (for example, mineral dust, soot, or organic aerosols each exhibit distinct densities). While precise density estimation is essential for accurate mass quantification, this work focuses primarily on the methodological framework for deriving particle size distributions and relative mass relationships rather than on compositional variability.

Conclusions

This study presents the design, deployment, and validation of a low-cost, sensor-based air quality monitoring system developed within the METER.AC network. By combining open-source tools, automated data acquisition, and multi-sensor integration, the system provides an affordable yet scalable alternative to conventional high-cost air quality monitoring stations. The comparative analysis between professional and low-cost sensors revealed that while low-cost modules such as the Sensirion SEN55 and Honeywell HPMA115S0 demonstrate good consistency and strong intercorrelation ($R^2 \approx 0.90$), establishing a direct linear relationship with the FLUKE 985 reference particle counter remains challenging. The results indicate that particulate mass concentrations cannot be reliably inferred from single-channel particle count data alone due to complex influences such as particle density, composition, and environmental conditions.

Nevertheless, multivariate linear regression models (MLRA) yielded satisfactory to excellent performance under certain conditions in controlled environments. These findings suggest that site-specific calibration and adaptive modeling approaches can substantially improve the accuracy of low-cost sensor data.

The developed methodology for estimating particle size distribution based on power-transformed particle counts, Np (optimal p around -0.25), provides a promising foundation for further mass approximation models and deeper aerosol characterization.

In summary, this research confirms the potential of low-cost sensor systems as complementary tools in environmental monitoring networks. Continued refinement of calibration models, coupled with large-scale deployment across diverse environments, will enable more reliable, transparent, and FAIR-compliant air quality data collection—enhancing both scientific research and citizen-driven environmental initiatives.

Acknowledgments

This study is financed by the European Union-NextGenerationEU, through the National Recovery and Resilience Plan of the Republic of Bulgaria, project № BG-RRP-2.004-0001-C01.

References

Lee, J., Hopke, P., Holsen, T., & Polissar, A. (2005) Evaluation of Continuous and Filter-Based Methods for Measuring PM2.5 Mass Concentration. *Aerosol Sci. Technol.*, 39(4), 290–303. doi: 10.1080/027868290929323

Feng, G., Yu, M., & Liu, Y. (2025) A comprehensive review of the latest research progress in optical particle counters, *Measurement*, 251, 117177. doi: 10.1016/j.measurement.2025.117177

Shukla, K., & Aggarwal, Sh. (2022) A Technical Overview on Beta-Attenuation Method for the Monitoring of Particulate Matter in Ambient Air. *Aerosol and Air Quality Research*, 22(12), 220195. doi: 10.4209/aaqr.220195

Terziyski, A., Tenev, S., Jeliazkov, V., Jeliazkova, N., & Kochev, N. (2020). METER.AC: Live Open Access Atmospheric Monitoring Data for Bulgaria with High Spatiotemporal Resolution. *Data*, 5(2), 36. doi: 10.3390/DATA5020036

Received: 14.10.2025
Accepted: 10.12.2025